

**ADDIS ABABA UNIVERSITY**  
**ADDIS ABABA INSTITUTE OF TECHNOLOGY**  
**SCHOOL OF CIVIL AND ENVIRONMENTAL ENGINEERING**



**Regional Regression Model of Mean Annual Streamflow For Ungauged  
Catchments**

**(Case study: Awash River Basin, Ethiopia)**

---

**A Thesis in Hydraulic Engineering**

By Zemene Kidanemariam

ID No: GSR/7167/10

Oct, 2019

Addis Ababa

A Thesis

Submitted in Partial Fulfillment of the Requirements for the Degree of Master of Science

The undersigned have examined the thesis entitled ‘**Regional Regression Model of Mean Annual Streamflow For Ungauged Catchments( case of Awash River Basin, Ethiopia)**’ presented by **Zemene Kidanemariam**, a candidate for the degree of **Master of Science** and hereby certify that it is worthy of acceptance.

Yilma Seleshi (PhD)

---

Advisor

---

Signature

---

Date

---

Internal Examiner

---

Signature

---

Date

---

External Examiner

---

Signature

---

Date

---

Chair person

---

Signature

---

Date

## UNDERTAKING

I certify that research work titled '**Regional Regression Model of Mean Annual Stream flow For Ungauged Catchments (Case of Awash River Basin, Ethiopia)**' is my own work. The work has not been presented elsewhere for assessment. Where material has been used from other sources it has been properly acknowledged / referred.

Zemene Kidanemariam

Signature\_\_\_\_\_

Email: [zeedagim21@gmail.com](mailto:zeedagim21@gmail.com)

Phone no: +251926463418

## ABSTRACT

Water resource developers and modelers focus on sustainable water resources management and availability of watercourses into account during design of hydraulic structures such as small dams for use at present and the future. Since most of the rivers in Ethiopia are ungauged, knowledge of predicting annual streamflow to be used for water resource development and other multiple purposes requires a study on stream flow of ungauged catchments. In this study an attempt was made to develop multiple linear regression model of mean annual stream flow for ungauged catchments of Awash River Basin using some statistical tests to secure the performance of the model. First the annual rainfall homogeneity test, trend analysis, and monthly and annual rainfall characteristics and variation were analyzed, based on this study the mean monthly rainfall of Awash River basin for the 19 stations varied from 3.15 mm to 411.8 mm in the period of 1981-2013 G.C and the mean annual rainfall distribution varies from 352.9 mm to 1766.4 mm, the mean annual flow varies from 36.68 MCM to 2026.82 MCM. Second development of equation for each gauged catchments based on different forms of equations (i.e. power, polynomial and linear functions) were performed and only one form of equation was chosen depending on model performance indices like  $R^2$ , NSE and RMSE results. The majority of the results show remarkably good fits with  $R^2$ -values ranging from 0.51 to 0.72 and NSE ranging from 0.53 to 0.7. The third and main task of this thesis work was developing a Regression model that could be used for estimation of mean annual flow at ungauged catchments, which was trained for Region-I using 9 gauging stations and for Region-II using 8 gauging stations. The performance and model fit was checked by statistical indicators and graphical methods. For Region-I it is found that the model with prediction variables Catchment area, mean annual rainfall and mean slope result the best fit giving ( $R^2=0.89$ ), (NSE=0.94) and (RMSE=107.24), Similarly For Region-II it is found that the model with prediction variables Catchment Area, mean annual Rainfall, mean Elevation and mean Slope result the best fit giving (NSE=0.88) and (RMSE=48.51). These model evaluation indicators were under a Very Good performances for Region-I and Good performances for Region-II.

**Key words:** ungauged catchments, River basin, regression analyses, multiple linear regression (MLR),

## **ACKNOWLEDGMENTS**

This thesis would not have been done without the support and encouragement of different individuals and organizations. First and foremost, I would like to express my deepest gratitude to my advisor, Dr. Yilma Seleshi for their serious follow up, proper guidance, constant encouragement, and shrewd and timely comments that have been helpful in shaping and enriching the thesis. I have learned many other qualities from him which would be important for my career as an academician and researcher.

My acknowledgements also go to the Ministry of Water Resources, the National Meteorological Agency of Ethiopia and Awash Basin Authority for their cooperation in providing relevant hydro meteorological data. I am appreciative to different staffs of Addis Ababa University, Faculty of Technology and Civil Engineering Department for having simplified various administrative matters.

Last but not least I am truly grateful for all the support I have received from my families next to the one during preparation of this thesis and of course during all the years I have studied to reach this place, and to my friends for their constructive thoughts on my entire work and wishes of successful accomplishment.

## TABLE OF CONTENTS

<b>ACKNOWLEDGMENTS</b> .....	<b>IV</b>
<b>TABLE OF CONTENTS</b> .....	<b>V</b>
<b>LIST OF TABLES</b> .....	<b>VIII</b>
<b>LIST OF FIGURES</b> .....	<b>IX</b>
<b>ABBREVIATION AND ACRONYMS</b> .....	<b>XI</b>
<b>1 INTRODUCTION</b> .....	<b>1</b>
1.1 Background .....	1
1.2 Statement of the Problem .....	2
1.3 Objective of the Study .....	3
1.3.1 General objective .....	3
1.3.2 Specific objectives .....	3
1.4 Research Questions .....	3
1.5 Significance of the study .....	3
<b>2 LITERATURE REVIEW</b> .....	<b>4</b>
2.1 Hydrology of ungauged catchments .....	4
2.2 Streamflow prediction in ungauged catchments .....	5
2.3 Regionalization of Mean Annual Flow .....	6
2.4 Related Studies in Ethiopia .....	6
2.5 Application of Regression analysis .....	8
2.6 Regression analysis .....	8
2.6.1 Simple Linear Regression .....	9
2.6.2 Multiple Linear Regression .....	9
2.6.3 MLR Implementation .....	10
2.6.4 Assumptions of OLS regression .....	10
2.6.5 Model evaluation statistics .....	11
<b>3 MATERIALS AND METHODS</b> .....	<b>13</b>
3.1 Description of the study area .....	13
3.2 Material and data .....	15

