ADDIS ABABA UNIVERSITY

SCHOOL OF GRADUATE STUDIES

ADDIS ABABA INSTITUTE OF TECHNOLOGY

SCHOOL OF CIVIL AND ENVIRONMENTAL ENGINEERING

MSc. THESIS

ON

ESTIMATION OF LOW FLOW QUANTILES IN UNGAUGED RIVER CATCHMENTS

(CASE STUDY OF DEDESSA RIVER BASIN)

FEBURARY 2015

ADDIS ABABA UNIVERSITY

SCHOOL OF GRADUATE STUDIES

ADDIS ABABA INSTITUTE OF TECHNOLOGY

SCHOOL OF CIVIL AND ENVIRONMENTAL ENGINEERING

MSc. THESIS

ON

ESTIMATION OF LOW FLOW QUANTILES IN UNGAUGED RIVER CATCHMENTS

(CASE STUDY OF DEDESSA RIVER BASIN)

BY: TATEK WORKU

ADVISOR: Dr.Ing. DEREJE HAILU

FEBURARY 2015

ADDIS ABABA UNIVERSITY

ADDIS ABABA INSTITUTE OF TECHNOLOGY

SCHOOL OF CIVIL AND ENVIRONMENTAL ENGINEERING

ESTIMATION OF LOW FLOW QUANTILES IN UNGAUGED RIVER CATCHMENTS

(CASE STUDY OF DEDESSA RIVER BASIN)

Thesis Submitted to the School of Graduate Studies in Partial Fulfillment of the Requirements for the Degree of

Master of Science

In

Hydraulic Engineering

By

Tatek Worku Marga

APPROVED BY BOARD OF EXAMINERS

Dr. Ing. Dereje Hailu ----------------------- -------------------------

(Advisor) Signature Date

Dr. Geremew Sahilu ----------------------- -------------------------

(Internal) Signature Date

Dr. Agizew Nigussei ----------------------- -------------------------

(External) Signature Date

Dr. Isayas G/Yohannes ----------------------- -------------------------

(Chairman) Signature Date

DECLARATION AND COPY RIGHT

Tatek Worku Marga, declare that this Thesis 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___________________________

DEDICATION

I dedicate this work to my family, my friends and especially to my parents who are the reasons for who I am!

ACKNOWLEDGEMENT

I would like to express my sincere gratitude to my advisor, Dr.Ing Dereje Hailu, for his unreserved assistance, constructive and timely comments at all stages of my work and also for supplying me relevant materials to carry out the research. I should strongly appreciate his patienceful guidance in a lot of discussions we made on various problems I faced during the course of the work.

I am very grateful to Addis Ababa Institute of Technology, Department of Civil & Environmental Engineering for allowing me to take part in the Master Program and for my employer Oromia Irrigation Development Authority (OIDA) for providing me to learn my MSc. Program. I am also grateful to Ethiopian Road Authority (ERA) and Ministry of Water, Irrigation and Energy for their financial support and their hospitality in all my works.

I also acknowledge, Hydrology Department of the Ministry of Water, Irrigation and Energy and the National Meteorological Service Agency, who supplied me data free of charge.

Finally, I would like to express my warm feeling of appreciation and tank to my friends who helped me in all stages especially to Kidist Endale, Roba Muhyedin, and all staff member of Oromia Irrigation Development Authority (OIDA), who were spiritually with me, and gave me the strength to finalize my duties successfully. Thanks for your encouragement and true friendship! Your friendship meant a great deal to me.

ABSTRACT

Estimates of low flow statistics are required for a variety of water resource applications. At gauged river sites, the estimation of low flow statistics requires estimation of annual n-day minimum stream flows, low flow estimates, as determined by probabilistic modeling of observed data sequences, are commonly used to describe certain stream flow characteristics. Unfortunately, reliable low-flow estimates can be difficult to come by, particularly for gauging sites with short record lengths. The shortness of records leads to uncertainties not only in the selection of a distribution for modeling purposes but also in the estimation of the parameters of a chosen model. In flood frequency analysis, the common approach to mitigate some of these problems is through the regionalization of frequency behavior. The same general approach is applied here to the case of low-flow estimation, with the general intent of not only improving low-flow estimates but also illustrating the gains that might be attained in doing so.

In this study low flow regionalization has been performed using three approaches: flow duration curves, base flow index and low flow frequency analysis. Low flow frequency analysis of Dedessa river basin have been carried out using 7-day annual minimum flow series extracted from the average daily stream flows observed at 10 stream flow gauging sites across Dedessa river basin. L- Moment ratio diagrams were used for identifying and grouping of stations in to hydrological homogeneous regions and hence the underlying statistical distribution. Accordingly, the basin was delineated in to two homogeneous regions. The homogeneity of the delineated regions was tested using the Cv-based and discordance tests. The goodness of fit test (Z-statistics) and the minimum standard error of estimates have been used for assessing the suitability of the selected distribution. Consequently, the Wakebay lower boundary distribution is found to be a suitable candidate for region one and the generalized extreme value distribution for region two. The method of probability weighted moments was considered as the best parameter estimation procedure as compared with the method of moments and the maximum likelihood method.

Flow duration curves of 1, 7, 10, 30 days have been derived for 10 stations. The basins flow duration curves are characterized in to two homogeneous classes based on the characteristics shown on the right tail end of the basins flow duration curves. Base flow separation has been

made by the standard method called the institute of hydrology method. Accordingly the base flow index is calculated and the basin is grouped in to three classes based on their range of BFI. Regression analyses were conducted on the delineated regions to predict the index low flows (mean annual 7-day minimum flow) for ungauged catchments using catchment characteristics. The multiple coefficients of determinations R^2 was used as a measure of the ability of the regression model to describe variations in the dependent variables. The value of R^2 was found to be 0.997 for region one and 0.84 for region two. In addition the standardized low flow frequency curves which will be used for estimating low flow quantiles at ungauged catchments have been established for the two delineated regions using the standardized flow data.

ABBREVIATIONS

LFFC	Low-flow Frequency Curve’s
BFI	Base-flow index
DEM	digital elevation model
FDC	Flow duration curve
LMRD	L-moment ratio diagram
MAM7	Mean annual 7day minimum flow
MOWIE	Ministry of water Irrigation and Energy
MAMF	Mean annual minimum flow
OIDA	Oromia Irrigation Development Authority
NMSA	National Metrological Service Agency
PPCC	Probability plot correlation coefficient
PET	Potential Evapotranspiration
PWM	Probability weighted moments
LFFA	Regional low flow frequency analysis
B/t	Between
RMSE	Root mean square error
DRB	Dedessa River basin
AAit	Addis Ababa Institute of Technology
WMO	World metrological organization

Contents **Pages**

LIST OF FIGURES

LIST OF TABLES

1 INTRODUCTION

1.1 Background

Sustainable water resources planning and management requires data to enable quantification of water quality and quantity. Information is required on the rates of transfers and storage of water within a catchment. Lack of adequate hydrological data introduces uncertainty in both the design and management of water resources systems. As a result of low conversion of rainfall to runoff a precarious balance between available water resources and water demand. The rapid population growth characteristic of the country, which is increasing water demand for domestic, agricultural, and industrial purposes, is causing water scarcity. The magnitude of this scarcity and its variation in both space and time are largely unknown because of lack of hydrological data. Catchment degradation in its various forms continues without effective control measures due partly to uncertainty regarding the adverse effects on water resources. This uncertainty again arises from lack of adequate hydrological data that should enable quantification of effects of specific land use practices on quality and quantity of water resources. In addition, floods and droughts occur with frequencies and magnitudes because of lack of relevant hydrological data. These cause annually major social, economic and environmental tragedies.

There are many possible definition of low flow, in this study low flow is the simple lowest mean flow over durations ranging from one to D days in a year (FREND 1989).

Low flows of a stream vary randomly. Also, the durations of low flows are important. A prolonged period of low flows is the drought. Thus, any analysis of low flows must consider both the magnitude and duration. The duration can be taken into account by averaging the flows over a period of time (1 day, 7 days, 10 days, 30 days, etc.) The most critical low flows on record must be examined for meaningful regulatory decisions. Because the historical record is only a sample from an ensemble of possible occurrences, the analysis of low flows must be based on probability theory. Frequency analysis provides a relationship between events and their probabilities of occurrences.

In general, this relationship is determined in the form of a theoretical probability distribution (Loganathan et al., 1985).

The estimation of flow regimes at un-gauged sites may be achieved by transfer of statistics derived from gauged catchments using regionalization procedures. The term regionalization in hydrology refers to grouping catchments in to homogenous regions.

A homogenous region is one which has sites with similar low flow characteristics and therefore has the same standardized frequency distribution form and parameters. Low flow frequency analysis is used to estimate low flow quintile (Q) at the interest site or project location. The rarity of the low flow event being estimated is expressed by its return period (T) (Mkhandi et al., 1997).

Regional low flow frequency analysis is one of the practical means providing low flow information at sites with little or no local data. In the regional approach, available low flow data series from hydrological homogenous region are pooled in dimensionless form and a frequency distribution is fitted to combined data. However, the main problem of in low flow frequency analysis is the determination of probability distribution that can provide a curve that defines the regional average relation between standardized flow magnitudes (Q_T) and return period (T) (Raymond, 2001).

1.2 Statement of the problem

Due to increased food insecurity and recent food price hikes in the country, it is envisaged that there would be huge demand for irrigated agriculture by the farmers themselves and as a matter of policy priority by the government at all level. This will undoubtedly create huge demand on the water resources particularly during the lean flow season. The extra demand on the river may create undesirable environmental as well as upstream downstream conflict. Therefore, for informed decision making and appropriate policy recommendations on the minimum environmental and downstream requirements, estimation of low flow is required that enables generation of flows at any required reach of the river system, low flow estimates are vital for planning water supplies, water quality management issuing and renewing of waste disposal permits, hydropower, and the impact prolonged drought on aquatic ecosystems. Now a day's extreme low flow events are more diligently analyzed in the emerging field of Eco-Hydrology. Estimation of low flow is also important for small scale irrigation projects that contribute significantly to poverty alleviation by means of increased crop production and generation of rural employment (Adane, 2003).So analysis of low flows would provide an accurate understanding of the demand that may safely be placed on a stream flow.

There is any study haven't been done concerning to low flow characteristic of the dedessa river basin so far. Also the volume and duration of dependable base flow is not quantified. Most of all the availability and quality of information is not adequate so further development of any water resource project within the basin are difficult and unreliable unless the low flow characteristics is well known.

Therefore, this study will try to analyze and characterize low flow of the un-gauged river catchments of Dedessa river basin and provide the necessary information about the low flow of the basin.

1.3 Objective of the research:-

The main objective is to identify and assess the suitability of statistical methods and flow duration curves analysis to estimate low flow characteristics of un-gauged catchments of dedessa river basin.

The specific objectives are:-

- To assess the potential of using hydrological homogenous regions as the basis for estimating low flow characteristics of un-gauged catchments.
- Develop daily, weekly, decadal, monthly flow duration curves.
- Analysis of low flow frequencies.
- Developing base flow index
- Developing regression model

1.4 Description of the study area

1.4.1 Location

The study area, Dedessa sub basin is situated in the south-west part of the Abbay Basin and it is located between 36^0 02’ and 36^0 46’ east longitude, and between 7^0 43’ 55’’ and 8^0 13’ 53’’ North latitude. as shown in Fig 1.1 and the largest tributary of the Blue Nine River in terms of volume of water contributing roughly a quarter of the flow as measured at the Sudan border. Draining an area of nearly 34,000km^2, the Dedessa river basin originates in the Mt. venin and Mt. wache ranges, flowing in an easterly direction for about 75Kms, and then turning rather sharply to the north until it reaches the Blue Nile River. The major tributaries the Dedessa River are the Wama entering from the ester the Dabana from the west and the Anger from the east

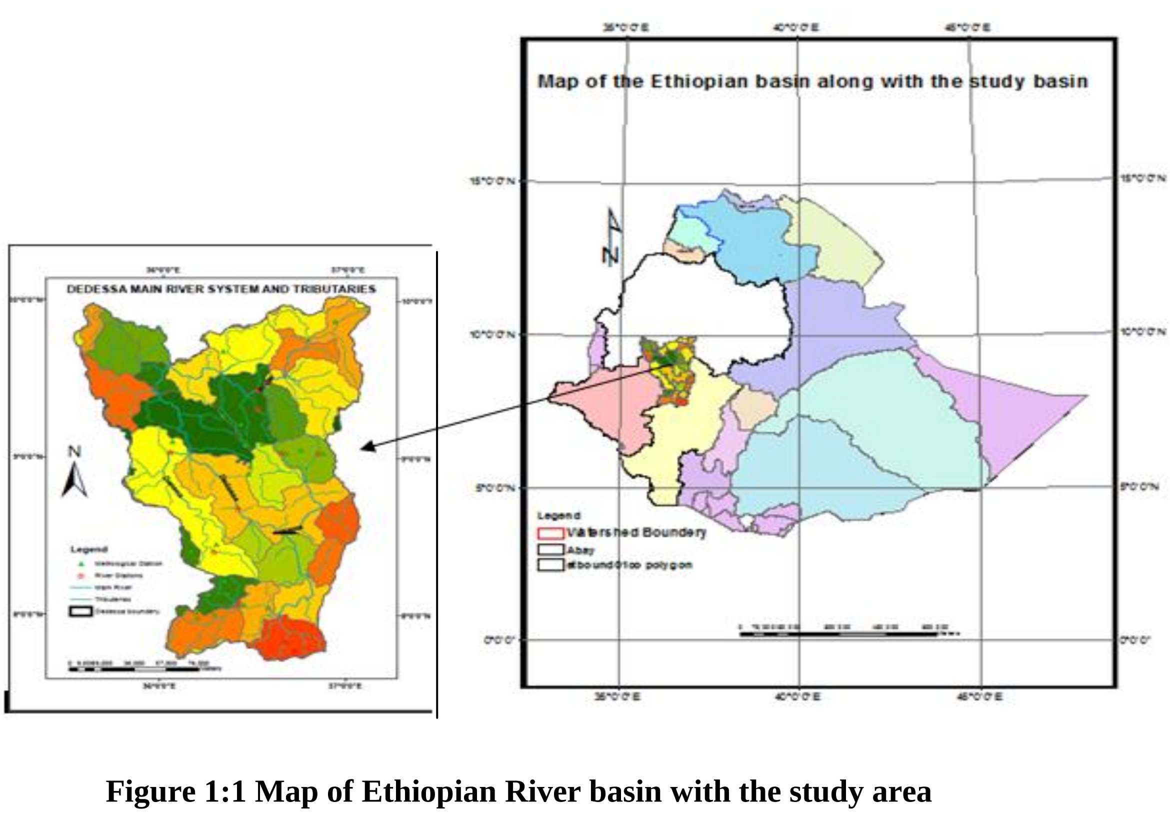


Figure 1:1 Map of Ethiopian River basin with the study area

1.4.2 Climate

The climate of Dedessa river basin is dominated by two factors: its near-equatorial location and an altitude ranging from 626 meters above sea level to more than 3000meters.

The annual rainfall in the basin ranges between 1535 to 2214 mm, average being 1835mm. the area is characterized by a humid tropical climate of heavy rainfall in its upper highlands. Almost 70% of the total annual rainfall is received during kiremt. The maximum and minimum monthly temperature varies between 22.4^0 C-30.9^0 C and 8.4^0 C- 14.4^0 C respectively.

The rainfall in the dedessa river basin is uni-modal type, and most of the rainfall is concentrated in the June to September period with virtual drought from November through February, Annual totals from less than 1500mm to more than 2000mm. the five wettest months and the dry season four months cover 63% and 7% of the total annual rainfall respectively.

1.4.3 Soil

The 2007 soil survey/study and the 2012 soil survey have identified nine major soil units, classified based on the World Reference Base for Soils Resources (ISSS- ISRIC-FAO, 1998). These soil units are Vertisols, Cambisols, Luvisols and Nitisols, as identified by the 2007 survey and Major portion of the soils in the catchment area of Dedessa river basin are vertisols. Occur on flat and almost flat lands, and developed on basic rocks and alluvial deposits. The Vertisols of have clay fraction throughout the profile and the clay content ranges from 24 to 75%. The nature of clay of the Vertisols is extremely hard when dry and very sticky and very plastic when wet. These soils crack widely to more than 15 cm width during the dry season. However, the cracks close up again on rewetting and as a result of this expansion and contraction the soil forms grooved shiny pedfaces, known as slickensides and wedge shaped structure at depth of 20-100 cm.

The seasonal characteristic changes of Vertisols causes the development of micro-reliefe, known as gilgai, comprising low mounds and hollows. The Vertisols are difficult to till and require greater skill in management for better production than other soils.

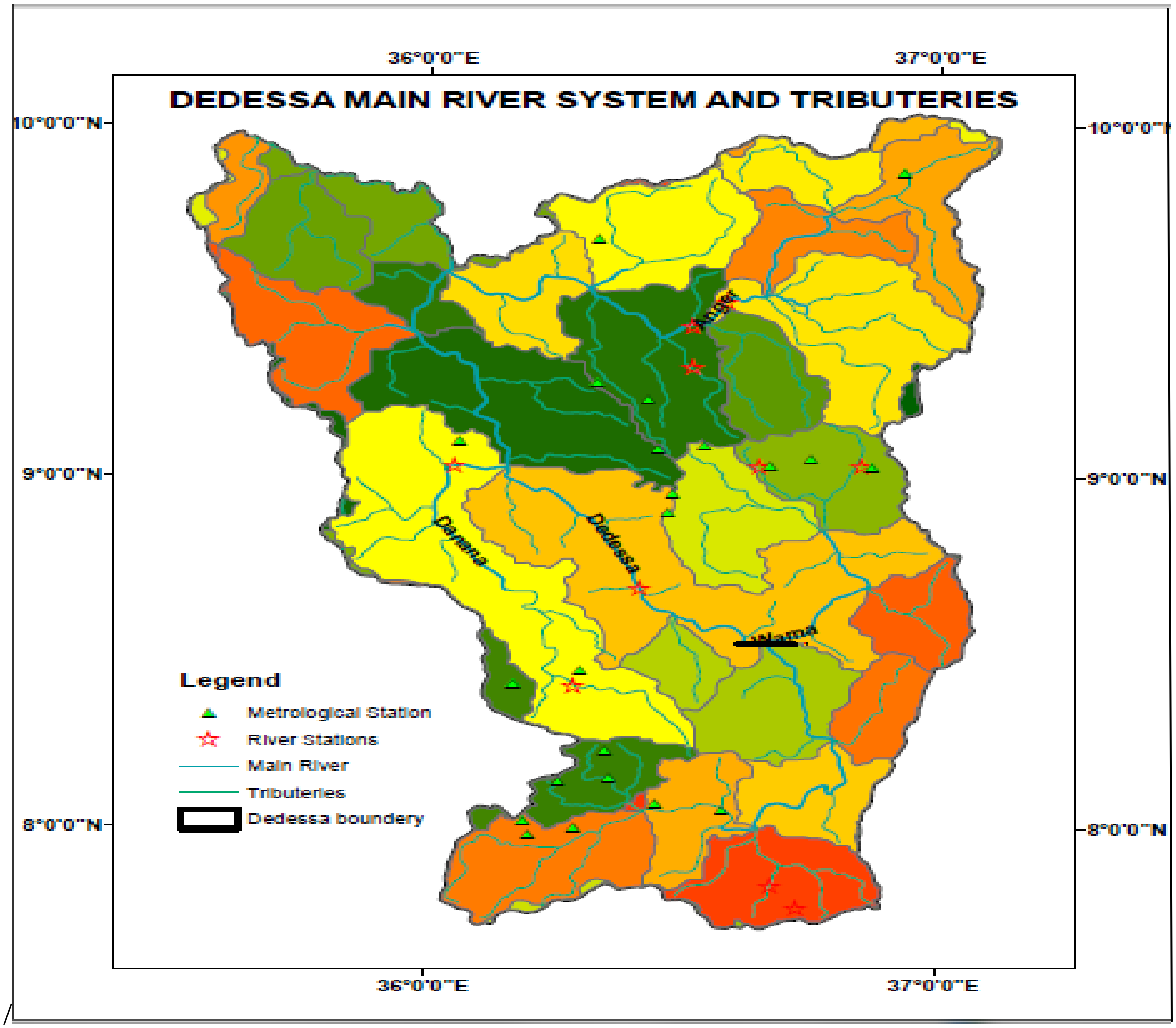


Figure 1:2 Dedessa Main River System & Sub Basin

1.4.4 Land Cover

The dominant land cover of the dedessa river basin is savanna woodland. The flatter areas along the Dedessa River are cover by tall grass while the sloping lands and hilly areas mostly covered by woodland vegetation. In addition, there are riparian forests covering narrow strips of land on banks of Dedessa River and its tributaries. Furthermore, the recently established settlement areas and cultivated lands comprise a significant proportion of the land cover.

Land Form

The dedessa river basin are divided in to thee physiographic regions. These are the dedessa plane ranges between 1270-1400m, upland or chain of ridges ranges between 1400-2500m and plateau greater than 2500m.

The catchment characterized by mountainous highly rugged and dissected topography with deep slops. The lowest part of the catchment is characterized by valley floor with flat to gentle slopes. The dedessa plane is found to the Dedessa River. The chains of ridges are rugged and dissected by the tributaries of Dedessa Rivers. The plateau is found on the top of the basaltic ridges. The dedessa plane is covered by alluvial soils. They are mainly silt to sandy reddish brown and black cotton soils are products of the breakdown of granitoids and basalts. While the plateaus are being made of tertiary basalts the uplands are made up of Precambrian granitoids and tertiary volcanic rocks. Pediments are gently slope erosion surface.

1.5 Methodology and procedure

Generally the study comprises the following methods and procedures.

- Collection of relevant data.
    - Hydrological date
    - Meteorological data
- Topographic and digitized map of the basin.
- Checking of adequacy consistency and independency of the data.
- Computing of statistical parameters of selected stations within the basin.
- Regionalization of the basin into homogenous regions based on statistical value using Arc View GIS.
- Selection of frequency distribution for the delineated regions.
- Derivation of frequency curves using the standardized flow data for all regions.
- Regional homogeneity test describing the delineated region.
- Developing Regression model for estimation of annual minimum flow for each regions.

1.6 Structure of the thesis

This thesis is composed of eight chapters. Chapter 1 gives a brief general back ground and introduction on low flows and a detailed description of the study area It also outlined the methodology and procedures used in this study.

Chapter 2 describes methods of low flow analysis, historical background of low flow frequency analysis, flow duration curve techniques, parameter estimation methods and criteria to choose a frequency distribution. The chapter goes on describing low flow regionalization, low flow estimation of un-gauged catchments, base flow definition and method of base flow separation.

Chapter 3 deals with source and availability of data, screening of low flow data and computation of different statistical parameters.

Chapter 4 presents the results of FDC's and base flow groupings.

Chapter 5 gives result of regionalization that is identification of homogeneous region and test of homogeneity of region, delineation of the basin into homogeneous low flow regions is described ,the methods used ,description of the delineated regions is also given.

Chapter 6 outlines the identification of the underlying statistical distribution of low flows, selection and evaluation of best fit statistical parent distribution of low flow, the use of L-moment diagram approach is explained and derivation of regional low flow frequency curves.

Chapter 7 deals with regional regression models of low flow and the result is presented in this chapter.

Chapter 8 gives a summary, conclusions and recommendations of the work.

1.7 Limitations

World Meteorological Organization recommended as minimum record length of data of 30 years' duration or more against which for estimation and prediction of low flow analysis. Owing to data constraint, this study considered an average of 10 years and above as a baseline period. The limitation of short record data is that the data are unlikely to capture the natural variability of the catchment, also insufficient number and distribution of gauging station in the Dedessa river basin difficult to regionalize the catchment in to additional homogeneous region/group.

2 LITERATURE REVIEW

2.1 Introduction

There are three primary scenarios when low flow information is required. The first is when a water resources scheme is being developed. This normally requires a pre-feasibility study to determine whether the objectives of the proposal can be achieved. For scheme with huge capital investment, this is followed by much more detail design work, including estimate of the frequency of low flow. The second is during operational phase of a water resource scheme once it is constructed, and including decision how to manage the scheme on the day by day basis. For example it may be necessary to determine how much water can be diverted from river for hydropower purpose without infringing legal abstraction conditions. These may be depends on the time of the year or on the river discharge at the downstream point. In many countries, such operational constraints are not formalized and the operator must be aware of, and sympathetic to, the needs of many downstream water users. For example, water could be needed by households, or agriculture, electricity production, navigation, industrial abstraction, tourism, the dilution of industry or domestic effluent, maintaining ecosystem for food production, or maintaining natural biodiversity of the river. The third scenario is when it is necessary to make operational decision today based on estimates of future river flows which looks days, weeks, or sometimes months ahead. These forecasting increasing the efficiency of water uses and are economic importance in terms of reducing the operational cost of water resource scheme.

Historically, most applications of low flow information have been in design and operation of schemes for specific water sector such as public water supply, irrigation, energy, navigation and industry. It is now recognized that there is a need to provide a long-term baseline monitoring and analysis low flow to support integrated river basin management. This provides a framework for environmental agencies to make decisions regarding the catchment wide development of water resource.

2.2 Definition of low flow

A discussion of low-flow hydrology and related issues should ideally begin with a definition of what 'low flow' really is. This term could mean different things to different interest groups. To many it may be considered as the actual flows in a river occurring during the dry season of the year, others may be concerned with the length of time and the conditions occurring between flood events. Yet others may be interested in the effects of changes in the total flow regime of a river on sustainable water yield and riparian ecology. The latter may perceive 'low flows' not only as discharges occurring during a dry season, but as a reduction in various aspects of the overall flow regime.

International glossary of hydrology (WMO, 1974) defines low flow as 'flow of water in a stream during prolonged dry weather'. This definition does not make a clear distinction between low flows and droughts. A low flow is a seasonal phenomenon, and an integral component of a flow regime of any river. Drought, on the other hand, is a natural event resulting from a less than normal precipitation for an extended period of time. But here in this study low flow is the simplest lowest mean flow over durations ranging from one day to D days in a year (FREND, 1989).

2.3 Methods of low flow analysis

Low-flow regime of a river can be analyzed in a variety of ways dependent on the type of data initially available and the type of output information required. Consequently there exist a variety of low-flow measures. The term 'low-flow measure' used here, refers to the different methods that have been developed for analyzing, often in graphic form, the low-flow regime of a river.

The most common means of quantifying low-flow information of streams is with, statistical estimates of the magnitude and frequency of occurrence. Methods using regression analysis to relate low-flow frequency characteristics and selected basin characteristics could provide significant benefits for managers responsible for protecting surface-water quality and allocating surface-water supplies. Low-flow frequency characteristics with different consecutive-day averaging periods and recurrence intervals provide quantitative information that can be used in the management of a variety of additional water-quality and supply problems

2.3.1 Flow duration curve

The cumulative frequency distribution of daily mean flow shows the percentage of time during which specified discharge are equaled or exceeded during the period of record. The relationship is normally referred to as the flow duration curve and although it does not convey any information about the sequencing properties of flows, it is one of the most informative methods of displaying the complete range of river discharges from low to flood flows. The curve is most conveniently derived from daily discharge data by assigning daily flow values to class intervals and counting the number of days within each class interval. The proportion of the total number of days above the lower limit of any class interval is then calculated and plotted against the lower limit of the interval. Sometimes in plotting the flow duration curve the discharge axis is standardized by the mean flow. This facilitates comparison between catchments. Because it reduces differences in the location of the flow duration curve which are caused by differences in the mean annual runoff.

FDC may also be obtained by applying a D day moving average to the original hydrograph. This enables the portion of D day periods when the average discharge over D days is greater than a given value to be estimated. The low flow indices derived from the flow duration curve would be Q 95(1), Q95 (10) Q70 (1), Q70 (10) etc….This is determined based on countries adoption and the purpose of the indices.

A flow duration curve presents the relationship between the magnitude and frequency of daily, weekly, monthly or any other interval of stream flow. It shows the percentage of time during which specified discharges are equaled or exceeded during the period of record. Generally a normal probability scale is used for the frequency axis and a logarithmic scale for the discharge axis. Because of the assumed lognormal distribution of daily discharges, these transformations assist in linear zing the plot. A one-day flow duration curve can be derived manually. The method involves assigning daily flows for the complete period of record to class intervals and counting the number of occurrences within each interval. The total number of occurrences above the lower limit of each class interval is expressed as a percentage of the total number of days in the record and this percentage is plotted against the lower limit of the interval to give the percentage of time flow is equaled or exceeded. Flow duration curves of other durations can be obtained by applying a D-day moving average to the original hydrograph

2.3.2 Low flow frequency analysis

Unlike the FDCs, which shows the proportion of time during which a flow is exceeded, a Low-flow Frequency Curve's (LFFC) shows the proportion of years when a flow is exceeded or equivalently the average interval in years ('return period' or 'recurrence interval') that the river falls below a given discharge.

LFFC is normally constructed on the basis of a series of annual flow minima (daily or monthly minimum discharges), which are extracted from the available original continuous flow series (one value from every year of record). Similar to FDCs, LFFCs may be constructed using the flow minima series of different averaging intervals. In the case of daily data, the minima of 1, 3, 7,10, 15, 30, 60, 90, 120, 150, 180 or 183, 273 and 284 may be analyzed FREND (1989).

The independence of extracted annual flow minima averaged over long intervals (i.e. over 7 days) must be ensured. The average of the annual series of minimum 7-day average flows is known as Mean Annual 7-day Minimum flow (MAM7) Gustard (1992).

The 7-day period covered by MAM7 eliminates the day-to-day variations in the artificial component of the river flow. Also, an analysis based on a time series of 7-day average flows is less sensitive to measurement errors. At the same time, in the majority of cases there is no big difference between 1-day and 7-day low

flows. For seasonal low-flow frequency analysis, minimum flow values are selected from the season of interest (winter or summer low flows). FREND (1989) have examined Weibull, Gumbel, Pearson Type III, and log- normal probability distributions for fitting the sequences of annual minimum flows and methods for estimating distribution parameters in different regions and for minima of different averaging intervals.

Low flow frequency analysis is one of the practical means of providing low flow information. In the low flow frequency analysis, available low flow data series are fitted to a frequency distribution. However, the main problem of in low frequency analysis is the determination of probability distribution that can provide a curve that defines the regional average relation between standardized flow magnitudes (QT) and return period (T) Raymond (2001).

The slope of LFFC may be considered as a low-flow index. It is represented by the difference between two flow values (normalized by the catchment area), one from high and another from low probability domains. The larger this slope is, the more variable is the flow regime of the river and vice versa. The issue, which often arises in low-flow frequency analysis, is that the

observed stream flow time series (and, consequently, low-flow time series) often contain zero flow values. In arid climates stream flow may naturally frequently fall to zero. Similar situation arises in cold regions, where the streams may be completely frozen in winter. On the other hand, the recorded zero flows are often those which are below the measuring limit of a stream flow gauge (censored flows). Zero values may not be ignored in statistical analysis of low-flow series.
Distributions fitted to series with zero flows will result in a positive probability of negative stream flows, unless the distribution is explicitly constrained to have a lower bound of zero. Such results are physically meaningless. Also, constraining some distributions to have a zero lower-bound may reduce their flexibility. Fitting a probability distribution for low-flow series containing zero values may be performed using conditional probability adjustment procedure suggested by Haan (1977). An extra parameter describes the probability of encountering a zero flow in a sample, while a continuous distribution is established for non-zero flows. The results are then adjusted to the full sample. Low-flow frequency indices are widely used in drought studies, design of water supply systems, estimation of safe surface water withdrawals, classification of Streams' potential for waste dilution (assimilative capacity), regulating waste disposal to streams, maintenance of certain in-stream discharges, etc. Numerous specific indices may be obtained from a LFFC. These indices are similar to characteristics derived from FDC, but unlike FDC, time sequencing of discharges used in the analysis is not disturbed. The most widely used indices are 7-day 10- year low flow (7Q10) and 7-day 2-year low flow (7Q2), which are defined as the lowest average flows that occur for a consecutive 7-day period at the recurrence intervals of 10 and 2 years, respectively.

2.3.3 Definition and history of flow frequency analysis

A flow frequency curve shows the portion of years when a flow is exceeded, or equivalently the average interval in years that a river falls below a given discharge.
Application of techniques on flow frequency analysis goes back to 1914 when Fuller (1914) derived empirical frequency curves from annual maximum flood peaks by graphical method. From here further developments were made with analytical methods to get different types of distributions such as the skewed Pearson Type III distribution, Foster (1924) and Log-Normal distribution by Hazen (1924).In low flow frequency analysis, Gumbel (1941) introduced the extreme value theory to hydrology. The statistical method was boosted further by the

development of an extreme value probability paper by Powell (1943), which he used to construct flood frequency curves. In the development of low flow frequency curves, Matalas (1963) investigated the fitting of four theoretical distributions to low flow data.

Riggs (1968) prepared a very useful review of frequency curves and later produced a manual that describes in detail the procedures for defining and evaluating low flow frequency curves.

2.3.3.1Statistical distributions

Several statistical distributions are available for use to date but the problem arises in choosing the right one to fit the observed annual minimum flow data. Some of the distributions used for estimating low flow quantiles with a particular probability of exceedance or non exceedance are:

1. Normal distribution (N)
2. Log-Normal (LN2, LN3) (Hazen, 1914)
3. Exponential distribution (EXP)
4. Gamma distribution (G2)(Moran, 1957)
5. Pearson Type III (P3) (Foster, 1924)
6. Log Pearson Type 3 (LP3) (USWRC, 1967)
7. General Extreme Value (GEV) (Jenkinson, 1955)
8. Gumbel (EV1) (Gumbel, 1941)
9. Wake by (WAK4. WAK5) (Houghton, 1977)
10. Weibull (Wu & Goodridge, 1959)
11. Generalized Pareto distribution (Gpar)
12. Log - Logistics (LLG) (Ahmed *et al*, 1988)

2.3.2.1.2 Parameter estimation methods

Selection of a distribution is followed by estimation of parameters. Since the Parameters are to be estimated from the sample data, the estimates are subject to sampling errors. The three most popular parameter estimation methods used are:

1. Method of Moments (MOM)

This method is one of the most commonly and simply used method for estimating the parameters of a probability distribution. However, it is well known that some moment estimates are biased, especially for the cases of short sample series or higher order moments (Wallis 1974).

It is confirmed theoretically (Kirby, 1974) that coefficient of variation (CV) and skewness (Cs) in random samples are bounded and that the bounds are a function of sample size. Much research work has been done in deriving factors or functions by which the bias can be corrected.

2. Maximum Likelihood (ML)

In this method the parameter estimates are determined by maximizing the sample log likelihood function. The unknown parameters may be obtained by setting each of the partial derivatives with respect to each parameter equal to zero and solving the resulting equations simultaneously. These equations unfortunately do not often take a simple closed form and the numerical solutions have to be used. Even though the maximum likelihood estimator is asymptotically unbiased, sufficient and consistent, sometimes there is a failure to obtain proper solutions due to the complexity of log likelihood functions, particularly when sample size is small or when the distribution has more than two parameters. This method is statistically the most efficient one in large samples.

3. Probability Weighted Moments (PWM)

Probability weighted moments (PWM) were introduced by Greenwood et al. (1979) and further analyzed by Hosking (1986). PWMs were proposed for the purpose of derivation of expressions for the parameter of distributions whose inverse forms $x = x(F)$ can be explicitly defined, such as Extreme Value type 1, General Extreme Value, Log Logistic and Wakeby distributions. The use of PWMs was later extended to be used to estimate P3 parameters (Song and Ding, 1988) and other distributions whose inverse forms cannot be explicitly defined (Hosking, 1986).

The PWM method has recently come to be regarded as the best method by research hydrologists (Hosking, 1986). This procedure leads to relatively unbiased quantile estimates due to the fact that the PWMs being linear functions of the data, suffer less from the effects of sampling variability than do the conventional moments, because the latter involve higher powers of the data.

A study by Gunasekara (1990) shows that regional estimates by PWMs yield quintile estimates with low variability and low bias.

2.3.3.2 Quantile Estimation

After a distribution or a number of distributions are selected to fit the data, their parameters must be estimated. The estimated parameters are used to calculate quantile estimates (Q_T) for different return periods.

This is achieved by using the distribution function in which the parameter of the distribution are replaced by their estimates and the relationship between return periods (T) and the probability of exceedance (F) is expressed as

$$F=1/T \quad \ldots\ldots 2.1$$

Where, F=F (Q_T) is the probability of having a low flow of magnitude Q_T or larger.

Thus the problem reduces to evaluating Q_T for a given value F.Chow (1964) proposed a general form for calculating QT as

$$Q_T = U_1' + K_{T*}\sqrt{\mu_2} \quad \ldots\ldots 2.2$$

Where $K_{T=}$ frequency factor which is a function of return period & parameter of distribution.

U_1', μ_2 = moments of distribution which can be calculated by using the estimated parameters.

2.4 Selection of the most robust low flow estimation method

Descriptive ability tests were applied for evaluating the proposed statistical distribution procedures and come up with the most robust low flow frequency model for each region in the basin. In addition to descriptive ability test selection of the most efficient method which gives the smallest standard error of estimate of different statistical distribution are used.

2.4.1 Descriptive ability tests

The indicators that are used in selecting one procedure from another from descriptive ability point of view vary from the traditional chi – square and kolmogorov – simronov tests to the recently introduced L- moment ratio diagrams.

The objective of all indicators is to assist the hydrologist in the choice of a distribution that best describes a set of sample data.

Some recent techniques are introduced that can be grouped under descriptive ability tests. One of them is goodness of fit, here; this method is discussed as a methodology in discriminating the best statistical distribution.

Goodness of fit measure

For the already identified regions, goodness – of – fit measure (Z) helps to test whether a given distribution fits the data acceptably or not. The common approach is to choose a distribution from a number of distributions that fits the data best. This measure is based on L-moments of the at-site data. If the observed sites do form a homogenous region

(a) L- moments of the sites are well summarized by regional average, indicating that the scatter of the individual sites' moments about regional average represents no more than sampling variability and

(b) It can be assumed that the region's population L – moments are likely to be close to the average of the sample L – moments of the observed data.

In most cases the distribution being tested will have location and scale parameters which can be chosen to match the regional average mean and L – CV. The goodness of fit will therefore be judged by how well the L – skewness and L – kurtosis of the fitted distribution match the regional average L – skew ness and L – kurtosis of the observed data.

For a chosen distribution, the goodness of fit measure Z^{dist} is defined as

$$Z^{dist} = \left(\frac{\tau_4^{dist} - \bar{t}_4 + \beta_4}{\sigma_4} \right) \qquad \text{..2.3}$$

Where $\bar{t}_4$ is the average L – kurtosis value computed from the data of a given region, σ_4 denotes the standard deviation of $\bar{t}_4$, which can be obtained by repeated simulation of a homogenous region with the distribution under test and sites having low flow record lengths the same as those of the observed data.

β_4 is the bias of $\bar{t}_4$ and τ_4^{dist} is the average L – kurtosis value computed from simulation for a fitted distribution.

For short record lengths (n ≤ 20) and large population L skewness $(T_3 \geq 0.4)$, a bias correction factor for t_4 is required. Hence in a goodness of fit test the fitted L – kurtosis τ_4^{dist} is compared not with the regional average $\bar{t}_4$ itself but with the bias corrected version $\bar{t}_4$ - β_4 .

The steps involved in a goodness of fit test are:

Step 1. Consider a distribution.

Step 2. Fit each distribution to the group average L – moments

$\bar{t}$, *and* $\bar{t}_3$ .Denote the L – kurtosis of the fitted distribution by τ_4^{dist}

Where,

$$\bar{t} = \frac{\sum_{i=1}^{N} n_i \; t^{(i)}}{\sum_{i=1}^{N} n_i} \qquad \ldots\ldots 2.4$$

$$\bar{t}_r = \frac{\sum_{i=1}^{N} n_i \; t_r^{(i)}}{\sum_{i=1}^{N} n_i} \qquad r = 3, 4 , \ldots\ldots 2.5$$

$t^{(i)}$ = L – Moment ration at site i $\left(t^{(i)} = \frac{L_2}{L_1} \right)$

L_2 and L1 are as defined in chapter 2 sections 2.5.1.2

N = number of sites in a region

n_i = record length at site i.

Step 3. Fit the considered distribution to the group average L – moments.

Step 4. Simulate a large number $Nsim_1$ (here 500) of regions from the fitted distribution. For the mth simulated region calculate the regional average L – skewness $\bar{t}_3^{(m)}$ and L – kurtosis $\bar{t}_4^{(m)}$.

Step 5. Calculated the bias and standard deviation of $\bar{t}_4$ using the following expressions.

$$\beta_4 = N_{sim1}^{-1} \sum_{m=1}^{N_{sim1}} \left(\bar{t}_4^{m} - \bar{t}_4 \right) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 2.6$$

and

$$\sigma_4 = \left[\left(N_{sim1} - 1 \right)^{-1} \left(\sum_{m=1}^{N_{sim1}} \left(\bar{t}^{4m} - \bar{t}_4 \right)^2 - N_{sim1} \beta_4^2 \right) \right]^{1/2} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots .. 2.7$$

Small values of Z^{dist} imply that the considered distribution can be accepted as the true underlying frequency distribution for the region. According, to Hosking and Wallis (1993) if $\left| Z^{dist} \right| \leq 1.64$, the distribution is acceptable.