3.2.1	Hydrological data analysis.....	15
3.2.1.1	Rainfall data .....	15
3.2.1.2	Stream flow data.....	29
3.2.1.3	Physiographical data.....	37
3.3	Method.....	39
3.3.1	Software used.....	39
3.3.2	Method of analysis.....	40
3.3.2.1	Selection of climatological and physiographic variables .....	40
3.3.2.2	Classification of Regions.....	40
3.3.3	Computing the Regression Analysis.....	43
3.3.3.1	Test correlation between variables .....	44
3.3.3.2	Model specification .....	46
3.3.3.3	Estimation of model parameters.....	47
3.3.4	Model performance and Validation .....	47
<b>4</b>	<b>RESULTS AND DISCUSSIONS .....</b>	<b>49</b>
4.1	Estimation of Annual flow for the selected catchments .....	49
4.2	Estimation of runoff coefficient for the gauged stations .....	52
4.3	Regression Analysis of mean annual flow for Region-I.....	53
4.3.1	Results for Regression using all predictor variables.....	53
4.3.2	Results for Regression using the most appropriate predictor variables.....	56
4.4	Regression Analysis of mean annual flow for Region-II .....	57
4.4.1	Results for Regression using all predictor variables.....	57
4.4.2	Results for Regression using most appropriate variables .....	58
<b>5</b>	<b>CONCLUSIONS AND RECCOMENDATIONS.....</b>	<b>62</b>
5.1	Conclusions.....	62
5.2	Recommendations.....	63
	<b>REFERENCES .....</b>	<b>64</b>
	<b>APPENDICES.....</b>	<b>67</b>

Appendix A: Table Summarizes rain gauge missing data analyses.....	67
Appendix B: Sample rain gauge stations with filled missing data.....	69
Appendix C: Table showing Developed equations of Annual flow for each Gauging stations .....	73
Appendix: D Correlation coefficient and scatterplot of Region-II .....	75
Appendix E: Sample R code used in development of regression model .....	76

## LIST OF TABLES

Table 1 Models used for streamflow estimation in Gauged/Ungauged catchments .....	5
Table 2 Selected Rainfall Stations of Awash River Basin .....	16
Table 3 homogeneity test results using different methods at 95% of significance level..	19
Table 4 Mann-Kendall and Sen's slope estimator test results .....	23
Table 5 Selected Rainfall stations of Awash River Basin for the study.....	24
Table 6 Mean rainfall (mm) and CV (%) .....	26
Table 7 the selected gauging stations of Awash RB.....	31
Table 8 Outlier check of MAF for kesem at beke station.....	33
Table 9 Mean and CV of selected gauging stations .....	36
Table 10 List of variables used as predictors in the model.....	40
Table 11 Gauged stations used for training the model of Region-I.....	43
Table 12 Gauged stations used for training the model of Region-II .....	43
Table 13 VIF and Tolerance values between predictor/Explanatory variables .....	45
Table 14 Summary of Annual Streamflow equation for the selected Gauged catchments .....	49
Table 15 Value of runoff coefficients for the streamflow gauged stations .....	52
Table 16 Summary of all models with model performance indicators.....	59
Table 17 Criteria for the goodness-of-fit evaluation.....	59

## LIST OF FIGURES

Figure 1 map of the study area .....	14
Figure 2 Regression analyses for Addis Ababa bole rain gauge using Addis Ababa observatory as a donor rain gauge .....	17
Figure 3 Results for station Addis Ababa bole using different homogeneity test methods. (* $\mu$ represent the mean rainfall).....	20
Figure 4 change year in annual rainfall series/break year for station Boneya (* $\mu_1$ and $\mu_2$ represent the mean rainfall before and after the change point) .....	21
Figure 5 Monthly rainfall distribution of Awash River basin at the selected station .....	25
Figure 6 Annual Rainfall (mm) in Awash RB of the selected stations .....	27
Figure 7 Kriging map showing annual areal rainfall of Awash River basin for the year 2000 .....	28
Figure 8 Kriging map showing annual areal rainfall of Awash River basin for the year 1981-2013 .....	29
Figure 9 Map of selected gauged stations in Awash RB .....	30
Figure 10 Monthly stream flow pattern of Akaki gauging station .....	34
Figure 11 Monthly stream flow pattern of moja gauging station .....	34
Figure 12 Monthly stream flow pattern of Awash at meleka werer gauging station .....	35
Figure 13 Monthly stream flow pattern of Awash at wonji gauging station .....	35
Figure 14 The annual river discharge of Awash RB at 27 stations .....	36
Figure 15 Slope map of Awash river basin .....	38
Figure 16 Map showing the regions of the study area.....	41
Figure 17 Figure Map of upper Awash River basin .....	42
Figure 18 Test of correlation between variables for Region-I.....	45
Figure 19 The R script used to check multicollinearity using VIF and Tolerance.....	45
Figure 20 Scatter plots of the training variables for Region-I.....	46
Figure 21 Plot of observed and estimated annual flow awash at melkahomble gauging station.....	50
Figure 22 Plot of observed and estimated annual flow for kesem at awramelka gauging station.....	50
Figure 23 Plot of observed and estimated annual flow for little akaki gauging station ...	51

Figure 24 Summary of results for the regression using all predictor variables for Region-I .....	53
Figure 25 Residual plot of the model using all predictor variable .....	54
Figure 26 Plot of observed and estimated mean annual flow fit with 1:1 line .....	54
Figure 27 Plot of MLR model using all a predictor variable.....	55
Figure 28 Estimation of NSE and RMSE for Region-I .....	55
Figure 29 Plot of observed and estimated mean annual flow using most suitable variables .....	56
Figure 30 Summary results for the regression using all predictor variables of Region-II	57
Figure 31 Plot of observed and estimated mean annual flow using all variables.....	57
Figure 32 Summary results of the regression using area for Region-II.....	58
Figure 33 Mean Annual observed and predicted flow of Regression for Region-II .....	58

## ABBREVIATION AND ACRONYMS

HBV	Hydrologiska Byrans Vattenavdelning model
NSE	Nash-Sutcliffe Efficiency
RMSE	Root Mean Square Error
NMA	National Meteorological Agency
MLR	Multiple Linear Regression
MCM	Million Cubic Meter
MAF	Mean Annual Flow
SD	Standard Deviation
CV	Coefficient of Variation
MAR	Mean Annual Rainfall
GIS	Geographical Information Systems
MWIE	Ministry of Water, Irrigation and Electricity
OLS	Ordinary Least Squares
RB	River Basin
VIF	Variance Inflation Factor
TOL	Tolerance
UH	Unit Hydrograph
A	Catchment Area
E	Mean Elevation
S	Mean Slope

# Regional Regression Model of Mean Annual Streamflow for Ungauged Catchments

---

## 1 INTRODUCTION

### 1.1 Background

One of the world's most important and vital natural resources is water. The access to clean water and sanitation was recognized as a human right by the United Nations General Assembly on the 28 of July 2010. It is therefore important that water resources are managed in a way that allows for this right to be fulfilled by everyone, everywhere. Water has a wider range of use than for drinking and sanitation; for example irrigation, hydropower, navigation and recreation. This use has to go hand in hand with watershed management. In order to obtain a responsible management, estimates of annual watershed runoff volumes are required (Vogel, Wilson, and Daly 1999). Regionalization is a process of grouping basins or sub basins of the smallest order into homogeneous regions on the assumption that extreme flow populations at several sites in a region are similar in statistical characteristics, regardless of their catchment sizes. One alternative method that has been recognized is regional regression models, these models relate mean annual streamflow with geomorphic and climatic variables and have been widely used in the western and northern parts of the world where, in general, a similar climate regime can be found. The question is, however, if regional regression models would show as good results for other types of climate, specifically tropical ones.

Quantifying mean annual flow of rivers at ungauged sites is essential for assessments of global water supply, ecosystem integrity and water footprints. In ungauged catchments, model parameters are usually transposed from gauged catchments as no runoff data are available for calibrating them. Parameters can be either transposed from nearby catchments or, alternatively, functional relationships between model parameters and available catchment attributes can be derived from gauged catchments that may be further away. In this study model parameters were derived from the selected gauged catchments that can be used for estimation of mean annual flow in ungauged catchments using physiographic and climatologic predictor variables of the targeted ungauged catchments.

# Regional Regression Model of Mean Annual Streamflow for Ungauged Catchments

---

## 1.2 Statement of the Problem

Majority of the catchments throughout the world are ungauged (i.e. they lack stream flow observations) therefore a challenge for hydrologist's is to develop tools and strategies for predicting stream flow at these ungauged catchments (Wagener et al. 2004).

In Ethiopia most river basins have less coverage of hydro meteorological gauging stations. However, the country is on the way of exploiting its water resource potential. Full exploitation of the available water resource potential requires knowledge of the basin water balance. This in turn requires knowledge of the contribution of ungauged catchments of the basin. Since, in most of the basins of Ethiopia, there is lack of gauging stations and only few gauging stations have long recorded stream flow data. Consequently, there is a need to develop a means for predicting mean annual flow at the ungauged sites.

Based on Ethiopian water, irrigation and electricity research and development directorate researchable issues and list of research outputs/index, publication No 5<sup>th</sup> 2018, developing better tools and methods to estimate discharge for un-gauged catchments that are compatible to the country's situation should be promoted in no time. Most catchments found in the country are un-gauged therefore it needs the development of discharge estimating based on the climatological and physiographical factors dominating the particular area. In Ethiopia hydrology of ungauged catchments is a major challenge for researchers since there is no past experienced method to develop a model for ungauged basins, therefore in this study regional regression equation of mean annual flow for ungauged catchments will be developed using regression analysis methods for Awash River basin of Ethiopia which can be used by water resource development managers and decision makers on the current and future use of water resources of the country and meet the needs of water engineers who currently face tremendous challenges in designing small and medium hydraulic structures in the basin.

# Regional Regression Model of Mean Annual Streamflow for Ungauged Catchments

---

## 1.3 Objective of the Study

### 1.3.1 General objective

The general objective of this study is to develop a regional regression model of mean annual flow for ungauged catchments of Awash River basin using MLR model.

### 1.3.2 Specific objectives

The following are some of the specific objectives of this thesis;

- Homogeneity Testing, Trend Analysis In Long Term Rainfall Data For Awash River Basin over 30 years records and to asses monthly and annual rainfall and streamflow characteristics of awash river basin
- To formulate a reliable Rainfall runoff equation of annual flow for each of the selected gauged catchments and to estimate the runoff coefficients of these stations.