2.) Standard error of estimates

The standard error of estimate accounts for the error due to small samples, but not the error due to the choice of inappropriate distribution. The standard error of estimate depends in general on the method of parameter estimation. Consequently, each method gives a different standard error of estimate; the most efficient method is that which gives the smallest standard error of estimate (Rao.A.R, 2000).

2.5 Low flow statistics

Preference to L-moment statistics (Lcv, Lcs, Lck) over conventional moments is given because of the fact that conventional moments exhibit substantial bias and variance for the small samples encountered in hydrological applications. L- moment method is a powerful and efficient method to compute any statistical parameters, Also it cannot be influenced with the presence of outliers (Roa &Hammed, 2000).

- The statistical parameters computed includes
- Mean
- Standard deviation
- Coefficients of Skewness
- Coefficients of Variations and
- Coefficients of Kurtosis

2.5.1 Moments of distributions and their sample estimates

2.5.1.1 Conventional moments

Moments about the origin or about the mean are used to characterize probability distribution .Moments about the origin are the expected values of power of random variable .For a distribution with a probability density function f(x), the rth moment about the origin is given by (Roa., 2000)

$$\mu'_r = \int_{-\infty}^{\infty} x^r f(x)\, dx, \qquad \mu'_1 = \mu = \text{mean} \ldots\ldots 2.8$$

The central moment's μ_r are computed by Eq.2.8

$$\mu_r = \int_{-\infty}^{\infty} (x - \mu')^r \text{ f (x) dx}, \qquad \mu_1 = 0 \ldots\ldots 2.9$$

It can be easily proved that the relationship between μ'_r and μ_r is given by Eqs 2.8 and 2.9 (Kendall and Stewart, 1967).

$$\mu_r = \sum_{j=0}^{r} \binom{r}{j} \mu'_{r-j} \; (-\mu'_{r-1})^j \ldots\ldots 2.10$$

$$\mu'_r = \sum_{j=0}^{r} \binom{r}{j} \mu_{r-j} \; (\mu'_{r-1})^j \ldots\ldots 2.11$$

Sample moments $m_r^`$ and $m_r^`$ on the other hand are calculated by using Eqs.1.7.and 1.8

$$m_r^` = \frac{1}{n} \sum_{i=1}^{n} x_i^r \quad m'_1 = \bar{x} = \text{sample mean} \ldots\ldots 2.12$$

$$m_r = \frac{1}{n} \sum_{i\ 1}^{n} (x_i - \bar{x})^r, m_1 = 0 \ldots\ldots 2.13$$

The relationship as in Eqs.2.8 and 2.9 hold for sample moments as well. These sample moments are often biased and may be corrected (Cunnane, 1989). For example, some of the corrected central moments are given below

$$\hat{m}_2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2 \ldots\ldots 2.14$$

$$\hat{m}_3 = \frac{n}{(n-1)(n-2)} \sum_{i=1}^{n} (x_i - \bar{x})^3 \qquad 2.15$$

$$\hat{m}_4 = \frac{n^2}{(n-1)(n-2)(n-3)} \sum_{i=1}^{n} (x_i - \bar{x})^4 \qquad 2.16$$

However, in small samples the bias may be larger than can be properly corrected by using simple expressions in n. The conventional moment ratios are defined as below.

Moments of distributions and their sample Estimates

Coefficient of variation Cv = $\mu_2^{1/2} / \mu'_i$ ……………………………………………2.17

Coefficients of skewness C_s = $\mu_3 / \mu_2^{3/2}$ ………………………………………2.18

Coefficient of kurtosis C_k = μ_4 / μ_2^{2} ……………………………………..2.19

2.5.1.2L-moments

Hosking (1990) introduced the L-moments, which are linear functions PWMs. The L-moments are more convenient than PWMs because they can be directly interpreted as measure of scale and shape of probability distribution. In this respect they are analogous to conventional moments.

L-moments have the following advantage (Cunane,1989).

Compared to conventional moments, L-moment can characterize wide range of distribution

Sample estimates of L-moments are so robust that they are not affected by the presence of outlier in the data set

They are less subjected to bias in estimation

L-moments yield more accurate estimates of the parameters of a fitted distribution.

As Hosking (1986) defined the L-moments of a real value random variable x in terms of probability weighted moments (PWMs) as:

$$L_r = (-1)^{r-1} \sum_{k=0}^{r-1} P_{r-1,k} M_k \qquad 2.20$$

Where

$$P_{r,k} = (-1)^{r-k} \binom{r}{k} \binom{r+k}{k} \qquad 2.21$$

L^r = the r th L- moment

$$M_{10k} = \sum_{j=0}^{k} (-1)^j \begin{pmatrix} k \\ j \end{pmatrix} M_{1j0} \quad \ldots\ldots 2.22$$

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

$$\hat{M}_{10k} = \frac{1}{N} \sum_{i=1}^{N} \left[\begin{pmatrix} N-i \\ k \end{pmatrix} / \begin{pmatrix} N-1 \\ k \end{pmatrix} \right] x_i \qquad k = 0,1,2 \ldots\ldots N-1 \ldots\ldots 2.23$$

$$\hat{M}_{1j0} = \frac{1}{N} \sum_{i=1}^{N} \left[\begin{pmatrix} i-1 \\ j \end{pmatrix} / \begin{pmatrix} N-1 \\ j \end{pmatrix} \right] x_i \qquad j = 0,1,2,\ldots N-1, \ldots\ldots 2.24$$

where i = rank of observed flow data in ascending order.

The first few moments are:-

$L_1 = M_{100}$

$L_2 = M_{100} - 2* M_{101}$

$L_3 = M_{100} - 6* M_{101} + 6* M_{101}$

L_4= M100 - 12* M_{101}- 30* M_{102} - 20* M_{103}

Like the conventional moments, L- moments can be used to specify and summarize probability distributions. In particular L_1, the first L- moment, is the mean of a statistical distribution and is identical to the first conventional moment, and L_2 is a linear measure of spread or dispersion analogous to standard deviation. L-moment ratios, which are analogous to conventional moment ratios, are defined by Hosking (1990), Roa & Hamed, 2000) as:

$$t = L_2 / L_1 \quad \ldots\ldots 2.25$$

$$t_r = L_r / L_2 \quad r \geq 3 \quad \ldots\ldots 2.26$$

Where L_1 = measure of location

r = measure of scale and dispersion (LC_V)

r_3 = measure of skewness (LC_k)

r_4 = measure of kurtosis (LC_k)

2.6 Low flow regionalization

Regionalization refers to grouping catchments into homogenous regions. A homogenous region is one which has sites/stations with similar low flow characteristics and therefore has the same standardized frequency distribution form and parameters. Moreover such a region must be geographically continuous.

Several studies have shown that, delineation of regions in the past has often relied on physiographic, political or administrative boundaries (Getachew, 1996).The resulting regions were assumed to be homogeneous in terms of hydrologic response.

This assumption actually is not true as it may have very different relief and stations within the same geographic region, which have high correlation that will cause some bias in the regionalization (Wiltshire, 1985: Cunnane, 1989, Roa & Hamed, 2000).The importance of homogeneity has been demonstrated by Hosking et al (1985), Wiltshire (1986).

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

Raymond (2001) tried to regionalize Tanzania In to homogeneous region and he grouped Tanzania in to eight homogeneous region based up on the low flow statistics of the gauged station.

In regionalization, a regional frequency curve is derived. Regional frequency curves show the portion of years when a flow is exceeded, or equivalently the average interval in years that a river falls below a given discharge for those region.

The curve is derived from annual minimum discharge of seven days extracted from daily flow time series. In response to the practical requirements for estimating the flood discharge of extreme events having return periods several times greater than the length of hydrological records is used but for low flow usually return period in the order of fifty years or less is used. In frequency analysis of low flow, two values are associated. The first value is the duration of the low flow used in the generation of the sample and the second is the desired return period. For example, if duration of seven days is used along with a ten-year return period, the calculated discharge (Q), will be the 7-day, 10-year low flow.

No matter how the regionalization is done, checking the homogeneity of stations (which made a homogeneous region) by some homogeneity test is a must.

2.6.1 Procedure for regionalization

The grouping of homogeneous regions is done with the aim of achieving regional low flow frequency relationships with lower standard errors than those of combined area (Kite, 1977).

The step-by-step procedure is as follows

1. Statistical values of each station under analysis (Lcs, Lck) for the flow is computed
2. LMRD of each station is drawn
3. Moment ratio diagram of each station together with the LMRD of the distribution are plotted.
4. Station that will lie in the same region are identified from the graph

This is done on the basis that station from the same homogenous region will be distributed along the same distribution of LMRD. (Dalrymple, 1990, URL)

2.6.2 Criteria for Low Flow Regionalization

2.6.2.1 Homogeneity

Regardless of the way regions are formed they have to satisfy the homogeneity criteria (Zrinji & Burn, 1996). Each of the regions should have two basic properties: dissimilarity from other regions and homogeneity of low flow characteristics within the region. Hosking et al (1985), Wiltshire (1986) and Lettenmaier et al. (1987) have demonstrated the importance of regional homogeneity. There are several tests available to examine regional homogeneity in terms of the hydrologic response of the stations in a region.

Another way of testing is by the use of the likelihood ratio test of whether the frequency of an event can be represented by a single regional frequency relationship (Acreman and Sinclair, 1986). Wiltshire (1986) suggested the use of the R-statistic, which is based on similarity of the geometry.

Later, Vogel and Kroll (1989) proposed the test, which is based on the probability plot correlation coefficient (PPCC). Hosking and Wallis (1991) suggested a homogeneity test, which is based on L-moments.

Their test statistic is a standardized measure of the inter site variance of the Lcv. Tests used for determining the degree of homogeneity of regions are given below.

- Cv and L-cv Homogeneity test
- Discordance measure test

2.6.2.1.1 CV based test

A simple test based on the variability in at site Cv values is used .The main statistic considered is the coefficient of variation of the Cv of the individual sites within the region (CC). In a region of m stations and n_j years of records at the j th station, having a standardized low flow discharges (q_{ii},i= 1,2,3,…. n $_j$) the coefficient of variation is defined as.

$$cv_j = \frac{\left[\sum_{j=1}^{nj} \frac{(q_{ji} - \bar{q}_j)^2}{n_j - 1} \right]}{\bar{q}_j} \quad \ldots\ldots 2.27$$

The weighted regional coefficient of variation of all the Cv_j (CC) is defined as

$$CC = \frac{1}{\underline{\bar{\bar{cv}}}} \sum_{j=1}^{m} \frac{(cv - cv^{-})^2}{(m-1)} \quad \ldots\ldots 2.28$$

The procedure consists of setting up repeated trial grouping of sites until the best possible set is obtained. Sites with approximately close cv_j values and lying within the same geographical proximity are first grouped together in one region and the internal homogeneity of the regions expressed numerically in terms of the flow statistics CC determined.

Depending on the value of cv, sites were transferred from one region to the other until the right balance was achieved. The criterion used to check for regional homogeneity was based on the value of CC. For homogeneous regions a value of 0.3 or less was allowed.

2.6.2.1.2 Discordance Measure Test

The discordance measure is intended to identify those sites that are grossly discordant with the group as a whole. The discordance measure D estimates how far a given site i form the center of the group. (Roa & Hamed, 2000)

If $U_i = \left[t^{(i)}, t_3^{(i)}, t_4^{(i)} \right]^r$ is the vector containing the t, t_3 and t_4 values for site (i) ,then the group average for NS sites within the region is given by

$$\overline{U} = \frac{1}{NS} \sum_{i=1}^{NS} U_i \quad \ldots\ldots 2.29$$

The Sample covariance matrix is given by

$$S = (NS - 1)^{-1} \sum_{i=1}^{NS} \left(U_i - \overline{U}\right)\left(U_i - \overline{U}\right)^T \quad \ldots\ldots 2.30$$

The discordance measure is defined by

$$D_i = \frac{1}{3} \left(U_i - \overline{U}\right)^T S^{-1} \left(U_i - \overline{U}\right) \quad \ldots\ldots 2.31$$

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

2.7 Base flow

2.7.1 Definition and Introduction

Base flow is generally considered to be that component of flow which originates from stored sources within the catchments, principally groundwater aquifers, as distinct from that flow component which is rapidly transmitted to streams following rainfall events.

Base flow component tend to increase in wet season due to the rise of the depth of saturated subsurface flow. The base flow component of river flow is commonly expressed as a proportion of the total river flow, termed the Base Flow Index (BFI). A catchment's with BFI approaching 1 is highly dominated by base flow, while a catchment's with BFI approaching 0 receives little base flow contribution.

2.7.2 Base flow Separation Methods

In order to determine the contribution from overland flow in a watershed to the streams in the watershed or the contribution of flow from the stored sources it is necessary to separate out the base flow from stream gage data. The base-flow index is the total volume of base flow divided by the total volume of runoff for a period (Wahl, 1995).

Hydrologists are often asked to separate runoff into the components that result either from direct runoff or from ground-water discharge. However, doing so is necessarily an inexact science.

The processes that produce flow are not simple, and many factors influence the source, timing, and magnitude of stream discharges.

Basically there are two methods which are available for separations of base flow from a hydrograph .These are graphical method and deterministic method.

In graphical method the simple way to make a base flow separation is to draw a straight line from the point of rise to an arbitrary point on the lower portion of the recession segment of the hydrograph. This arbitrary point may be so chosen that the base flow separation line should not be too long and, on the other hand, the base flow should not rise too high. This point may also be chosen as the beginning point of the fitted part of a ground water recession curve being matched to the recession segment. According to Linsley, Kohler, and Paulhus, the point may be approximately taken as $A^{0.2}$ days after the time of peak flow, where A is the drainage area in square miles. Another simple procedure is to extend the approach segment forward to a point directly below, or a little beyond, the peak, and then to connect this point to the arbitrary point on the recession segment with a straight line. But doing so for the whole hydrograph is tiresome and needs drawings of line to each hydrographs also it is not possible directly calculate the base flow, hence for the calculation of base flow you have to adopt an area calculation methods.

To make the base-flow separation process less tedious and more consistent deterministic procedure was proposed in 1980 by the British Institute of Hydrology called the standard method or institute of hydrology method.

The method combines a local minimums analysis with a recession-slope test ,with this method we can estimates the annual base-flow volume of rivers and streams and computes an annual base-flow index (BFI, the ratio of base flow to total flow volume for a given year) for multiple years of data at one or more gage sites. (Swan, 1983) .So this method is adopted in this study.

2.7.3 Standard Method of Base Flow Separation Method

The standard method called the Institute of Hydrology Method begins by partitioning the year into N-day periods and determining the minimum flow within each period. These minimum flows are the potential turning points on the base-flow hydrograph. (If the year is not evenly divisible by N, any remaining days will be included in the last period of the year). The default value for N is 5 days. When several days within one period are tied for the lowest flow, the earliest day will be considered the minimum, except during the last N-day period of each year, when the minimum will be considered to occur on the latest day.

To determine the turning points on the base-flow hydrograph, the collection of N-day minimum flows is processed using a turning-point test.

Institute of Hydrology (Standard Method) turning-point test:

Given three adjacent N-day minimum flows, Q_0, Q_1, and Q_2:

Q_1 is a turning point IF:

$$Q_1 * f <= Q_0 \quad \text{and} \quad Q_1 * f <= Q_2$$

Where f is a turning-point test factor that must be greater than 0 and should be less than 1.

2.8 Low flow estimation of un-gauged catchments

Planning and designing hydraulic structures need adequate hydrological information of the specific area or the region at large. This includes observed stream flows at many sites. But many sites do not have adequate number of gauging station or that are recently established. There may not be gauging stations in the catchments at all. In such cases, transfer of required information from gauged sites to un-gauged sites becomes very essential.

Generated low flow in catchments, among other things, are dependent on the catchments characteristics. If an interest site has no record of flow, then a low flow, such as the seven day mean annual minimum flow, $\overline{Q}_7$, can be estimated from a relation between Q_7 and measurable catchments characteristics, which has been calibrated from gauged catchments in the region.

The physical characteristics of catchments may be grouped under the following headings (Cunnane, 1985):

- Size and shape
- Density and distribution system
- Overland and channel slope
- Base flow Index
- Soils and geology
- Rainfall and climate

The selection of catchments characteristics depends on judgment of the likely predictive success of variables, based on cause and effect or input and output relationship. The catchments physiographic characteristics must be simple enough so that they can be measured from maps easily and their accuracy should not be highly dependent on the scale of the map (Sine. 2004).

A number of measurable indices can be used to represent these characteristics. The most commonly used indices are Area, Rainfall, slope, and base flow index. The correlation between each and combinations of them with the index low flow will be established and then each correlation is evaluated using any of the available parameters.

In the case of multiple regressions, the degree of correlation of a dependent variable to many independent variables is measured by any of the following parameters: the standard deviation of residual, multiple correlation coefficients, and the coefficient of multiple determinations, the partial correlation coefficients, and the beta coefficients (Yevjevich, 1972).

The first step in regression analysis is the choice of the form of the predictive model. Most previous studies have used regressions of the form:

$$\overline{Q} = aA^{b} B^{c} C^{d} \qquad \text{…2.32}$$

Where $\overline{Q}$ = mean annual low flow

A = basin area characteristics

B, C… = catchments characteristics

a, b , c , d …. = regression parameters to be estimated.

The model indicated above is not linear but to make the treatment easier logarithmic transformation is applied to get a linear equation (Chow, 1964). Least square techniques are applied to determine the parameters and then coefficient of determination is used to choose significant parameters.

2.9 Previous Study

Little or no research has been done on the same topic of low flow analysis in the Dedessa river basin; rather two authors (Gebeyehu.A.1989 and Sine.A.2004) have done Regional Flood Frequency Analysis that can have a relation with this research paper in some degree. (Gebeyehu.A.1989) has done Regional frequency analysis of the North –western highlands of Ethiopian's basin. (Such as Abbay basin, part of Gibe-Omo and Awash basins together).

For regionalizing and delineation of the basin into homogeneous region (Gebeyehu.A.1989) adopted geographical location and climatic similarity. But (Sine.A.2004) used statistics of the observed flood data the one which is recommended in recent days.

The other research was done by Tegenu zerfu (M.sc. thesis) with a title of Low flow frequency analysis for Abbay River basin. The main objective of the study was to delineate the Abbay river basin into hydrological homogeneous regions which would form the basic units to form and develop frequency curves for each region. Low flow frequency analysis was done for the Abbay river basin using 7-day annual minimum flow series and six homogeneous regions were delineated using a simple test based on the variability of at-site values of Cv. In this research statistics derived from observed low flow data's from different stations are used for regionalizing and delineation of the basin into homogeneous region.

In addition to the above papers Abbay River Basin Integrated Development Master Plan Project was done by BCEOM covering major parameters i.e. hydrology, water resources, natural resources, land resources, demography, socio-economy and infrastructure in the basin.

3 DATA ANALYSIS FOR DEDESSA RIVER BASIN

3.1 Source and availability of data

To achieve the goal of the research various data have been collected from different agencies and individuals, which include time series data, topographical data and digitize map of the study area.

3.1.1 Data used

Flow data

The Dedessa river basin has eighteen (18) hydrological gauging stations. Out of which only the 12 stations data are available and collected from Minister of Water, Irrigation and Energy (MoWIE). The stations are as summarized in table 3.1 and only Eleven (11) gauging stations with a minimum of 10 years record length and above were used for the analysis and those stations have below 10 years record length rejected. Details of the stations are attached in the annex part (see annex A. Table 10.1)

The stations consist of daily flow series. They were used to produce different duration of low flows, flow duration curves, base flow, base flow index and also used for low flow frequency analysis for the whole region.