## 1.4 Research Questions

- Is there a possibility to formulate a rainfall-runoff relationship in any of gauged catchments using linear, power and polynomial functions?
- Which catchment attributes are more significant in development of MLR model for ungauged catchments?
- Can the developed regression model be used for estimation of mean annual flow of ungauged catchments of Awash River basin?

## 1.5 Significance of the study

Since water resource development managers and decision makers face tremendous challenges of knowing the amount of available volume of water at any ungauged catchments for the current and future use of water resources of the country, this study may have the tendency to resolve these problems. On the other hand water engineers who currently face difficulties in designing small and medium hydraulic structures in the basin may use any information addressed in this thesis work.

## 2 LITERATURE REVIEW

### 2.1 Hydrology of ungauged catchments

Flow estimation in ungauged catchments is probably one of the most basic and oldest tasks of hydrologists. This long-standing issue has received increased attention recently due to the prediction in ungauged basin initiative by the International Association of Hydrological Science. Predicting hydrologic quantities such as rainfall-runoff, sediment, nutrients etc. are a major challenges for hydrologists in developing countries where a large number of watersheds are ungauged (Razna R. & Shinna H., 2014). The history of hydrology of ungauged catchments starts more than 87 years since the inception of UH theory by Sherman (1932), it is still one of the methods most widely used for development of Flood Prediction and Warning Systems for gauged basins with observed rainfall-runoff data. Sherman paved the way for the advancement of watershed rainfall-runoff analysis with his UH technique (Singh, Mishra, and Jain 2014). The UH technique was able to develop flood hydrograph modelling with its advances and developments. The unit hydrograph developed from rainfall and streamflow data on a watershed applies only for that watershed and for the point on the stream where the observed Time streamflow data were measured on the other hand Synthetic unit hydrograph procedures are used to develop unit hydrographs from other locations on the stream to the same watershed or for nearby watersheds of a similar characteristics (Chow, Maidment & Mays, 2003). Snyder (1938), of Chow (2003) found synthetic relations for some characteristics of a standard unit hydrograph Additional such relations were found later U.S. Army Corps of Engineers (1959). According to Chow et al. (2003) One of the drawbacks of using UH for estimation of flow in ungauged catchments was the size of the watershed considered as a case study were limited varying in size from about 30 to 30,000 km<sup>2</sup>. Further developments were made by Clark, (1943), of Chow (2003) by the addition of parameters that is characteristic of watersheds reservoirs. In 1972, Soil Conservation Service (SCS) produces dimensionless synthetic unit hydrograph (Safarina et al. 2011). Unit hydrograph theory is frequently used for modeling the process of rainfall and runoff. A hypothetical relationship is formed between hydrologic and geomorphologic factors by Geographical Instantaneous Unit Hydrograph (GIUH) and

# Regional Regression Model of Mean Annual Streamflow for Ungauged Catchments

---

the direct runoff reaction of the catchment area to unit excess rainfall (Ghumman et al. 2014). Generally speaking using hydrograph models for estimation of flow in a watershed is site specific, suitable for relatively small watersheds and mostly used to estimate peak floods, but for a basin scale studies where multiple streamflow gauging stations considered may not be appropriate, due to this other regional hydrologic methods for estimating streamflow in ungauged catchments were used in different parts of the world as reviewed on the following sections.

## 2.2 Streamflow prediction in ungauged catchments

According to Saliha et al. (2011) most rivers in Africa are either ungauged or have limited hydro-meteorological data due to poorly developed hydrometric networks and lack of human and financial resources to develop and maintain such networks. Different studies world widely were conducted related with hydrology of ungauged catchments, using either of hydrologically model dependent or hydrologically model independent methods for different basins as well as in catchment level for estimation of a response variable based on the selected predictor (explanatory) variable of gauged catchments. The following are summaries of currently used models for streamflow forecasting in either of ungauged or gauged catchments (Swain, Jha, and Patra 2015).

Table 1 Models used for streamflow estimation in Gauged/Ungauged catchments

Distributed Physically-Based Models	Conceptual And Semi Distributed Models	Data-Driven Models
TOPMODEL, Mike 11 NAM and MIKE-SHE	HBV, SimHyd and IHACRES	Multiple Linear Regression (MLR), Autoregressive Moving Average (ARMA) and Artificial Neural Networks (ANNs).

According to Islam (2015) Physically based hydrologic models are based on known scientific principles of energy and water fluxes whereas, conceptual models are based on conceptual storages and model parameters that require calibration, or they are moisture

# Regional Regression Model of Mean Annual Streamflow for Ungauged Catchments

---

accounting models without explicitly considering energy fluxes, and so they mimic physical processes in a simplified manner. Data-Driven models, in contrast to the first two, involves mathematical equations that are not derived from physical processes in the catchment but from analysis of time series data. Examples include the unit hydrograph method, linear regression and ARIMA models (Sudheer, Gosain, and Ramasastri 2002).