Table 3.1 Hydrological gauging stations within the basin

NO	STATION NUMBER	NAME OF STATION	BEGINNING YEAR OF RECORD	LAST YEAR OF RECORD	LENGTH OF RECORD
1	114001	Dedessa @ Arjo	1980	2003	24
2	114002	Anger @ Nekemte	1994	2004	10
3	114003	Sifa @ Nekemte	1962	1967	6
4	114005	Dabana @ Abasina	1962	1984	23
5	114007	Anger @ Gute	1982	2004	23
6	114008	Yebu @ Yebu	1979	2005	28

7	114009	Urgessa @ Gembe	1980	2003	24
8	114010	Tato @ Gutie	1996	2009	14
9	114012	Indires @ Sire	1981	2006	26
10	114013	Dabana @ Bunobe	1984	2005	22
11	114019	Temssa @ Agaro	1989	2006	18
12	101006	Uke @ Uke	1980	2005	26

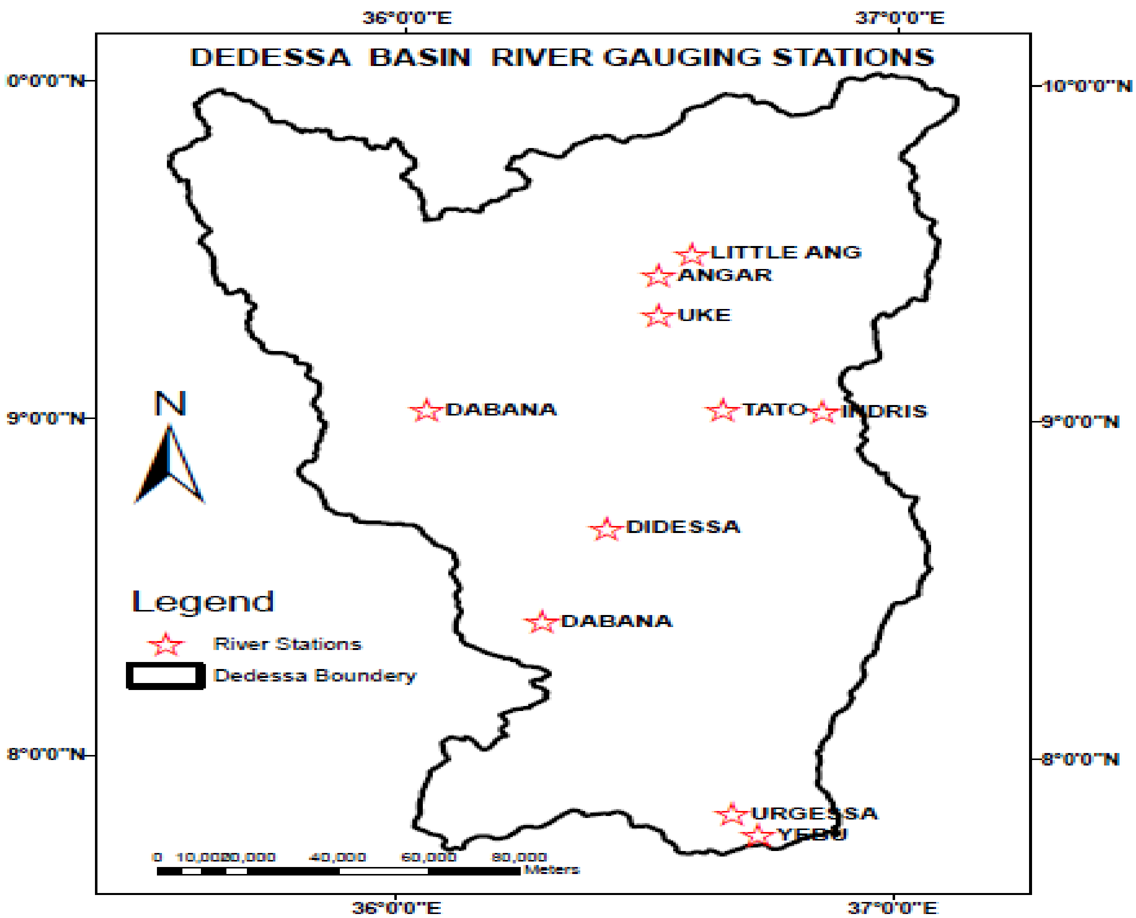


Figure 3:1 Hydrological Gauging Stations within the Basin

Table 3.2 Range of missing data in % from the record length

Range of missing data in %	No of stations
0-2	6
2-4	3
4- 6	3
>6	1

Rainfall data

Average annual and average monthly rainfalls of the basin were provided by the National Metrological Service (NMS) stations. These data were used to calculate catchment's average annual and monthly time series of rainfall in turn the output is used for regression model.

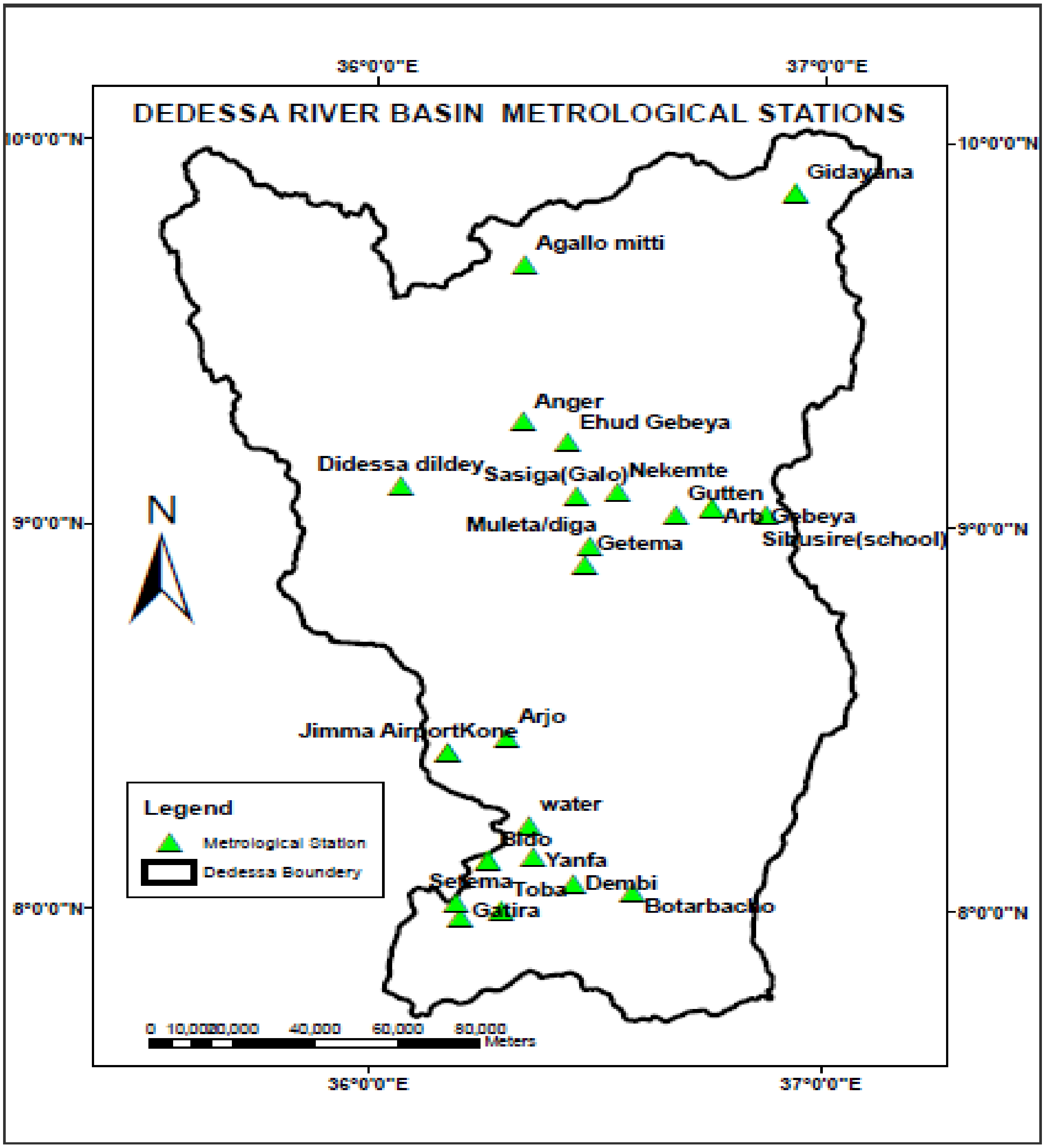


Figure 3:2 Metrological Stations within the Basin

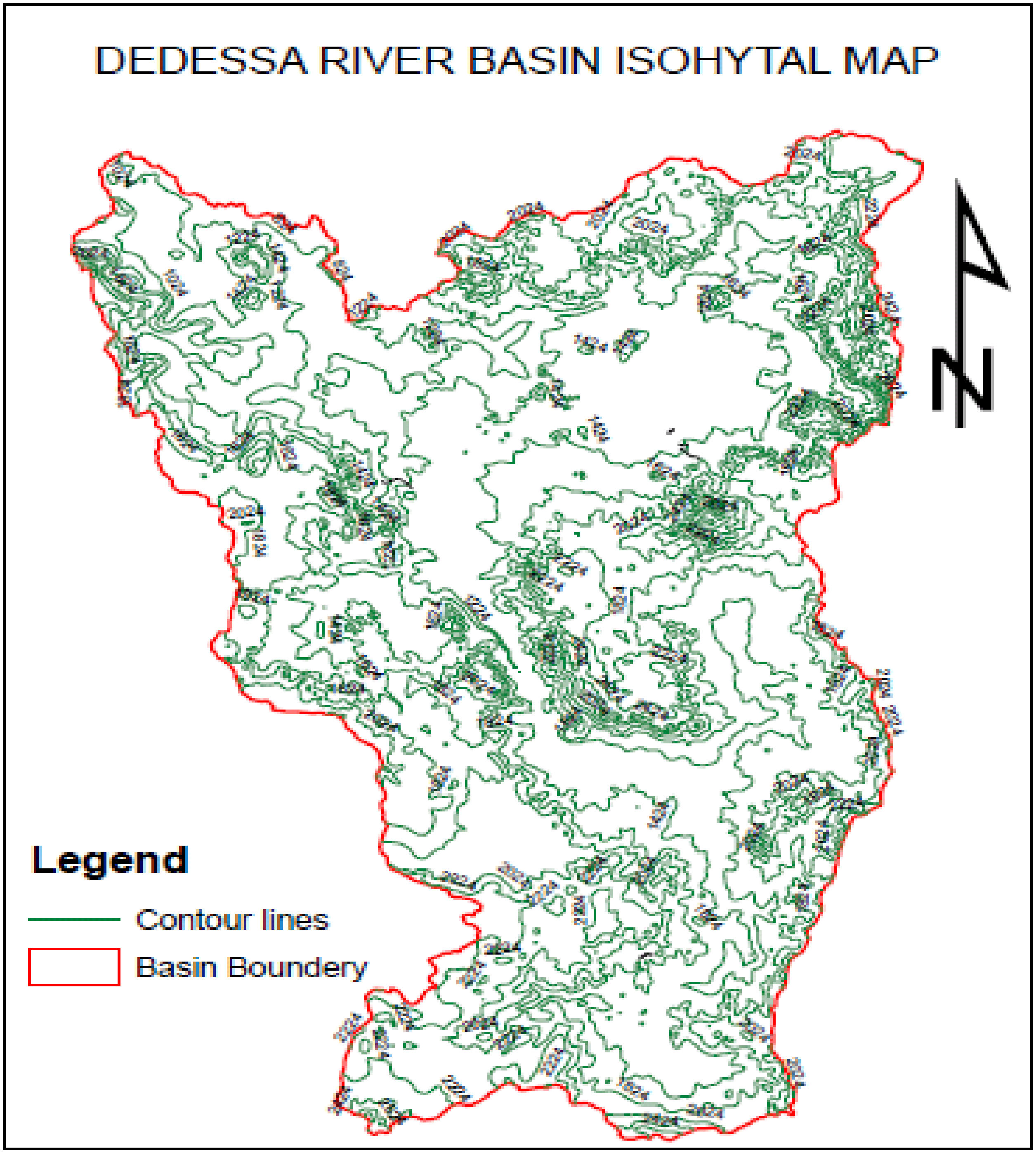


Figure 3:3 Isohytal Map of Basin

Topographic map

Topographic map of the basin with a scale of 1:250,000 is collected from Ethiopian Mapping Agency to be used as basic input for regionalizing the basin in to homogeneous region and also used to identify the correct name of the rivers within the basin.

Digitized map

Digital Elevation model DEM of the basin with a scale of 90x90m is collected from Ministry of Water, Irrigation and Energy of GIS Department. This data is a basic input for developing digital elevation model (DEM), flow accumulation, flow direction and also used in specifying the exact location of gauging station .All of which are necessary for regionalizing the basin in to homogeneous region.

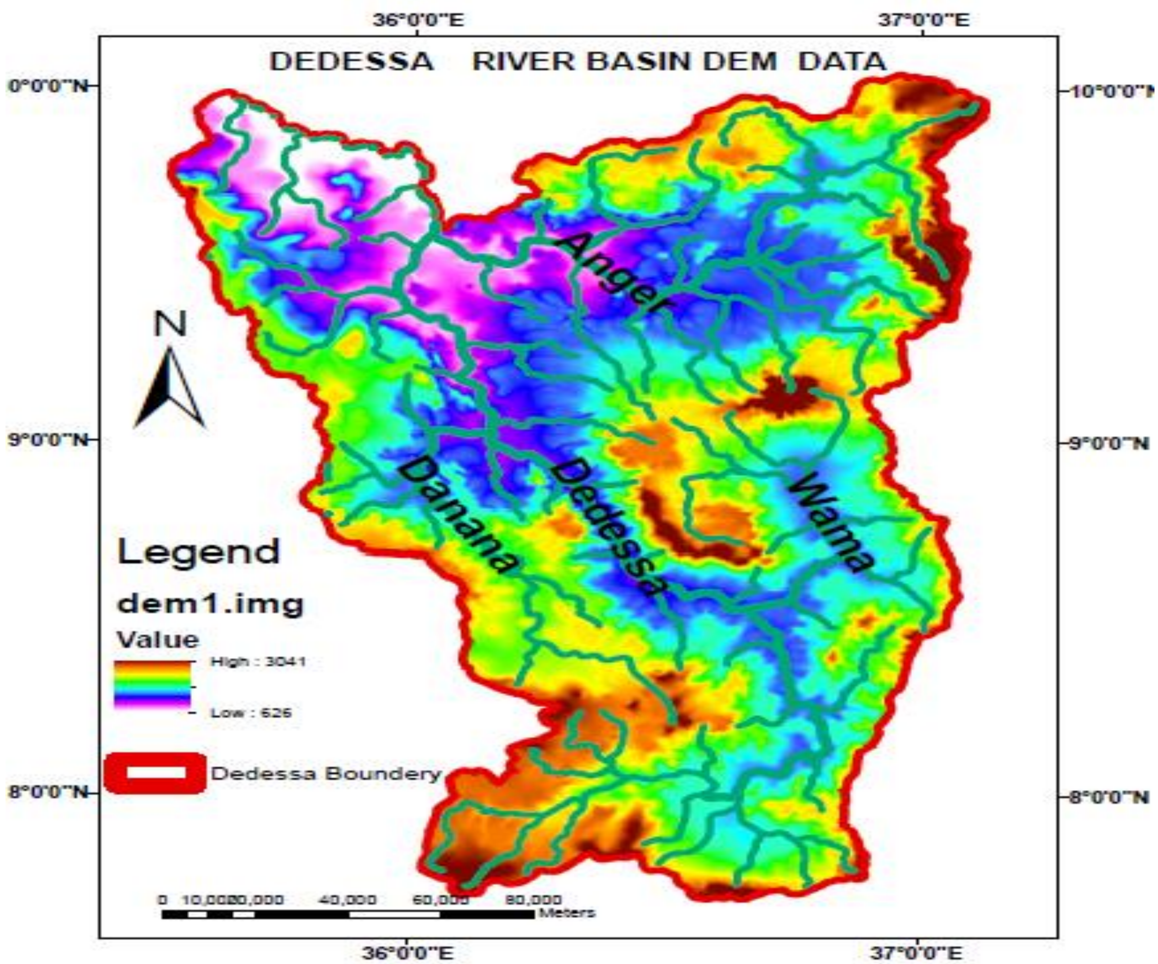


Figure 3:4 DEM Data of the Basin

3.2 Data Screening

3.2.1 Annual minimum series

The calendar year is used in defining an annual minimum value. Only stations without zero flow in their series are analyzed. The7day annual minimum flow series are used in frequency analysis, as it is usually necessary to nullify the effects of minor river regulations.

In abstracting annual minimum series from daily data, missing data was filled by using the following procedures (Teganu Zerfu (2009) low flow analysis and regionalization of Blue Neli River Basin, M.sc. thesis report Arba Minch University.)

(a) Missing periods of up to 15 days during the dry season were completed by interpolation.

(b) Missing data of up to 30 consecutive days in the dry & wet season were completed by seasonal mean value.

(c) Any calendar year containing periods of missing data and did not fall within any of the above categories was rejected for further analysis. A typical filled station graph is displayed in fig 3.1.

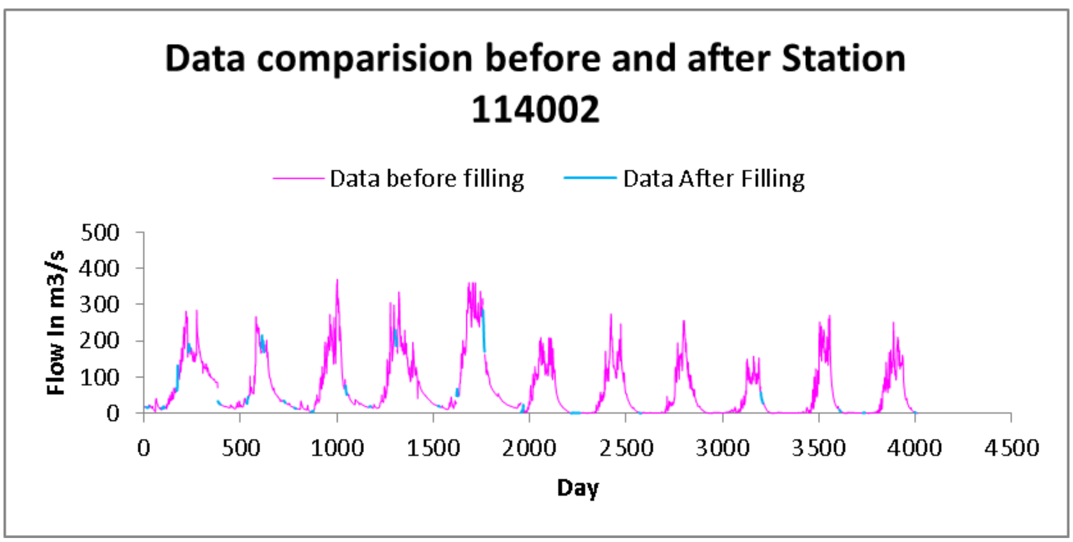


Figure 3:3 Data comparison before and after filling

3.2.2 Test for independency and stationery

One basic assumption in the use of statistical distribution in hydrological frequency studies is that the sample data are random variables without any serial correlation. As many hydrological sequences exhibit a departure from randomness. One of the means of measuring the extent to which a minimum in one year is dependent on the value of the previous is the lag one serial correlation coefficient denoted by r_1. It describes the strength of the relationship between a value in a series and that preceding it by one time interval. For a strict random series the value of r_1 must differ from zero only by sampling variation and for sequences showing strong persistence it is close to one.

So the following test is used to detect the randomness and the serial correlation of the data.

1. W-W test (Wald-wolfowitz test)
2. Lag –one serial correlation coefficient test

Out of the analyzed 11 stations one station were found to be dependent using both W-W test and serial correlation coefficient test. The rest of them were not found to be dependent using either of the randomness tests. The summarized result and criteria for both tests are show in table 3.3 and table 3.4 respectively.

Table 3.3 Result of data Independence & Stationarity using W-W test for Dedessa River basin

Station Name	Statistic mean	Critical Test Statistic	Remark
Dedessa	1.916	1.960	Independent
Anger	1.597	1.960	Independent
Dabana	0.728	1.960	Independent
Uke	1.743	1.960	Independent
Anger	1.538	1.960	Independent
Yebu	1.653	1.960	Independent
Urgessa	1.764	1.960	Independent
Tato	1. 659	1.960	Independent
Indires	0.771	1.960	Independent
Dabana	0.709	1.960	Independent
Temssa	2.942	1.960	dependent

Table 3.4 Result of data randomness using lag one serial Correction coefficient test for Didessa River basin

Station Name	L.1-Correlation Coefficient	Upper limit	Lower limit	Remark
Dedessa	0.31148	0.35623	-0.4437	Random
Anger	0.34338	0.48800	-0.6880	Random
Dabana	0.02900	0.36980	-0.4537	Random
Uke	0.25385	0.34408	-0.4241	Random
Anger	0.27023	0.36281	-0.4537	Random
Yebu	0.28090	0.41538	-0.3385	Random
Urgessa	0.19519	0.35623	-0.4432	Random
Tato	0.26640	0.44540	-0.5992	Random
Indires	0.07898	0.34408	-0.4241	Random
Dabana	0.07603	0.36980	-0.4650	Random
Temssa	0.60686	0.36281	-0.4537	Correlated

In any time series data outliers may or may not exist. These outliers may come due to personal error during recording and inadequacy of measuring devices or due to extreme condition of natural phenomena. Unless the source of the outliers is 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 annual minimum flows can be adequately modeled by a single distribution form (Cunnane, 1989).

So outliers test is not done in this study, but to avoid the effect of outliers an efficient method of parameter estimation like PWM was used. Even if outliers are retained in analysis, have only small effect if an efficient method of parameter estimation ML or PWM is used (Cunnane, 1989).