## **2.3 Regionalization of Mean Annual Flow**

Predictions of annual streamflow are needed in many different studies of hydrologic studies. Usually a streamflow gauge is unavailable at the location of interest hence regional methods that relate streamflow to readily measure geomorphic and climate characteristics provide a practical solution (Vogel, 1999). Streamflow regionalization has been done mostly through hydrologic models such as Conceptual hydrologic models, HBV have been the most frequently used tools for streamflow regionalization. In ungauged catchments, model parameters have to be estimated from other sources of information. An interesting way to estimate model parameters in ungauged catchments is to collect the model parameters from hydrologically similar catchments. The concept of hydrological similarity assumes that the runoff response to a given rainfall input in two different catchments will be similar if similar rainfall-runoff processes occur (Merz, Blöschl, and Parajka 2006). There are different regionalization techniques stated in many papers mainly Regionalization based on spatial proximity and Regionalization based on catchment attributes were widely used. In this study regionalization based on catchment attributes using regression analyses between model parameters and catchment attributes were conducted and analyses was made using different regions that were formed based on rainfall distribution similarities. The study conducted by The U.S. Geological Survey (USGS), in cooperation with the Colorado Water Conservation Board and the Colorado Department of Transportation, developed regional regression equations for estimation of various streamflow statistics that are representative of natural streamflow conditions in Colorado (Casella, Fienberg, and Olkin 2013).

## **2.4 Related Studies in Ethiopia**

Admasu G. (1989) developed procedures for flood frequency analysis for use in Ethiopia that provides vital information for the planning and designing of many hydraulic

## Regional Regression Model of Mean Annual Streamflow for Ungauged Catchments

---

structures and for risk assessment in flood plain use. The main objective of the research was to estimate a flood magnitude ( $Q$ ) corresponding to any required return period ( $T$ ) of occurrence. He uses the method of regionalization for improving flood quantile estimations and confined to the “index-flood” methods proposed by Dalrymple (1960). Empirical equations that relate the mean annual flood with catchment characteristics was obtained by a multiple linear regression model. Estimation of flow of ungauged catchments is usually based on transferring or extrapolating information from gauged to ungauged site, a process called ‘Regionalization’. The area ratio, spatial proximity, sub-basin mean and regional model regionalization methods are used to predict discharges from ungauged catchment.

Another research was conducted by Gebregiorgis & Moges (2013) entitled as “basin regionalization for the purpose of water resource development in limited data situation case of Blue Nile river basin, Ethiopia”. They have delineate the study area into five homogeneous regions based on statistical parameters of station data. Then they found generalized logistic model is the best fit for Regions I and IV. For all distributions, probability weighted moment parameter estimation method is most efficient, but for log-Pearson Type III, ordinary moments are chosen. For each region, a unique regional flood frequency curve was developed. These curves are important to estimate the flood quantiles of ungauged catchments in the data scarce area of the basin. However, only regression equations to estimate regional flood frequencies have been developed for a country wide application and basin water resource development. There were limited previous studies in Awash River basin that have developed regression equations to estimate mean annual flow of ungauged catchments. In developing countries like Ethiopia most of the Rivers are ungauged. Therefore, applying regression techniques for ungauged or poorly gauged river basin is crucial. In this thesis data driven techniques of estimating mean annual flow for ungauged catchments was conducted using multiple linear regression analysis by considering the basin as a region and developing a representative equation that may be used for estimation of mean annual stream flow of any ungauged catchment in the basin.

# Regional Regression Model of Mean Annual Streamflow for Ungauged Catchments

---

## 2.5 Application of Regression analysis

The river basin is increasingly being used as the unit for natural resource planning and management; to facilitate this, basin-scale models are being developed. For many basin-scale planning activities, detailed rainfall-runoff modelling is neither necessary nor tractable (Wilkinson, Young, and DeRose 2006). Regression analysis refers to a collection of statistical techniques commonly used for modeling the association between one or more independent variables, hereafter denoted explanatory variables or predictor variables and a dependent variable, which is expressed as a functional form of the independent variables and a set of parameters or coefficients, and is referred to as the response variable.

Regression analysis has become a widespread used tool for data description, estimation and prediction in many fields of science, such as physics, economics, engineering, and, of special interest for this thesis, due to the simplicity of its application framework and the appeal of its rigorous theoretical foundation. Regression-based approaches to calculate MAF are less time-consuming and computationally less intensive. Moreover, regression equations relating streamflow to explanatory catchment area, precipitation, mean elevation and mean slope may help to better understand general hydrological patterns and processes across different scales.

## 2.6 Regression analysis

The use of so called conceptual rainfall runoff models such as HBV or a physically based model such as MIKE SHE is more complex and require a relatively large amount of input data which has a high cost both in time and money. Physically-based distributed models particularly aim at predicting the effects of land use change. However, considerable debate on both the advantages and disadvantages of such models has arisen along with research and applications of those models (Dar 2017). Another possibility is to use empirical models which usually imply mathematical equations based on time series of input and output data. Regional regression models can produce an objective equation that can easily be integrated and implemented in water resource management as well as computer software's such as GIS. The major advantage is the models can give

# Regional Regression Model of Mean Annual Streamflow for Ungauged Catchments

---

information about both mean and variance of stream flow, this implies a wide range of information and guidance for planning and managing water resources (Vogel, Wilson, and Daly 1999).

Regression analysis is a widely used statistical tool, it is a conceptual method used to examine the relationship between a response variable and one or more so called explanatory or predictor variables. In this thesis the response variable is the streamflow in the region of interest and the parameters that were examined as predictor variables will be area, Mean Annual Rainfall, Mean elevation and Mean slope. In this study simple linear regression and multiple linear regression were used, the models was fitted by using ordinary least squares method of analysis to estimate the coefficients of predictor variables and compute equations for mean annual flow of ungauged catchments.

## 2.6.1 Simple Linear Regression

The simple linear regression is a model that results in an estimated line and its equation describes the estimated dependent variable. The line is a consequence of the equation since only one predictor or independent variable is used to explain the dependent or response variable.

## 2.6.2 Multiple Linear Regression

Multiple linear regression is a prolongation of simple linear regression, but consists of more than one explanatory or predictive variables thus it is harder to visualize the multiple linear regression because more dimensions are involved. MLR attempts to model the relationship between two or more explanatory variables and a response variable, by fitting a linear equation to the observed data.

In this thesis both simple and multiple linear regression were used to develop regression equation examined that incorporates various predictor variables. The performances of regression models in estimating the flow for ungauged sites have been assessed in. In order to estimates stream flows at ungauged sites, power form function such as:

$$Q = \alpha_0 B_1^{\alpha_1} B_2^{\alpha_2} \dots B_m^{\alpha_m} \varepsilon_0 ,$$

This form of equation is commonly used to build relation between streamflow and the catchment characteristics. Here,  $\alpha_0, \alpha_1, \dots, \alpha_m$  are the model parameters  $B_1, B_2, \dots, B_m$  are

# Regional Regression Model of Mean Annual Streamflow for Ungauged Catchments

---

the predictor variables ,  $\varepsilon_0$  is the error term,  $m$  is the number of predictor variables and  $Q$  represents Mean annual flow in MCM particularly used in this study, this equation can be solved using linear regressions by linearizing the power form model using a logarithmic transformation form. Transformations are used for three purposes; to make data more symmetric, to make data more linear and to make data more constant in variance. Logarithmically transforming variables in a regression model is a very common way to handle situations where a non-linear relationship exists between the independent and dependent variables. Using the logarithm of one or more variables makes the effective relationship non-linear, while still preserving the linear model (Benoit 2011).

### 2.6.3 MLR Implementation

(Badyalina and Shabri 2013) uses a jackknife multiple linear regression to simulate flood quantile using MATLAB software, they expressed the response variable  $QT$  as a function of catchment area ( $\text{km}^2$ ), annual mean rainfall (mm), elevation (m), longest drainage path (m) and mean catchment slope (%). Also (Engeland and Hisdal 2009) estimate low flow estimation for ungauged catchments using regression method and HBV model, they found that the regression method in general gave better prediction of  $QC$  in ungauged catchments than did the HBV-model.

In this thesis a step wise multiple linear regression was simulated using R software and Microsoft excel. The observed mean annual flow data are expressed as a function of catchment area ( $\text{km}^2$ ), mean annual rainfall (mm), Elevation (m), and mean catchment slope (%). the observed flow and the four explanatory variables are converted into the natural logarithm form. The model then is fitted by ordinary least squares procedures.

### 2.6.4 Assumptions of OLS regression

The criteria for selecting the best-fit OLS regression for the basin primarily were based on the regression equation with the largest  $R^2$ , NSE and smallest RMSE while also having explanatory variables with significant coefficients (p-value less than 0.05) for the regression equations, normality of the residuals, and Additional considerations in selection of the best-fit OLS regression included investigation of the normality of the explanatory variables (Streamstats 2015). In this study priority was given for the

# Regional Regression Model of Mean Annual Streamflow for Ungauged Catchments

---

significant p-value, normality of residuals and for statistical indexes  $R^2$ , NSE and RMSE values.

## 2.6.5 Model evaluation statistics

Regression models are powerful tools frequently used to predict a dependent variable from a set of predictors, they are widely used in a number of different contexts. An important problem is whether results of the regression analysis on the sample can be extended to the population the sample has been chosen from. If this happens, the model has a good fit and referred to this question as a goodness-of-fit analysis, performance analysis or model validation analysis for the model (Salmaso, 2007). A well-fitting regression model results in predicted values close to the observed data values as shown on the result and discussion part of (figure 26). The following are some of the recommended model evaluation statistics that were selected based on applicability to various models and climatic conditions; commonly used, accepted, and recommended in published literature and identified strengths in model evaluation.

**Coefficient of determination,  $R^2$ :** One of the important tests for examining the quality of the model is with the coefficient of determination also called goodness of fit index or as most commonly used  $R^2$ . This provides information about the accuracy of the regression line (simple regression) i.e. how much the dependent variable can be explained by the predictor variable(s). It is a coefficient that ranges from 0-1 where 1 means a perfect fit between the observed values and the fitted ones, in other words, that the predictor variable(s) perfectly explain the dependent variable.

$$R^2 = \frac{\sum(P_i - Q_m)^2}{\sum(Q_i - Q_m)^2}$$

**Root Mean Square Error (RMSE):** The RMSE is the square root of the variance of the residuals. It indicates the absolute fit of the model to the data how close the observed data points are to the model's predicted values. Whereas  $R^2$  is a relative measure of fit, RMSE is an absolute measure of fit. As the square root of a variance, RMSE can be interpreted as the standard deviation (SD) of the unexplained variance as shown on (table 17) of the result and discussion part and has the useful property of being in the same units as the response variable. Lower values of RMSE indicate better fit. RMSE is a

## Regional Regression Model of Mean Annual Streamflow for Ungauged Catchments

---

good measure of how accurately the model predicts the response, and is the most important criterion for fit if the main purpose of the model is prediction.

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (O_i - P_i)^2}{N}}$$

Where  $O_i$  and  $P_i$  represent the sample (of size  $N$ ) containing the observations and the model estimates, respectively.