4 LOW FLOW ANALYSIS

4.1 Flow Duration Curves

4.1.1 Characteristics of flow duration curves in the low flow regimes

In this study flow duration curves of 1, 7, 10, 30 day have been derived for 10 stations. From the flow duration curves starting from the 80th percentile one day discharge, Q_{80} (1) is used to generalize catchments .None of the stations have Q_{80} (1) equal to zero. Based on the characteristics shown in the right tailed end portion of flow duration curves the basin flow duration curves are characterized in to two homogeneous classes. This is done by visual inspection. The whole and grouped flow duration curves are shown below in fig 4.1, fig 4.2 and fig 4.3

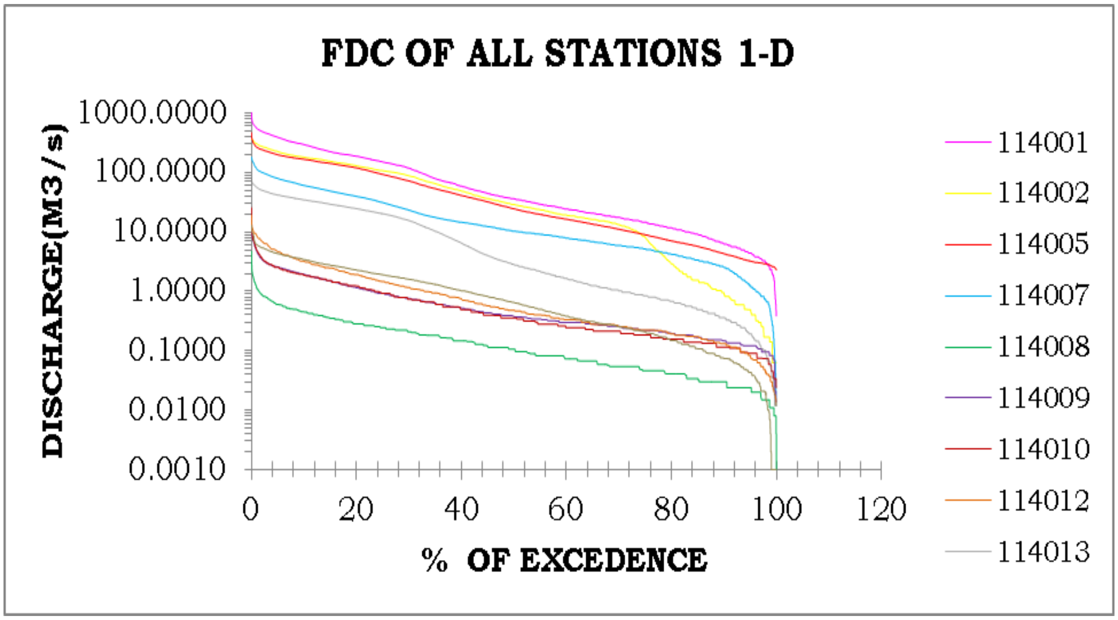


Figure 4:1 Flow Duration Curves of the Whole Stations

Four stations are grouped in group one FDC's, station 114005,114001, 114002 and 114007. The stations do not lie in the same region that is categorized in section 5.3.1 and 5.3.2 up on their low flow statistics. Two stations namely 114001 &114005 lie in region two.

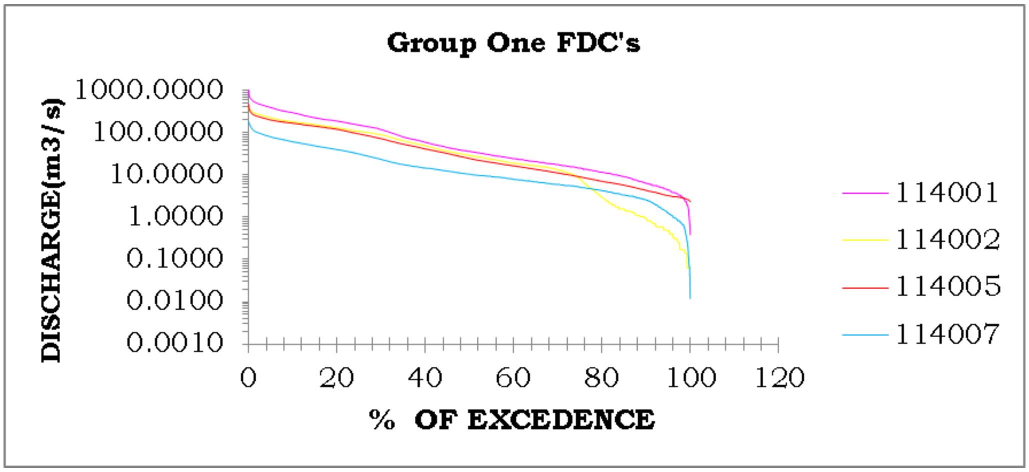


Figure 4:2 Group one FDC's

In group two FDC's six stations are available, station 114010, 114009, 114008, 101006,114012 and 114013. Based on the regions formed on their low flow statistics station 101006 lie in region One.

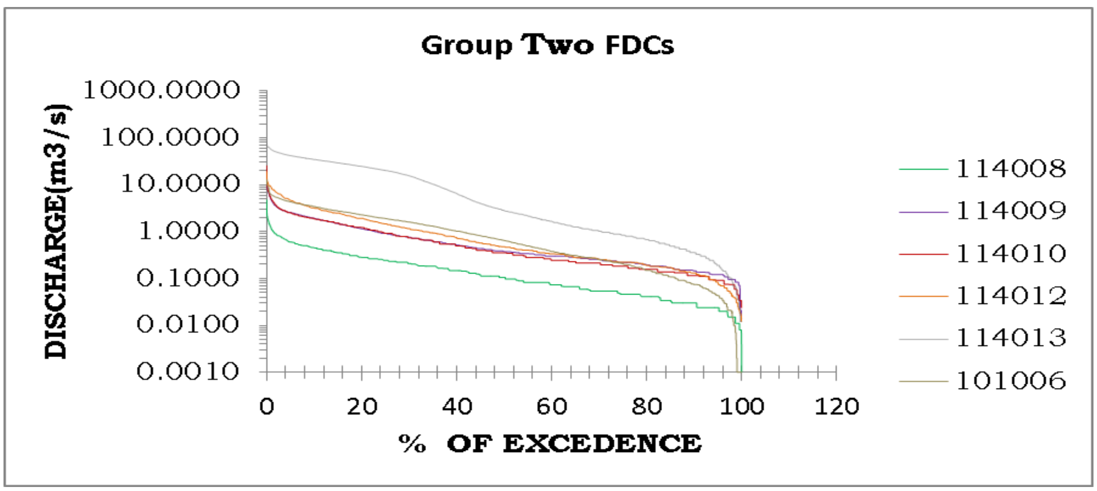


Figure 4:3 Group two FDC's

Note: - for this thesis grouping of station using flow duration curves simple for establishment of different durations and represents a measure of hydrological response which embraces the full regime from flood flows to low flows for each station when, comparing with grouping stations using statistical method (it show only the low flow regime) & due to this grouping of station using FDC's done.

4.2 Base Flow Separation and BFI

4.2.1 Base Flow Separation

In order to determine the contribution from overland flow in a watershed to the streams in the watershed or the contribution of flow from the stored sources it is necessary to separate out the base flow from stream gage data. The standard method called the Institute of Hydrology Method is used for separation of base flow from main hydrograph.

Adopting the prescribed method makes the base-flow separation process less tedious and more consistent.

For all the ten stations base flow hydrograph is separated and they are attached in annex E.For illustration typical separated base flow hydrograph is presented in Fig 4.6, 4.7 & 4.8.respectively.

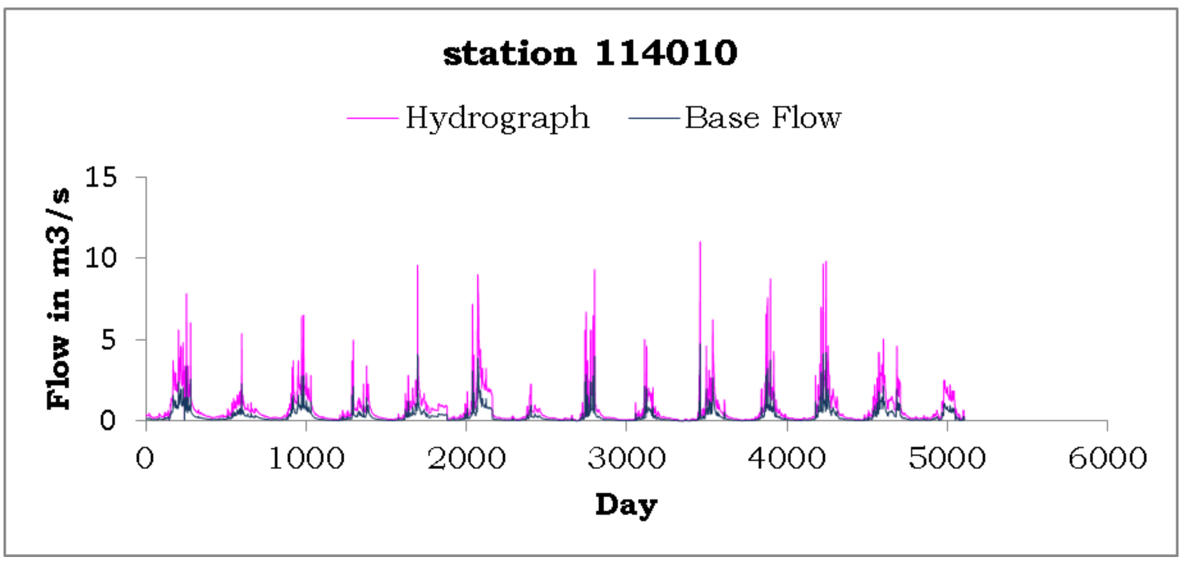


Figure 4:4 Typical Hydrograph and Base flow hydrograph

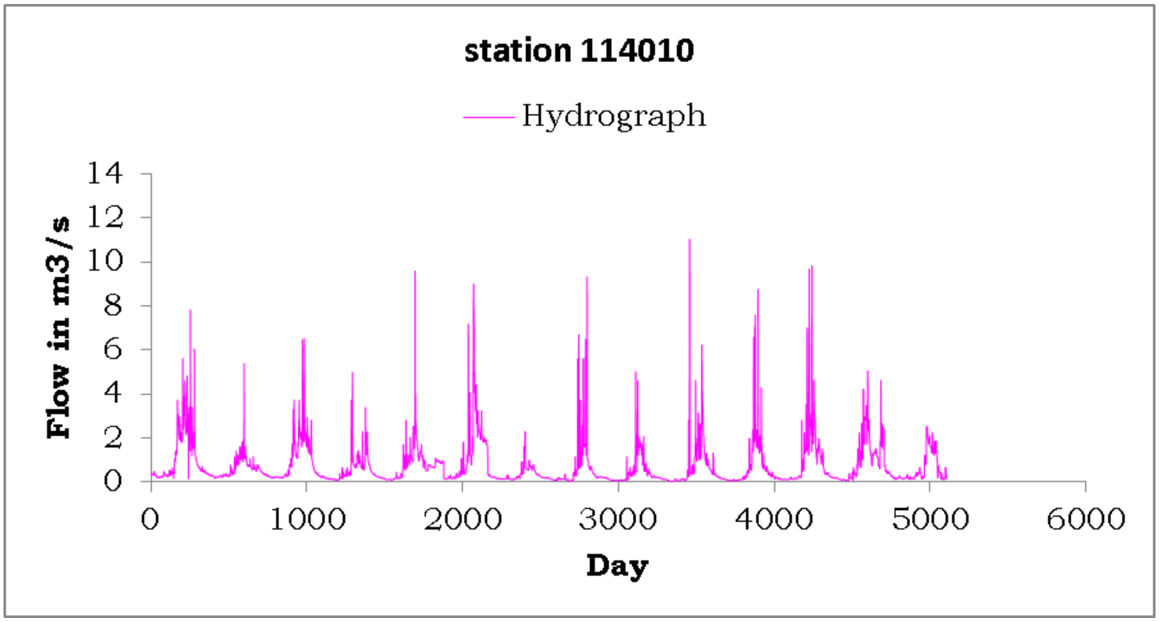


Figure 4:5 Typical Hydrograph

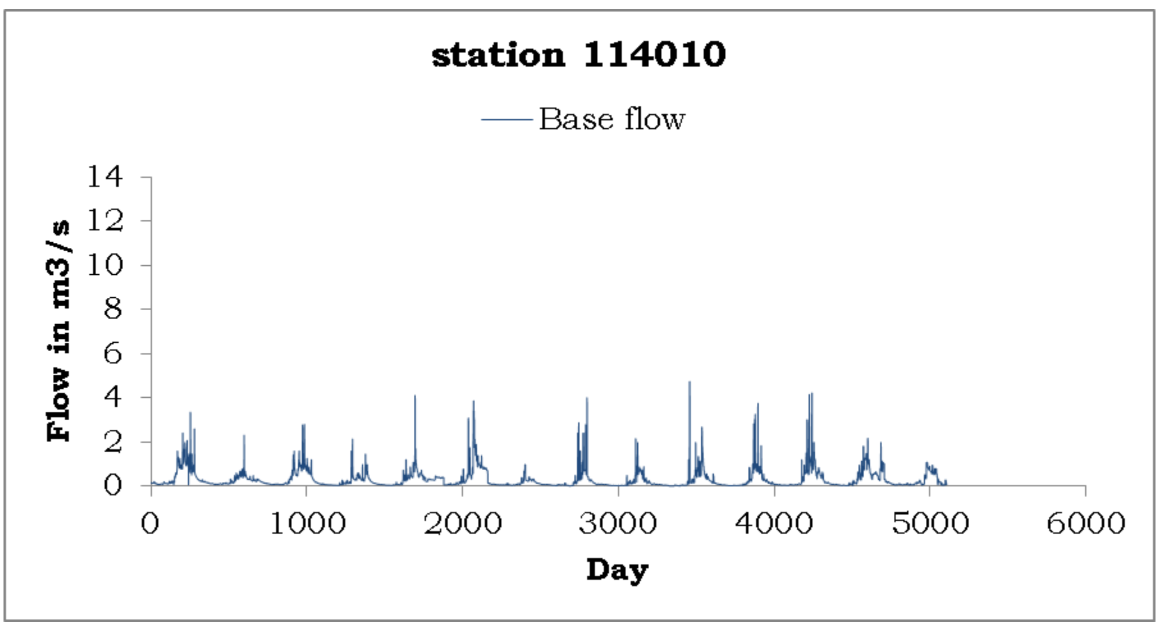


Figure 4:6 Separated Base Flow Hydrograph

4.2.2 Base Flow Index

The base-flow index is the total volume of base flow divided by the total volume of runoff for a period (Wahl, 1995). The base flow component of river flow is commonly expressed as a proportion of the total river flow, termed the Base Flow Index (BFI). A catchment's with BFI approaching 1 is highly dominated by base flow, while a catchments with BFI approaching 0 receives little base flow contribution.

From the main hydrograph, base flow hydrograph and base flow index is calculated. The obtained result that is BFI is used as one of the input parameter for the regression analysis. Grouped of stations within a range of BFI is shown in table 4.1.

Table 4.1 Stations Grouped based on the ranges of their BFI

Range of BFI	Station Number	Base flow index
	114008	0.234
	114009	0.262
	114010	0.346
	114012	0.247
	114013	0.350
0-0.35	101006	0.127
	114001	0.430
	114007	0.514
0.4-0.65	114002	0.641
>0.7	114005	0.725

4.3 Conclusive remarks

Those stations grouped in group one FDC's have a BFI of 0.43-0.725 which means 43% to 72.5% of the flow contributes from delayed storage and the flow mostly dominated by base flow .FDC's grouped in group two have a BFI of 0-0.35 which means the flow is mostly dominated by surface runoff.

5 RESULT OF REGIONALIZATION

5.1 Introduction

Before regionalization is done conventional moments, L-moments and L-moment ratios have to be computed. The computed values of conventional moments, L-moments and L-moment ratios are given on Appendix B, Tables B.1 and B.2.

For the computation of the above parameters using L-moments a Matlab routine program is used .These parameters are, later on are important and basic parameters for regionalization, selection of distribution and various tests. Detail computed statistical values are given on Appendix B, Tables B.1 and B.2.

On the basis of regionalization procedures shown in section 5.3 identification of homogeneous region and tests for their homogeneity is done on the following section.

5.2 Identification regions

The Lcs vs. Lck of standardized flow of the stations has been plotted with L-moment ratio diagram (LMRD) of various distribution function see fig 5.1 and stations which lie in the same distribution are preliminary grouped as homogeneous region. This is done on the basis that station from the same homogenous region will be distributed along the same distribution of LMRD (Dalrymple, 1990, URL).

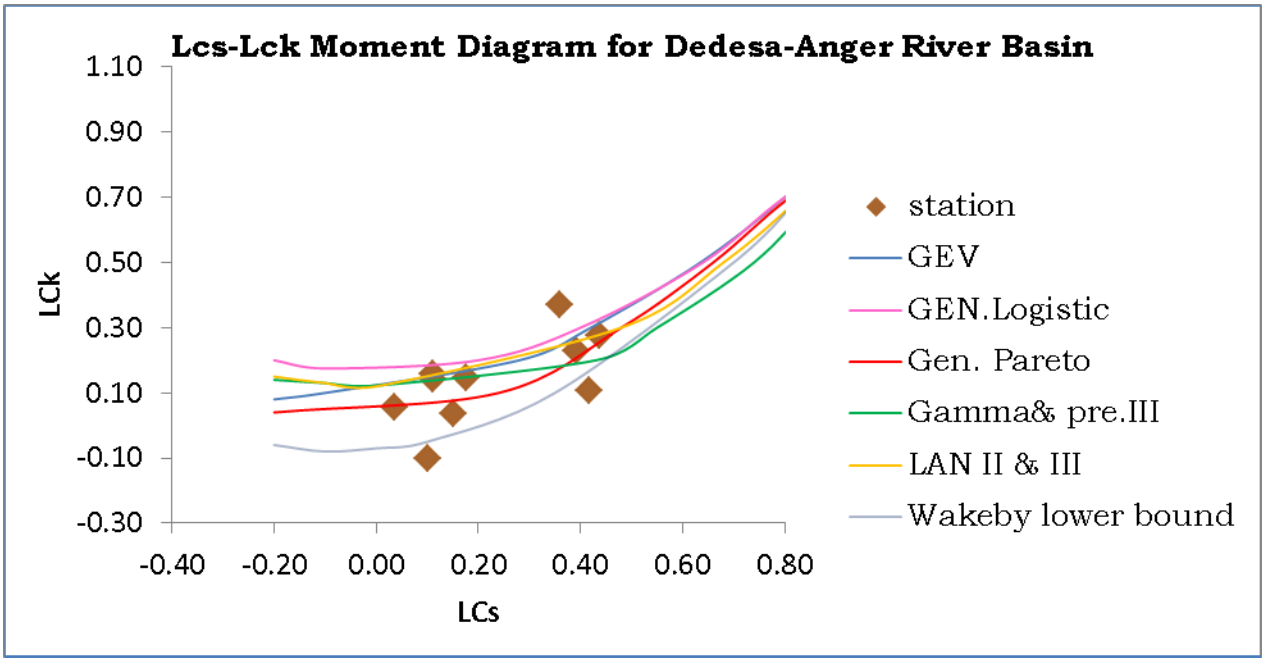


Figure 5:1 Lcs vs. Lck moment Ratio Diagram for 10 Standardized Station

5.3 Homogeneity test of region

5.3.1 CV and LCv homogeneity test

For the Cv and Lcv homogeneous tests a Matlab routine program used .As to the result obtained from the program all stations grouped preliminary as homogeneous region satisfies the Cv and Lcv homogeneity test criteria.

Table 5.1 Results of CV based homogeneity test for Dedessa river basin

Region name	cc value		Conclusion
	Conv.cv. based method	L-moment Cv-based method	
Region one	0.275	0.267	Homogeneous *
Region two	0.166	0.102	Homogeneous

***Note**: - The region declared to be homogenous if CC < 0.30 according to homogeneity test procedure stated in section 2.6.2.1

5.3.2 Discordance Measure Test

All of stations grouped preliminary as homogeneous region in each region satisfies the discordance test criteria which are stated in literature review part and a suitable criterion to classify a station as discordant is that D_i should be greater than or equal to 3.