**Nash-Sutcliffe Efficiency (NSE):** The Nash-Sutcliffe efficiency (NSE) is a normalized statistic that determines the relative magnitude of the residual variance compared to the measured data variance (Nash and Sutcliffe, 1970).

$$NSE = 1 - \frac{\sum_{i=1}^n (O_i - P_i)^2}{\sum_{i=1}^n (O_i - O_m)^2} = 1 - \left(\frac{RMSE}{SD}\right)^2$$

Where  $O_m$  is the mean of the observed values and  $SD$  represents the standard deviation of the observations.

## 3 MATERIALS AND METHODS

### 3.1 Description of the study area

The Awash River Basin is the fourth largest catchment in Ethiopia in terms of area, following Wabi- Shebele, Abbay and Genale-Dawa River Basins. It lies within 7°53'47" to 12°7'20" N latitude and 37°56'23" to 42°57'21" E. The basin is bordered by Danakil, Abbay, OmoGibe, Rift-Valley lakes and Wabi Shebele basins and Republic of Djibouti. The river originates near Ginchi in the central highlands of Ethiopia, and flows north east through the northern section of the Rift Valley to eventually discharging into Lake Abbe near Djibouti boarder, traveling a distance of about 1200km. The total catchment area of the basin is about 121961 km<sup>2</sup>.

The mean annual precipitation of the basin varies from about 1,600 mm, in the highlands north east of Addis Ababa, to 160 mm, in the northern point of the basin and Rainfall distribution is mostly bimodal in the Middle and Lower Awash and unimodal in the Upper Awash (Bekele et al. 2017).

There are three main dams in Awash River Basin; Koka, Kessem and Tendaho. Koka Dam is located in the upper Awash and is important for hydropower generation and irrigation development. The Koka reservoir is a main source of water supply for users of downstream dams from Amibara to Gewane. Kessem and Tendaho dams are located in the Middle and Lower Awash, respectively, for irrigation and household water supply. Major crops grown in the area are sugarcane, vegetables, fruit and cotton. The Awash River gives regular irrigation water to productive and larger agricultural areas. Moreover, it generates hydropower energy and provides water supply for different towns and cities along its course, which includes Legadadi, Gefersa, Dire and Aba-Samuel reservoirs (Tadese et al. 2019).

# Regional Regression Model of Mean Annual Streamflow for Ungauged Catchments

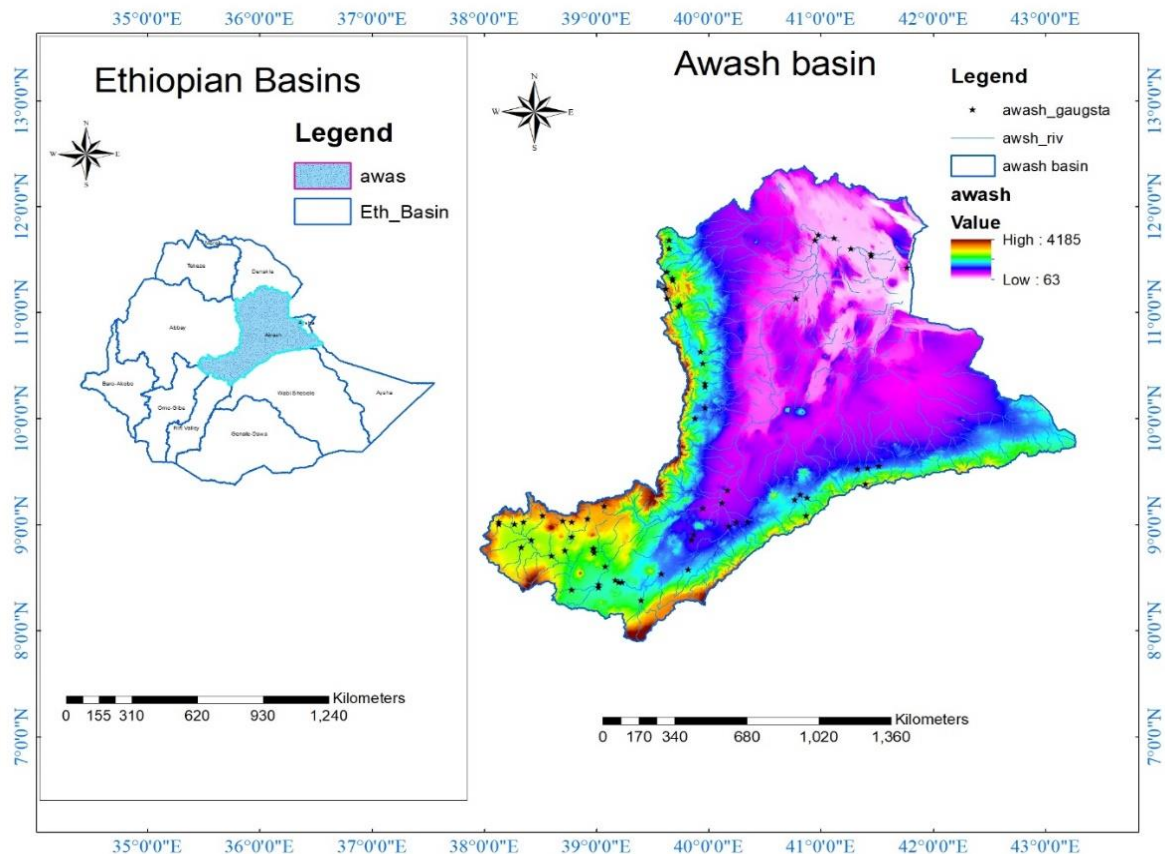


Figure 1 map of the study area

According to the Addis Ababa university team (2018) Report on Charge System for Irrigation Water Abstraction and Use, the Awash basin is the most utilized in Ethiopia, It serves as a source of drinking water, hydropower, industrial consumption, irrigation and disposal of waste water. The total irrigation area of the basin is found to be more than 200,000 ha. The irrigation practice in Awash River basin is dominated by surface irrigation which is the least efficient irrigation practice. A major user and consumer of the surface water in the Basin is irrigated agriculture through numerous private and government irrigation farms and stakeholders. The Basin has more potential for further irrigation development owing to suitable topography and ease of diversion of water along the river. But as the population and the economy grow rapidly, water is becoming increasingly scarce resources.