Table 5.2 Result of discordant measure test for the Dedessa river basin

REGION ONE			
S.No	Station Number	Discordant Measure (Di) > or = 3	Remark
1	114002	0.370	Homogeneous
2	114007	0.442	Homogeneous
3	101006	0.614	Homogeneous
REGION TWO			
S.No	Station Number	Discordant Measure (Di) > or 3	Remark
1	114001	0.160	Homogeneous
2	114005	0.573	Homogeneous
3	114013	0.355	Homogeneous
4	114008	0.061	Homogeneous
5	114009	0.100	Homogeneous
6	114010	0.100	Homogeneous
7	114012	0.219	Homogeneous

.

5.4 Delineation of the region

The tool used in delineation of homogenous regions where GIS (Arc-view 9.2) software. On the digitized map of the basins, all stations under analysis were identified according to their geographical location (latitude and longitude). For each stations the statistical values (LC_s , LC_k) which were computed in section 5.2 where given. In addition, the preliminary identification of the region using LMRD mentioned above where identified on separate field to geo-code with different color. It is assumed that (LC_S, LC_K) values of one station varies linearly with (LC_S, LC_K) values of the neighboring stations. So from GIS screen, the distance between one station and its neighboring station was determined and (LC_S, LC_K) values where interpolated to fix the boundary between two stations of different regions. The procedures followed in the delineation of the boundary of the region were as follows:

I. To compute the (LC_s, LC_k) value of each station
II. To identify the location of stations along the distributions of LMRD
III. To identify the group based on step II
IV. To interpolate between LC_s and LC_k Values of two stations of different groups to fix two boundaries, one from the LC_s and the other from LCk values.
V. The boundary of the region is fixed between the midway of the two boundaries found in step IV.

According to the result obtained above the basin is regionalized in to two homogeneous regions

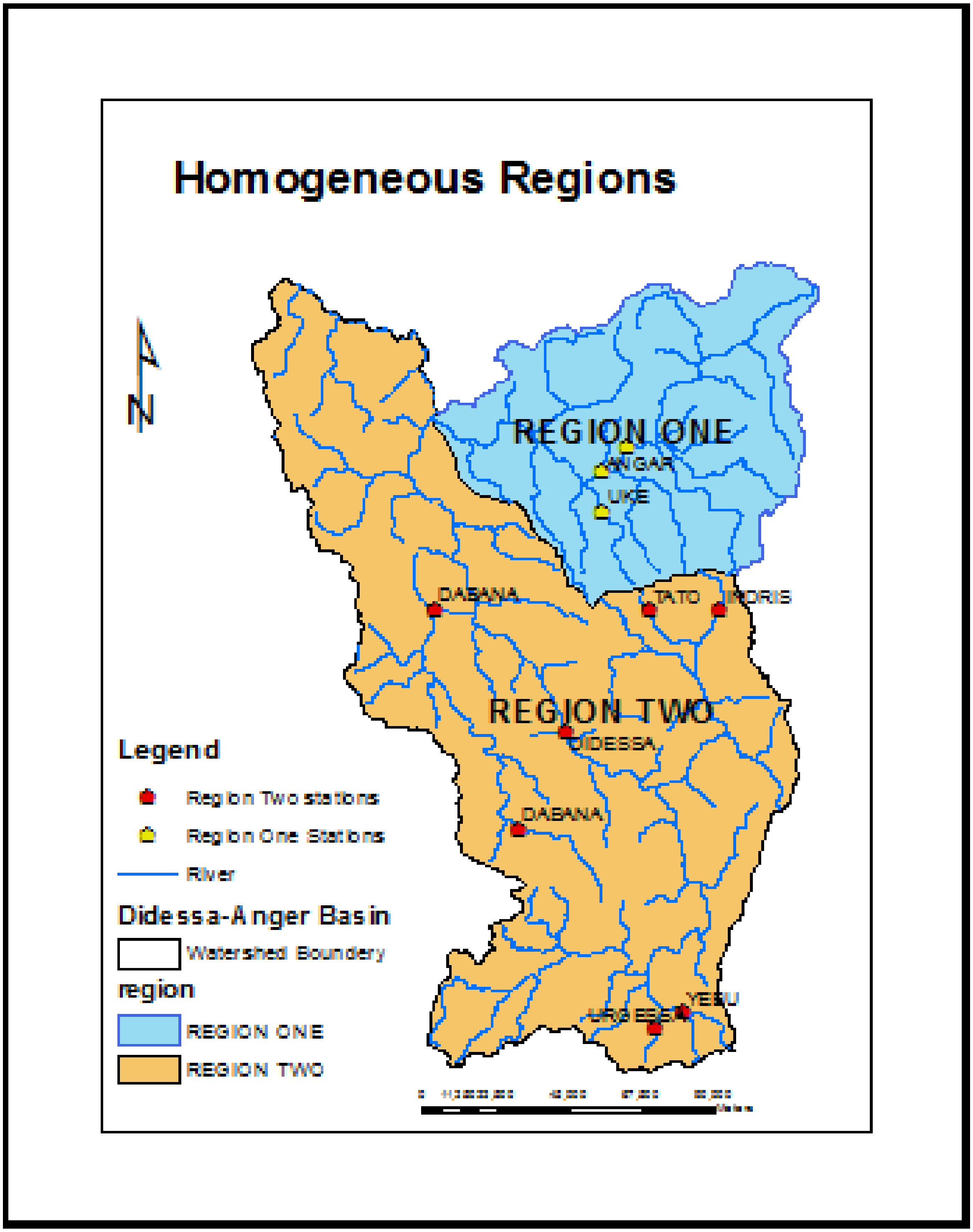


Figure 5:2 Map of the Established Homogeneous Region of Dedessa River Basin

Name of the region	**General characteristics**
Region One	Area 7,901.5km^2, Elevation ranges from 845 to 3000m, Rain fall varies from 1400 to 1800mm, slope ranges 5 to 15% and in some areas less than 5% land use of the area is dominantly cultivated and moderately cultivated.
Region Two	Area 19,610 km^2, Elevation ranges from 626 to 3041m, rain fall from 1200 to 1800mm, slope of the area varies from 5 to 10 %, land use of the area wood land , and cultivated

Table 5.3 Summary of land features, catchment characteristics.

6 SELECTION AND EVALUUATION OF BEST FIT STATSTICAL PARENT DISTRIBUTION OF LOW FLOW

6.1 Statistical distributions

The distributions used in developing models for estimating low flow quintiles with a particular probability of exceedance or non exceedance and the mathematical expression of the statistical distribution is shown in appendix C.

6.2 Selection of distribution by LMRD

According to Hosking (1990) the L moment ratio diagrams are based on relationship between the L moment ratios. A diagram based on Lcs verses Lck, can be used to identify appropriate distributions. The best parent distribution is the one that the average value of the point (Lcs, Lck) of all stations with in the region gets close to one of the drawn LMRD of the parent distribution. The result obtained is as shown on the graph below.

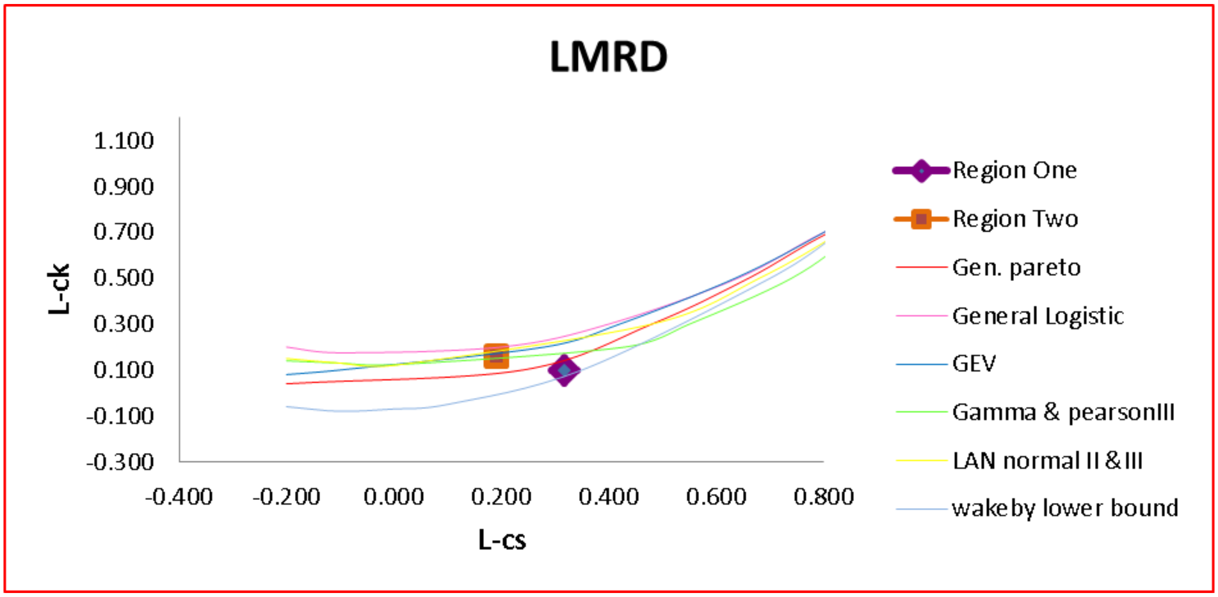


Figure 6:1 Regional Average of LMRD for the Regions of DRB

Accordingly, the most possible underlying candidate distributions are summarized on the table below.

Table 6.1Selected candidate distributions for the regions in DARB

Region Name	Average regional L-moments (t,t_3,t_4)	Selected Candidate Distribution
Region one	(0.243,0.319,0.094)	Wakeby Lower boundary
Region two	(0.267,0.190,0.162)	GEV /lognormal

So in the following section test for acceptability of the candidate distribution is done.

6.3 Selection of frequency distribution

Several statistical distributions are available for use to date but the problem arises in choosing the right one to fit the observed annual minimum flow data. The result of fitting a distribution also depends on the method of parameter estimation. There is no distribution that is universally accepted to fit low flows (Cunnane 1986) .The three parameter estimation methods used for evaluation are

Method of moment (MOM)

Method of maximum likelihood (ML)

Method of probability weighted moment (PWM)

For the candidate distributions which are chosen in the previous section the following method of estimation procedure were applied these are:

GEV (MOM, ML, PWM)

L Normal (MOM, ML, PWM)

Wake by (WAK4. WAK5) (PWM)

Table 6.2 Candidate distribution with adopted methods of distribution

Candidate distribution	Method of estimation		
	MOM	ML	PWM
GEV	*	*	*
L Normal	*	*	*
Wake by	-	-	*

For the Wakeby distribution, MOM and ML method is not used because moment estimates of parameters of wakeby distribution are not easily obtained.

6.4 Result for selection of parameter estimation method

1. Goodness of fit measure

The following results are obtained for goodness of fit

Table 6.3 Goodness of fit measure for the candidate distribution for each region

Region name	Candidate distribution	Goodness of fit measure (Z^{dist})	Remark
Region one	Wakeby	0.518	Desirable
Region Two	Log normal	0.182	Desirable
	GEV	0.105	Desirable

Standard error of estimates

Here the standard errors of estimate of the candidate distribution for different return period were computed to select the best statistical distribution and method of parameter estimation procedures.

Approximate standard error of estimate as well as parameter variances and covariance may be estimated by simulation on this particular case it is left out (Roa , 2000). The result is shown in table 6.4.

Table 6.4 standard error of estimate for different procedures and return period

Return period(T)	Region One			Region Two					
	Wekabay			GEV			Lognormal		
	MOM	ML	PWM	MOM	ML	PWM	MOM	ML	PWM
2	-	-	0.042	0.489	0.652	0.348	0.465	0.451	0.587
5	-	-	0.122	0.598	0.749	0.357	0.591	0.529	0.597
10	-	-	0.157	0.678	0.759	0.475	0.686	0.597	0.687
15	-	-	0.346	0.779	0.812	0.478	0.691	0.673	0.759
25	-	-	0.459	0.798	0.859	0.675	0.796	0.795	0.761
50	-	-	0.589	0.878	0.910	0.759	0.854	0.852	0.819

Based up on the Goodness of fit and minimum standard error of estimates the Wakeby lower boundary is the best distribution to model the annual low flow series for region one. Although both log normal and GEV distribution satisfied the goodness of fit test criteria for region Two, the GEV distribution fits the lower part of the data better than the log Normal distribution. Hence, the GEV distribution is selected for modeling the annual low flow series of region Two in Dedessa river basin for both regions PWM selected as best parameter estimated method.

Table 6.5 selected distribution and their method of parameter estimation for the different regions

Region name	Distribution and method of parameter estimation
One	Wakeby/PWM
Two	GEV/PWM

6.5 Derivation of regional frequency curve for the homogeneous region

Based on the result obtained in the above section the following distribution are found to be the best procedures for describing the annual minimum low flow and predicting acceptable low flow estimates for the delineated region.

Table 6.6 recommended method and procedure method for the different region

Region name	Distribution and method of parameter estimation
One	Wakeby/PWM
Two	GEV/PWM

For developing the frequency curves of the two regions a computer program (excel) which could calculate the standardized flow, non standardized flow and the parameter of the selected distribution is used.

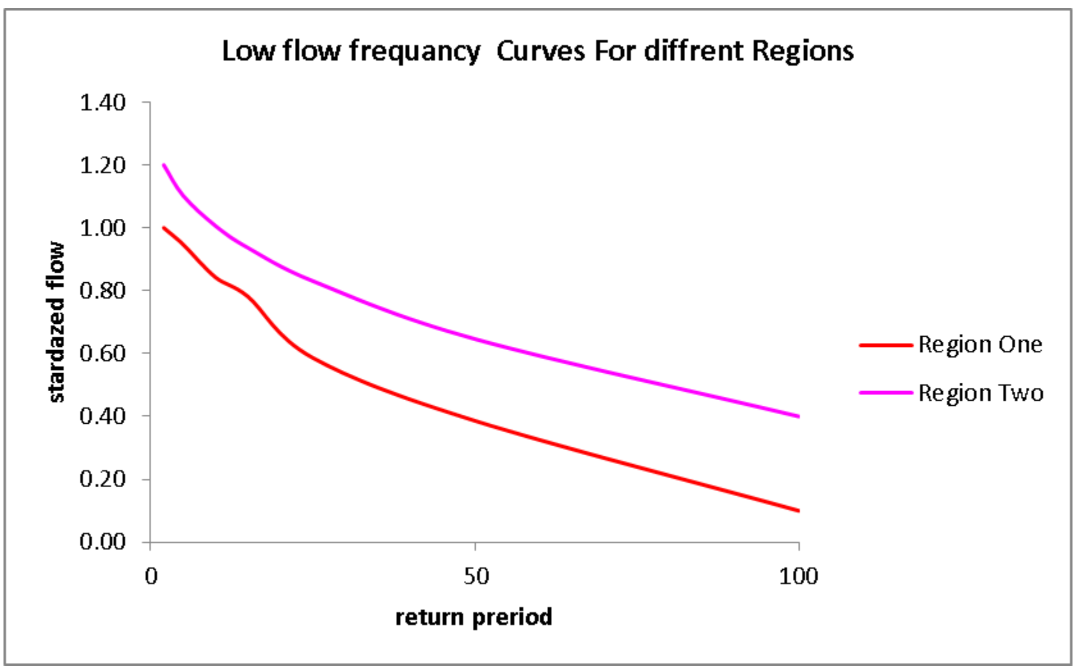


Figure 6:2 Developed Low Flow Frequency Curves of the Different Region

7 PREDICTION OF LOW FLOW FOR UN-GAUGED CATCHMENTS

7.1 Regression Model

A number of physical characteristics exist for a catchments and a number of numerical measures or indices have also been described in explaining and developing models that relate low flow and catchments characteristics. As was mentioned in chapter 2, there are a number of measurable physical characteristics of catchments that might have important relationship with the low flow. The availability of data representing them is very important in developing a model that relates index low flow and catchments characteristics is given by

$$\overline{Q} = c\ A^{a}\ S^{s}\ R^{r}\ BFI^{t} F^{f} \text{ ………………………………………………….7.1}$$

Where $\overline{Q}$ = index low flow, is the mean annual minimum flow.

A = the catchments area in square kilometers.

R = the mean annual rain fall in millimeters.

F = shape factor (L^2/A)

BFI = base flow index

S = Slope

And c, a, r, s, t and f are the regression parameters.

To make the analysis simpler, equation (7.1) is transformed to a linear form using logarithmic transformation

$$\text{Log } \overline{Q} = \log C + a \log A + s \log S + r \log R + t \log BFI + f \text{ Log } F \text{……. 7.2}$$

Then, the estimation will be performed on the parameters, namely, a, r, s, t and log c using multiple linear regression technique. Then the estimation was performed on the parameters.

An important consideration in multiple regression analysis is to check whether the data are adequately described or not. In this study the multiple coefficient of determination, R^2, is used as a measure of the ability of the regression model to describe variations in the dependent variable. The closer R^2 is to one, the better the regression model fits the data.

7.2 Result of regression model

Table 7.1 Derived regression equation for the different region in Dedessa river basin for prediction of mean annual low flow (MALF)

Table 7.1 Derived regression equation for the different region

Region name	Regression Equation	R^2
One	$Q = 6.98 A^{0.735} BFI^{5.699}$	0.997
Two	$Q = 0.00089 A^{0.9772} BFI^{0.9193} S^{-0.5534}$	0.840

Where A=area in Km^2

BFI =base flow index

S=slope in %

For all regions, the mean annual minimum flows are computed by the developed regression equation and compared with the actual annual minimum flows in the form of graph for each station of all regions. The graphs are shown in figure below.

The developed regressions equations have high coefficient of determination (almost 100% correlated) for region one. This may be due to unreliability or the number of the collected data. So, for realistic estimation of mean annual low flow, the input data has to be collected, processed and predicted properly. Inaccuracy is suspected in the input data particularly the areas of the catchment at the gauging stations. This needs collection of additional data and further analysis.

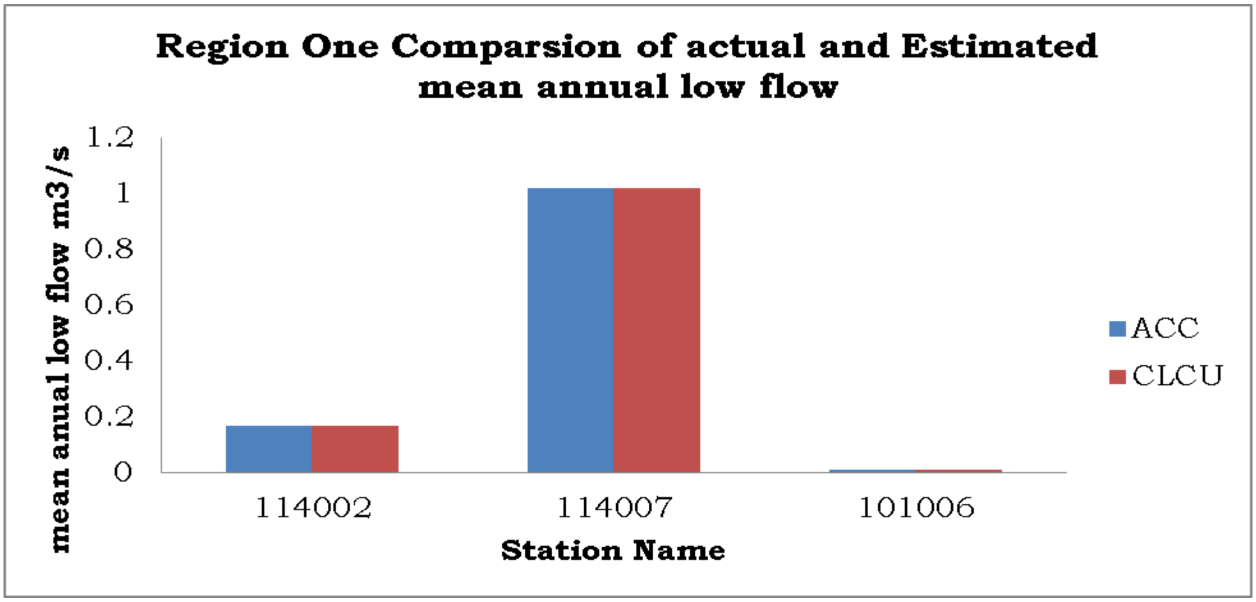


Figure 7:1 Comparison of Actual & Estimated Mean Annual low Flow for Region One

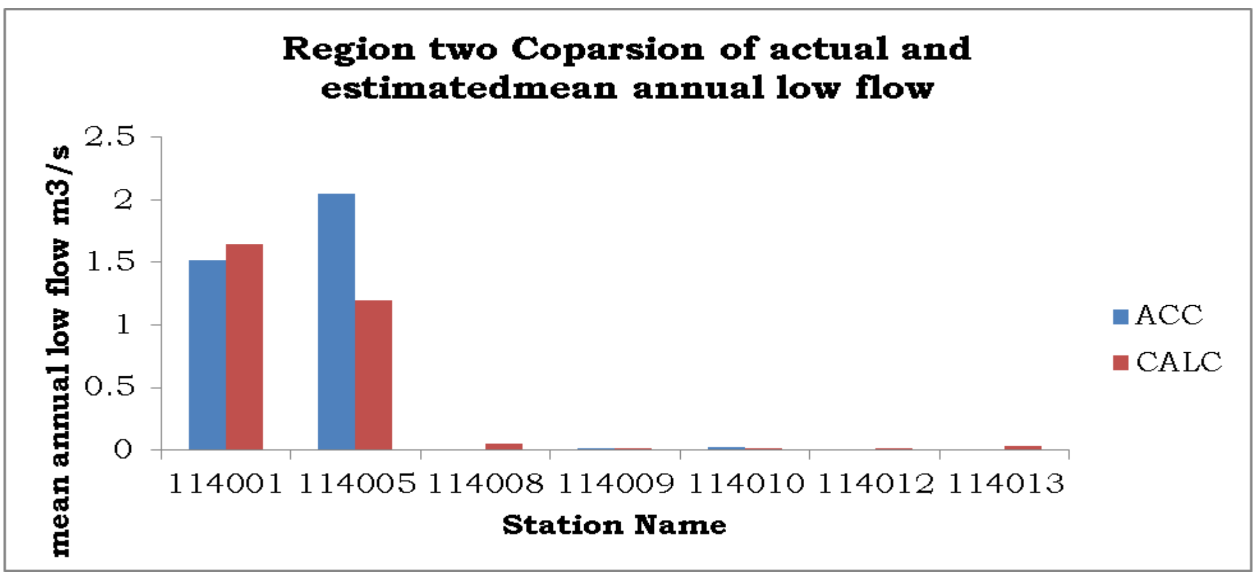


Figure 7:2 Comparison of Actual & Estimated Mean Annual low Flow for Region Two

8 SUMMARY, CONCLUSION AND RECOMMENDATION

8.1 Summary and conclusion

This study demonstrates the index low flow method of regional low flow frequency analysis. The main objective of the study was to delineate the Dedessa river basin into hydrological homogeneous regions which would form the basic units to form and develop frequency curves for each region.

The flow data which is collected for the analysis were checked for their consistency and missing data are filled by interpolation and seasonal mean value. The flow data was also tested for its independency and stationerity and those station which shows serial auto correlation or dependency behavior were discarded for further analysis.

Low flow frequency analysis was done for the Dedessa river basin using 7-day annual minimum flow series. Two homogeneous regions were delineated using a simple test based on the variability of at-site values of Cv. The delineated regions were tested for homogeneity using the Hosking and Wallis test. In assessing the suitability of a distribution selected for a particular region, the goodness of fit measure statistic (z) and minimum standard error of estimate were used. The Wakeby distribution provides a good fit to low flows in region one, while GEV distribution fits well in region Two. The method of probability weighted moments is considered as the best parameter estimation procedure compares with the method of moments and maximum likelihood.

Regression analysis has been applied to develop regression models to predict mean annual minimum flow from un-gauged catchments using catchments characteristics.

L- Moment ratio diagram provides a practical means to group stations in different region and also identify the underlying distribution for a given region.

So L- moment ratio diagram was used to group stations to same region and to identify the underlying statistical distribution.

Flow duration curves of different durations are established. The flow duration curve represents a measure of hydrological response which embraces the full regime from flood flows to low flows .The flow regime of DRB is thus described by studying flow duration characteristics and base flow contribution. Accordingly two groups of FDC's are established based on their pattern shown from the 80th percentile and the value greater than of it.

According to the results obtained from the three approaches used for low flow regionalization the author of this paper recommend that analysis and characterization of homogeneous region have to be done independently for each approach.

Regression analyses have been applied to develop regression models to predict mean annual low flow from ungauged catchment using catchment characteristics. The results of the analysis show that the area is the dominant factor, which is the common independent variable for all regions. During the analysis, for region one the coefficient of determination was not reliable when all variables of all station were correlated with the mean low flow at a time. In addition to this, the rainfall data is not correlated with the mean low flow satisfactorily for all regions. This is because the rainfall data, which was collected from NMA, is not representative of the whole basin. Therefore the rainfall data couldn't appear in the regression equation. In this regard further investigation is required by determining the appropriate areal rainfall distribution of the whole basin.

8.2 Recommendation

- Homogeneity test that consider not only higher moments of low flow but also some important basin characteristics should be investigated.
- Delineation of hydrological homogeneous regions based on statistical parameter of gauged sites could be one of the possible alternative methods of regionalization.
- Different photographic and climatic parameters like elevation, soil type, geology and hydrogeology indexes, etc of each region should be determined so that a better regression model will be resulted and the estimates of the mean annual, minimum flow will be accurately estimated.
- Low flow frequency analysis of stations which have zero flows have to be studied further.
- It is advisable to extend this approach of LFFA for other parts of Ethiopian river basins to establish the homogeneous regions of the country so that problems related of absence of

sufficient discharge data for water resources project planning and design could be reduced.

9 REFERENCE

Abebe A. (2003) "Ethiopian Journal of water Science and Technology"

Admas, J.C.,Brainerd,W.S.& Goldberg,C.H.(1992) Programmers Guide to Fortran 90.McGraw Hill, Singapore

A.Gustard et al (1989) Flow Regimes from Experimental and Network Data, volume 1 Hydrological Studies ,Institute of Hydrology ,Wallingford.

Baipai,A.C.,Calus,I.M.and Fairley,J.A.(1978). Statistical method for Engineers and Scientists

Brogan, L., Cunnane, C. (2006). Low flows and low flow distributions for Ireland. Irish National Hydrology Seminar (2006) pp. 85-92.

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

Cunnane, C. (1989) Statistical Distribution for Flood Frequency Analysis. World Metrological Organization Operational Hydrology Report No.33.

French Engineering Consultants and ISL Consulting Engineers (1990) Abbay River Integrated Development Master Plan Project, Main Report Volume 1.

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

G.L. Giese and M.A. Franklin,(1996) Water-Resources Investigations Report 96-4308 Magnitude and Frequency of Low Flows in the Suwannee River Water management District, Florida

Gubareva, T., van Nooijen, R., Kolechkina, A., et al., Extreme Flood Estimation by Different Probability Distribution Laws in Different Climatic Conditions, Geophysical Research Abstracts, 2008,

G. V. Loganathan et al (1985) Methods of Analyzing In stream Flows, Department of Civil Engineering Virginia Polytechnic Institute and State University Bulletin 148.

Haddad, K., Rahman, A. (2012), Regional flood frequency analysis in eastern Australia: Bayesian GLS regression-based methods within fixed region and ROI framework – Quantile Regression vs. Parameter Regression Technique, J. of Hyd. 430 – 43 1, 142-161.

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

Hosking, J. R. M., Wallis, J. R. (1997), Regional frequency analysis: an approach based on L-moments. Cambridge University Press, Cambridge, U.K.

Hosking,J.R.M.and Wallis,j.R.(1993). Some Statistics Useful in Regional Frequency Analysis .Water Resource Research, Volume 29,No.2. New York

Kozhevnikova, I.A., Distribution Parameter Estimation for Small Samples by *L*_moments,

Melesew,G.D.(1996). Regional Flood Frequency for Namibia and Zimbwabe, M.Sc thesis report. Dar e salaam.

Raymond J. (2001) Development of Low Flow Prediction Models for Southern Africa PHD, thesis report, Dar e salaam.

Sine, A. (2004) Regional Flood Frequency analysis, M.Sc. thesis report, Arba Minch University

Swan, D. R. and R. Condie. (1983) "Computation of the Base Flow Index," Water Resources Branch Inland Water Directorate, Environment Canada

S.H.Mkhandi,R.mngodo and Kachroo (1997) Analysis of Flow Regimes in Tanzania

USWRC,United State Water Resource Council, (1976) .guide lines for Determining Flood Frequency ,Bulletin No. 17 of the hydrology Committee, Washington, D.C.

Wu, Y.C., 2005. Establishing acceptance regions for goodness-of-fittest by stochastic simulation. Master Thesis. Department of Bio-environmental Systems Engineering,National Taiwan University (in Chinese).

10 APPENDICES

10.1 APPENDIX A

Table 10.1 Table Hydrological gagging station used for the study and their condition

Station _NO	catch_ area in km^2	no of year of data	mean flow m^3/S	Yield (Lt/Km^2)	LATITUDE	LONGITUDE	%_OF_MISSI	Period of recoded
114001	9981.0	24	100.58	10.08	8.683	36.417	2.59	1980-2003
114002	202	11	65.10	322.28	8.794	36.597	2.11	1994-2004
114003***		6						1962-1967
114005	2881.0	23	58.70	20.37			1.83	1962-1984
114007	3742.0	23	22.26	5.95	9.500	36.583	1.20	1982-2004
114008	47.0	28	0.18	3.94	7.800	36.700	1.15	1979-2005
114009	19.0	24	0.768	40.24	7.833	36.650	3.50	1980-2003
114010	42.0	14	0.750	17.86			3.40	1996-2009
114012	49.0	26	1.21	24.64	9.033	36.850	3.89	1981-2006
114013	47.0	22	11.353	241.56	8.400	36.283	3.97	1984-2005
114019**	47.5	14			9.317	36.517	1.59	1989-2006
101006	52.5	26	1.27	24.29			1.90	1980-2005

NOTE

*Stations not satisfy first condition data filling (stations having Zero flows)

** Stations failed to satisfy independency and stationerity

***Stations failed to satisfy minimum no of data

10.2 APPENDIX B

Table 10.2 Summary of site statistics by CV moments by CV moments Result sheet-1. LFFA-AS computer program

---

Mean flow and Conventional Moment ratios for

Dedessa river basin,

==

Station Name	Area km^2	N	Mean flow m^3/s	CV	Cs	Ck
114001	9981.0	24	3.118	0.310	1.052	1.299
114002	202.0	11	0.729	0.711	0.921	1.284
114005	2881.0	23	2.880	0.148	0.787	1.675
114007	3742.0	23	1.794	0.240	0.931	1.109
114008	47.0	28	0.019	0.100	0.236	1.122
114009	19.0	24	0.141	0.040	0.649	0.815
114010	42.0	14	0.101	0.030	0.144	0.645
114012	47.0	26	0.096	0.040	0.035	0.459
114013	47.5	22	0.160	0.210	0.399	0.987
101006	52.5	26	0.070	0.100	0.972	1.597

Table 10.3 Moments and L_ Moment ratios for Didessa river basin, Ethiopia; Result sheet-2 LFFA-AS computer programTable B.2 L-moments and L_ Moment ratios

Station Name	N	Unbiased t	t_3	t_4
114001	24	0.698	0.606	0.336
114002	11	0.651	0.400	-0.055
114005	23	0.307	0.198	0.178
114007	23	0.408	0.369	0.438
114008	28	0.326	0.348	0.317
114009	24	0.152	0.083	-0.086
114010	14	0.188	0.223	0.269
114012	26	0.255	0.038	0.013
114013	22	0.409	0.229	0.136
101006	26	0.624	0.482	0.134

10.3 APPENDIX C

Table 10.4 MATHEMATICAL EXPRESSIONS OF DISTRIBUTIONS

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

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 I Extreme value Type1 distribution (EV1)	F(x) = $\exp\left\{-e^{-\left(\frac{x-u}{\alpha}\right)}\right\}$	$-\infty < x < \infty$ $\alpha > 0$
Five parameters Wake by distribution (WAK 5)	x = m+ α [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) = $\left\{1+\left[1-\gamma\left(\frac{x-\alpha}{\beta}\right)\right]^{1/\gamma}\right\}^{-1}$, $\gamma \neq 0$ = $\left\{1+\exp\left[-\frac{x-\alpha}{\beta}\right]\right\}^{-1}$, $\gamma = 0$	γ <0, $\alpha+\frac{\beta}{\gamma} \le$ x $< \infty$ γ >0,- $\infty <$ x < $\alpha+\frac{\beta}{\gamma}$ -$\infty <$ x < ∞

10.4 APPENDIX D

Table 10.5 Result of Regression model for estimation of annual minimum flow from catchments characteristics

REGION ONE

MULTIPLE REGRESSION OF Z_2 ON 2 INDEPENDENT VARIABLES

CHECKLIST OF THE DATA USED HERE: Where Z1 =area in Km^2

	Z_1	Z_2	Z_3
1	3742.00	1.020	0.247
2	52.00	0.001	0.127
3	202.00	0.167	0.262

Z2 = flow in m^3/s

Z3 =base flow index

THE COVARIANCE MATRIX

Z_1	5.10		
Z_2	0.08	1.753	
Z_3	-0.20	0.29	0.08

THE CORRELATION MATRIX

Z_1	1.000		
Z_2	0.040	1.000	
Z_3	0.889	0.690	1.000

THE FITTED REGRESSION EQUATION:

$$Z_2 = 6.98\, Z_1^{0.735}\, Z_3^{5.6998}$$

THE MEASURES OF MULTIPLE CORRELATIVE ASSOCIATIONS:

THE MULTIPLE CORRELATION COEFFICIENTS: 0.951

CORRECTED FOR 1 DEGREES OF FREEDOM: 0.964

THE COEFFICIENT OF MULTIPLE DETERMINATIONS: 0.997

CORRECTED FOR 1 DEGREES OF FREEDOM: 0.893

THE STANDARDISED REGRESSION (BETA) COEFFICIENTS:

$B_2 = 6.980$
$B_1 = 0.735$
$B_3 = 5.6998$

REGION Two

MULTIPLE REGRESSION OF Z_2 ON 3 INDEPENDENT VARIABLES

CHECKLIST OF THE DATA USED HERE: Where z1=area Km2

	Z_1	Z_2	Z_3	Z_4	
					Z2= flow in m^3/s
1	9981	1.52	0.234	1.29	Z3=base flow index
2	2881	2.049	0.725	1.67	Z4= slope in %
3	47.0	0.001	0.350	6.09	
4	19.0	0.014	0.346	6.60	
5	42.0	0.024	0.43	1.07	
6	49.0	0.006	0.514	2.0	
7	47.0	0.012	0.641	2.46	

THE COVARIANCE MATRIX

Z_1	3.14			
Z_2	1.21	2.98		
Z_3	0.29	0.51	0.10	
Z_4	0.32	-0.35	0.02	0.25

THE CORRELATION MATRIX

Z_1	1.000			
Z_2	0.424	1.000		
Z_3	0.639	0.817	1.000	
Z_4	0.261	-0.301	0.109	1.000

THE FITTED REGRESSION EQUATION:

$$Z_2 = 0.00089\, Z_1^{0.9772}\, Z_3^{0.9193}\, Z_4^{-0.5534}$$

THE MEASURES OF MULTIPLE CORRELATIVE ASSOCIATIONS:
THE MULTIPLE CORRELATION COEFFICIENTS: 0.807
CORRECTED FOR 3 DEGREES OF FREEDOM: 0.859
THE COEFFICIENT OF MULTIPLE DETERMINATIONS: 0.844
CORRECTED FOR 3 DEGREES OF FREEDOM: 0.804

THE STANDARDISED REGRESSION (BETA) COEFFICIENTS:

B_2 =0.00089

$B_{1\ =}$ 0.9772

B_3 =0.9193

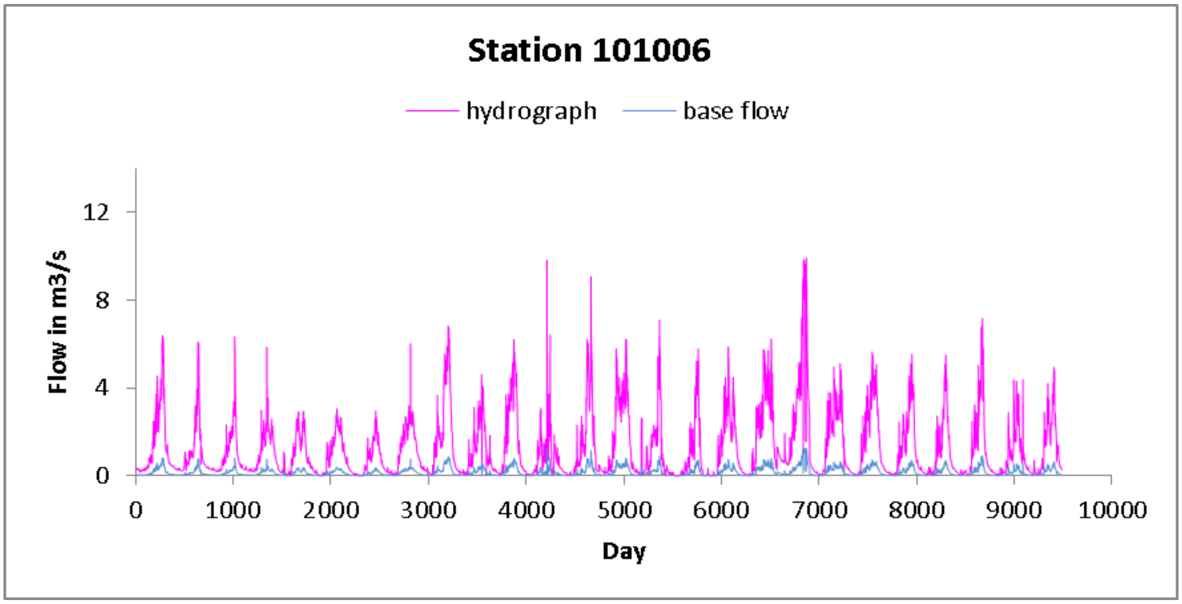
B_4 = -0.5534

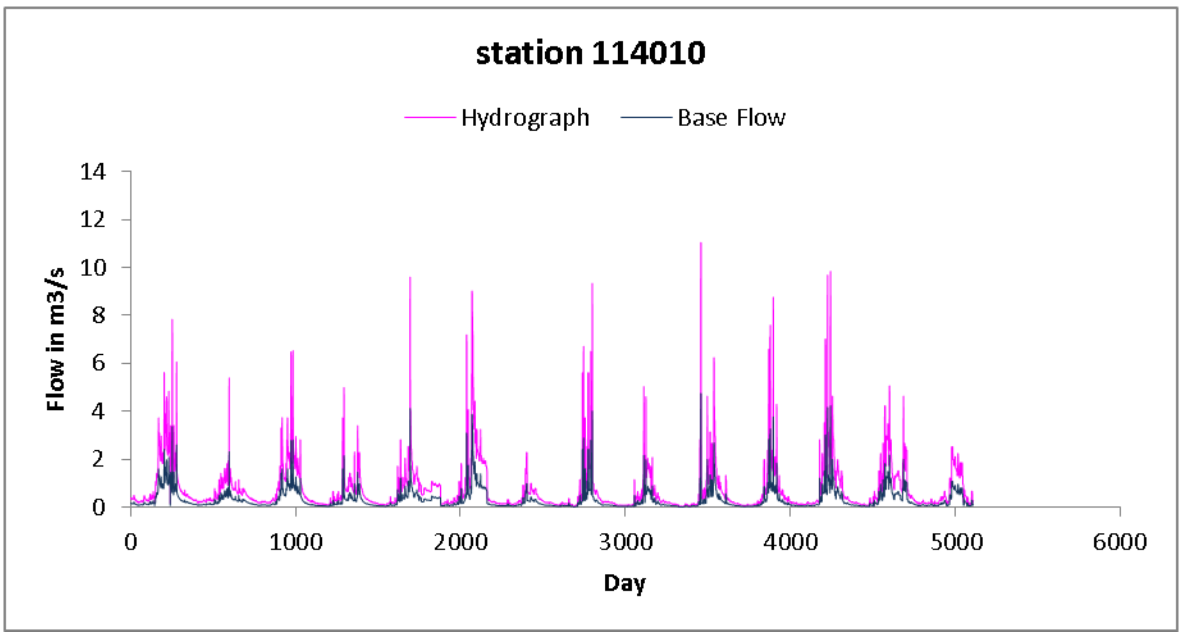
10.5 APPENDIX E

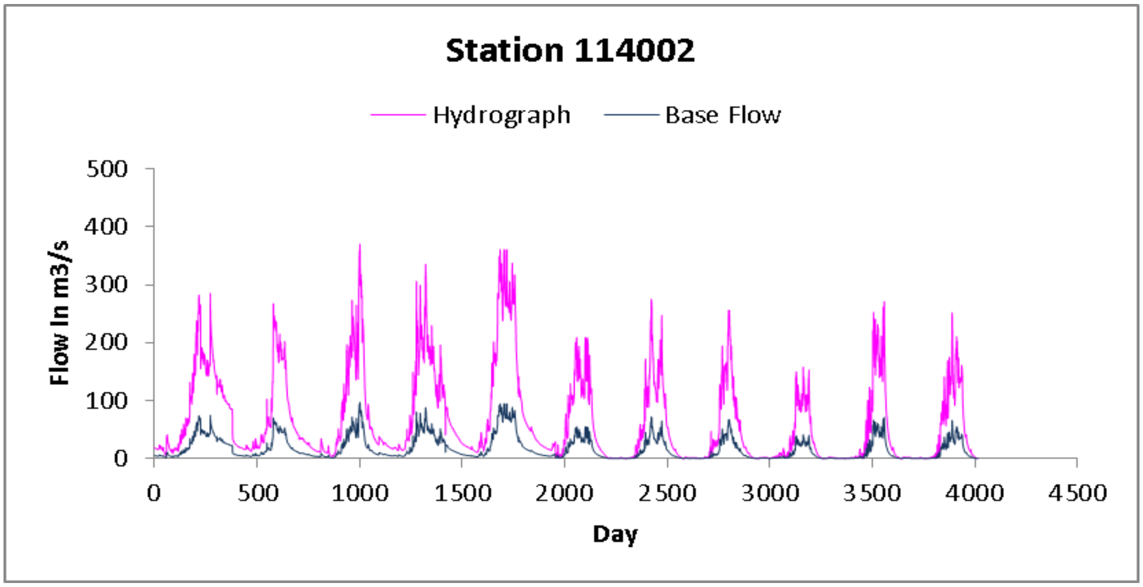
Base flow hydrograph and BFI of the analyzed station

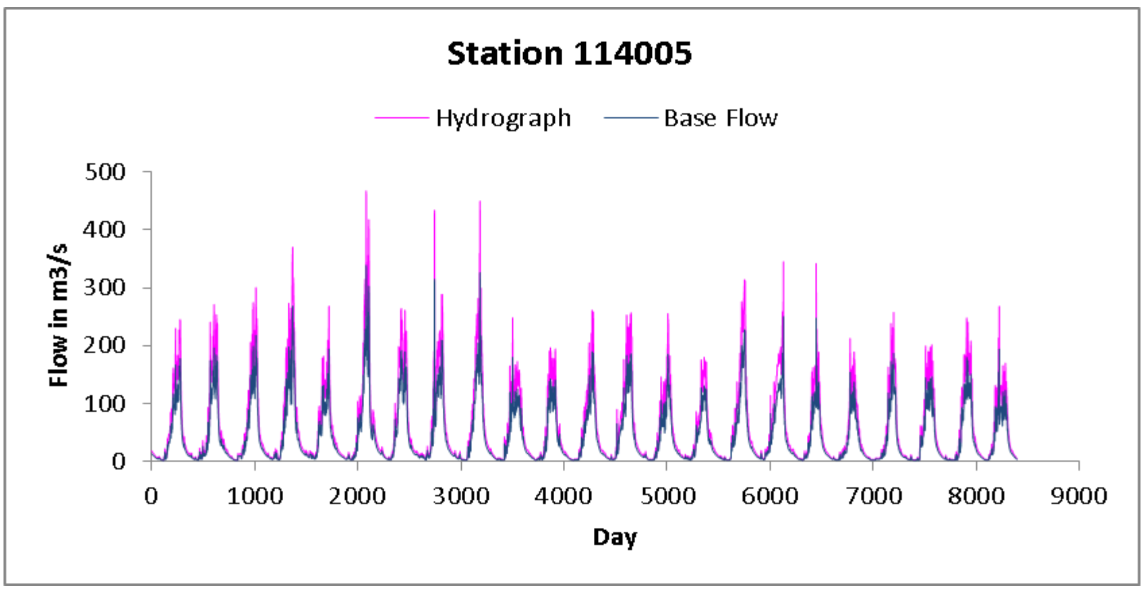
Table 10.6 Annual total flow, base flow and base flow index

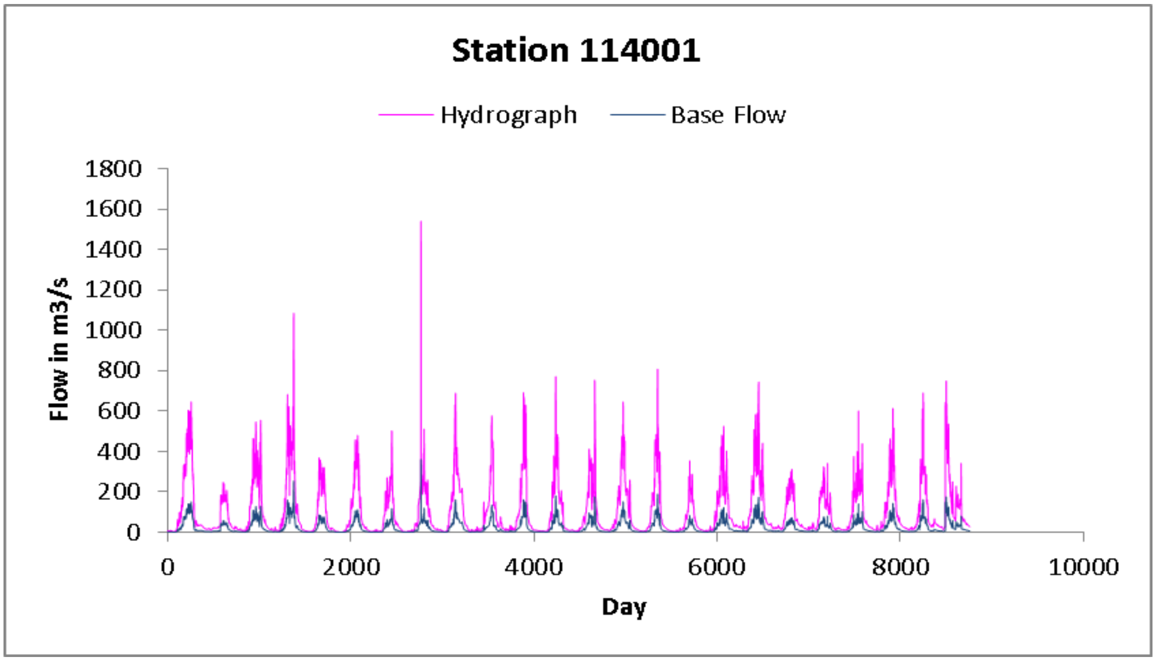
Station No	Total flow ,m^3/s per year	Base flow,m^3/s per year	Base flow index
114001	36709.05	8589.92	0.234
114002	23756.97	6242.24	0.262
114005	21429.16	15536.14	0.725
114007	8123.71	2006.56	0.247
114008	65.11	23.00	0.350
114009	280.32	96.99	0.346
114010	275.27	118.36	0.430
114012	440.62	226.48	0.514
114013	4144.52	2656.64	0.641
101006	465.54	59.12	0.127

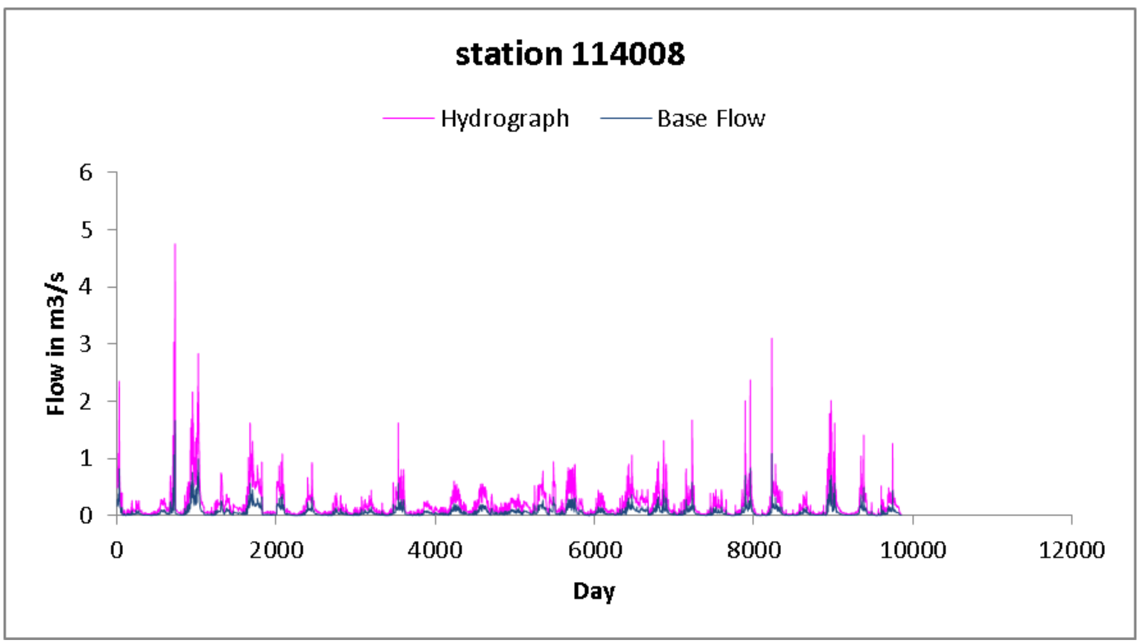

Station 101006
hydrograph
base flow
Flow in m3/s
12
8
4
0
0
1000
2000
3000
4000
5000
6000
7000
8000
9000
10000
Day

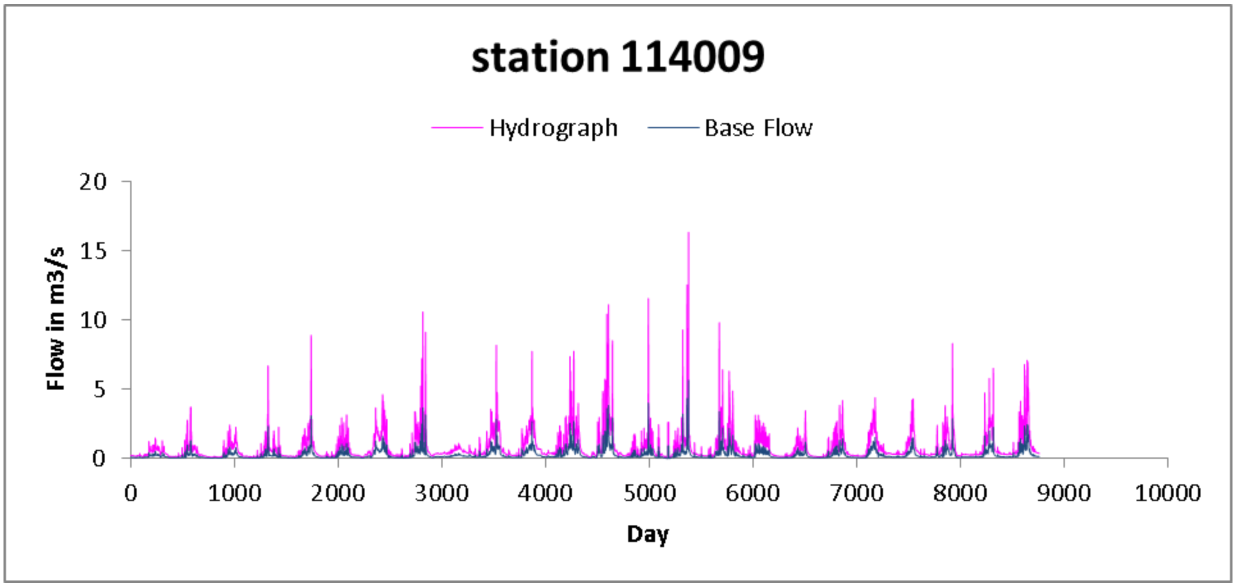

station 114010
Hydrograph
Base Flow
Flow in m3/s
14
12
10
8
6
4
2
0
0
1000
2000
3000
4000
5000
6000
Day

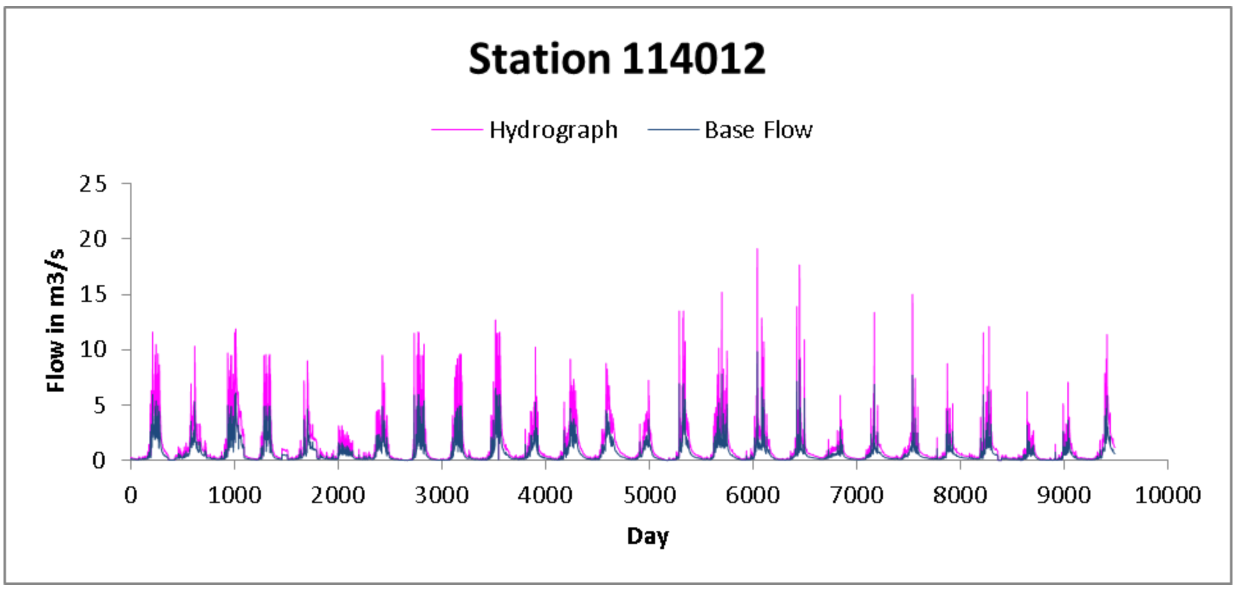

Station 114002
Hydrograph
Base Flow
Flow In m3/s
500
400
300
200
100
0
0
500
1000
1500
2000
2500
3000
3500
4000
4500
Day

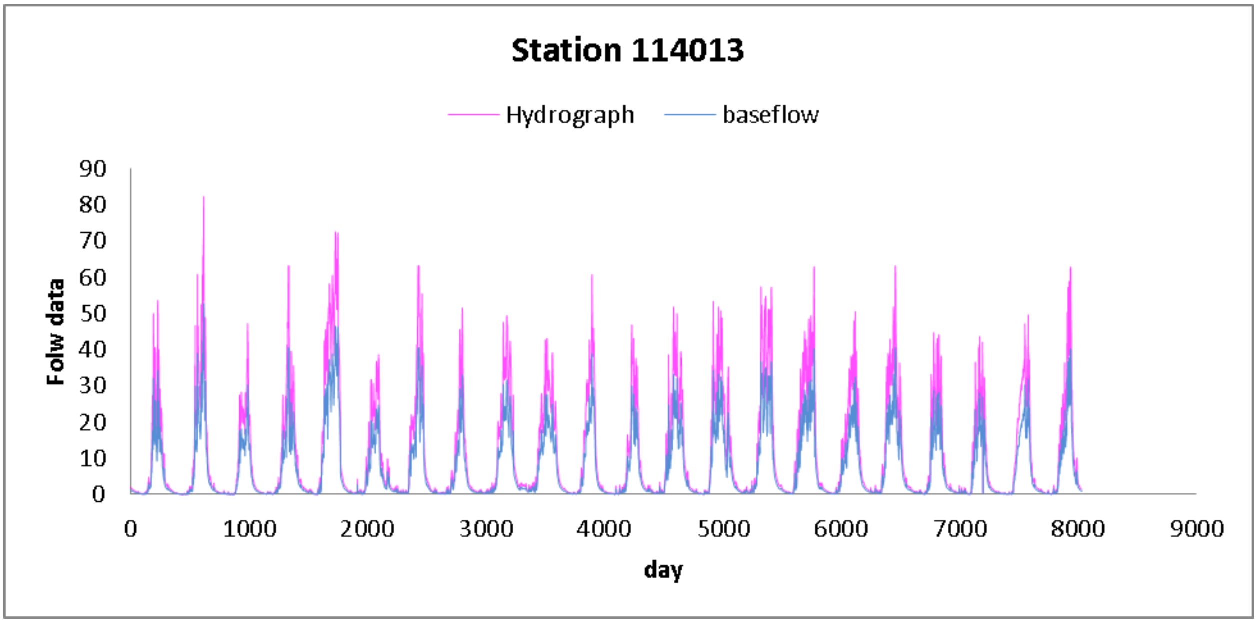

Station 114005
Hydrograph
Base Flow
Flow in m3/s
500
400
300
200
100
0
0
1000
2000
3000
4000
5000
6000
7000
8000
9000
Day

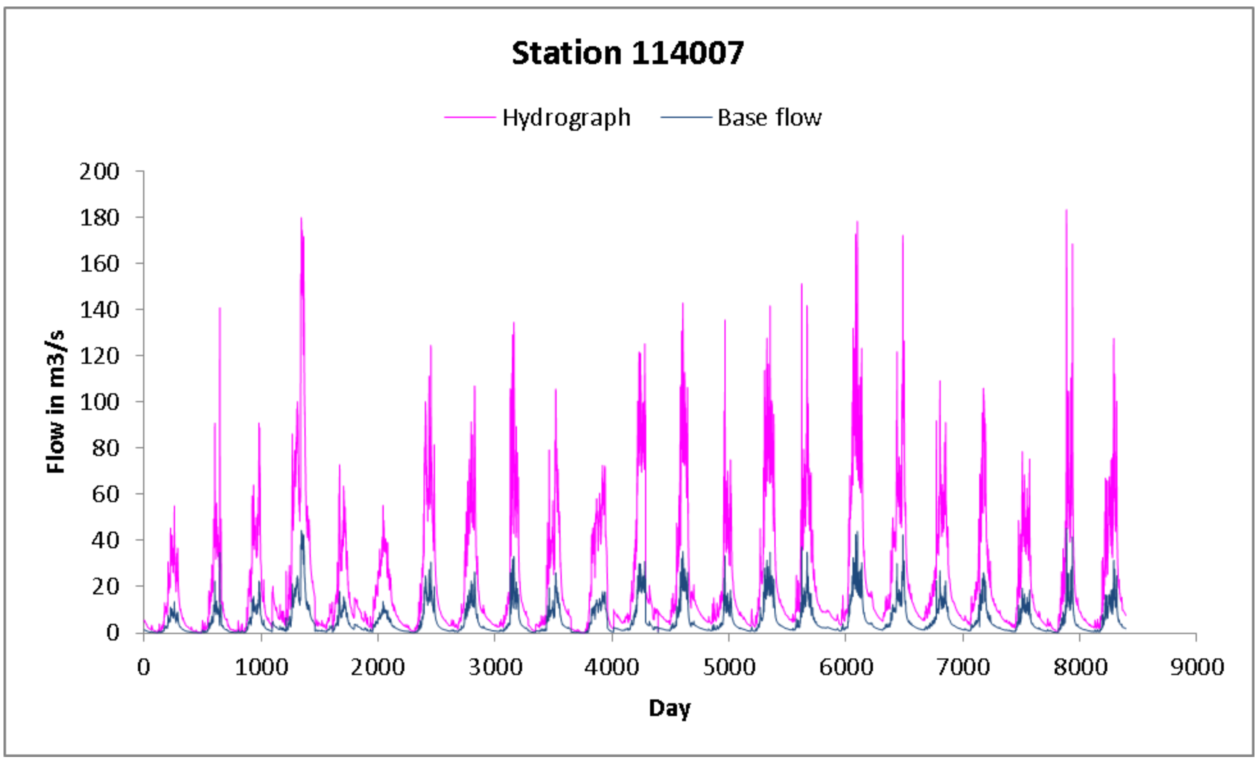

Station 114001
Hydrograph
Base Flow
1800
1600
1400
1200
1000
800
600
400
200
0
Flow in m3/s
0
2000
4000
6000
8000
10000
Day


station 114008
Hydrograph
Base Flow
6
5
4
3
2
1
0
Flow in m3/s
0
2000
4000
6000
8000
10000
12000
Day


station 114009
Hydrograph
Base Flow
Flow in m3/s
0
5
10
15
20
0
1000
2000
3000
4000
5000
6000
7000
8000
9000
10000
Day


Station 114012
Hydrograph
Base Flow
Flow in m3/s
0
5
10
15
20
25
0
1000
2000
3000
4000
5000
6000
7000
8000
9000
10000
Day


Station 114013
Hydrograph
baseflow
Folw data
day
0
10
20
30
40
50
60
70
80
90
1000
2000
3000
4000
5000
6000
7000
8000
9000


Station 114007
Hydrograph
Base flow
Flow in m3/s
Day
0
20
40
60
80
100
120
140
160
180
200
1000
2000
3000
4000
5000
6000
7000
8000
9000