Awash River basin ranks seventh from the twelve river basin of the country in terms of surface water resources. The basin has an annual flow of 4.9 billion cubic meters ( $\text{BM}^3$ ) and 0.14  $\text{BM}^3$  estimated groundwater potential and The total length of Awash River is

# Regional Regression Model of Mean Annual Streamflow for Ungauged Catchments

---

1200 km and starts flowing at an altitude of 3000 m in the Central Highlands of Ethiopia and flows through different locations, and finally joins Lake Abe at a 250 m altitude above mean sea level (amsl), The basin covers diverse parts of the country, namely: Amhara, Oromia, Afar, Dire Dawa and Addis Ababa (Tadese et al. 2019).

## **3.2 Material and data**

DEM, hydrological data, watershed and other shape files are obtained from MWIE, I have already get most of GIS related files. In order to create and develop regional mean annual flow for ungauged catchments require long records of data (greater than 30 years of recorded data) but there is high shortage of available data in both record length and quality. Therefore stream flow data having a record year of greater than 5 years were used in this study. Also metrological data (ground rainfall data and satellite based rainfall data) were obtained from NMA.

### **3.2.1 Hydrological data analysis**

#### **3.2.1.1 Rainfall data**

Currently in Awash River basin there are more than 113 rainfall gauging stations and the rainfall stations used in this study were selected based on long-term data availability, percentage of missing data and representativeness of the basin. Monthly rainfall data were collected from 28 rainfall stations located within Awash River basin, the duration selected was more than 30 years, from 1981 to 2013. The following table summarizes the selected rainfall stations of Awash River basin.

# Regional Regression Model of Mean Annual Streamflow for Ungauged Catchments

Table 2 Selected Rainfall Stations of Awash River Basin

S. No	Station Name	Elevation	long	lat	Record length	percentage of missing
1	Addis Ababa Bole	2354	38.79871	8.981081	1981-2013	1.77%
2	Addis Ababa Obs.	2386	38.7475	9.01891	1981-2013	3.54%
3	Akaki	2057	38.7862	8.8698	1981-2013	1.01%
4	Asgori	2072	38.3342	8.79	1981-2013	3.54%
5	Bologiorgis	1963	39.3585	8.836	1981-2013	3.79%
6	Boneya	2251	38.64167	8.7845	1981-2013	3.03%
7	Busa	2166	38.1382	8.7725	1981-2013	8.84%
8	Ginchi	2132	38.1333	9.01667	1981-2013	1.52%
9	Ginager	3100	39.566667	9.316667	1981-2013	4.04%
10	Guder	2011	39.754667	8.95767	1981-2013	0.76%
11	Guranda Meta	2187	38.5892	8.9072	1981-2013	1.77%
12	Kimoye	2150	38.33783	9.008	1981-2013	3.79%
13	Kora	1239	40.525	9.104667	1981-2013	14.39%
14	Nazeret	1622	39.283333	8.55	1981-2013	5.81%
15	Sendafa	2569	39.0215	9.152167	1981-2013	8.33%
16	Sebeta	2240	38.63	8.93	1981-2013	5.05%
17	Teji	2091	38.366667	8.8333	1981-2013	1.26%
18	Welenchiti	1458	39.43	8.67	1981-2013	6.82%
19	Welenkomi	2165	38.254667	9.001833	1981-2013	11.11%
20	Dertuliben	1991	38.1232	8.97	1981-2013	6.82%
21	Ejere	2254	39.25683	8.7733	1981-2013	7.83%
22	Jara	1960	39.988707	10.3098	1981-2013	2.02%
23	Adaiytu	507	40.783	11.11	1981-2013	0%
24	Combolcha	1857	39.717633	11.083899	1981-2013	1.26%
25	Dire Dawa	1180	42.5333	9.9667	1981-2013	1.77%
26	Haik	1985	39.680205	11.305316	1981-2013	3.79%
27	Dessie Zuria Met	2553	39.633333	11.12	1981-2013	21.20%
28	Awash Melka	916	40.15	8.983333	1981-2013	19.19%

### 3.2.1.1.1 Data Quality

#### Missing data analysis

Missed rainfall data were present in the rain gauges chosen and are filled using linear regression using nearby rain gauges (i.e. donor gauges). Monthly rainfall was plotted between the target rain gauges and nearby rain gauges and a linear regression equation

# Regional Regression Model of Mean Annual Streamflow for Ungauged Catchments

was established. Figure 2 shows an example of such types of plots for Addis Ababa Bole rain gauge and Addis Ababa observatory as a donor rain gauge. The donor rain gauge was selected based on the geographical proximity from the target rain gauge and the strength of  $R^2$  value.

As shown on (Figure 2) the donor rain gauge selected to infill missing monthly rainfall data for each rain gauge and the associated regression equation. As shown on the appendix: A mostly, the donor gauge set as priority 1 is the primary gauge used to infill missing data of the target gauge, but if the primary gauge itself also has missing data at that time step, then the secondary donor rain gauge was used to infill and so on.

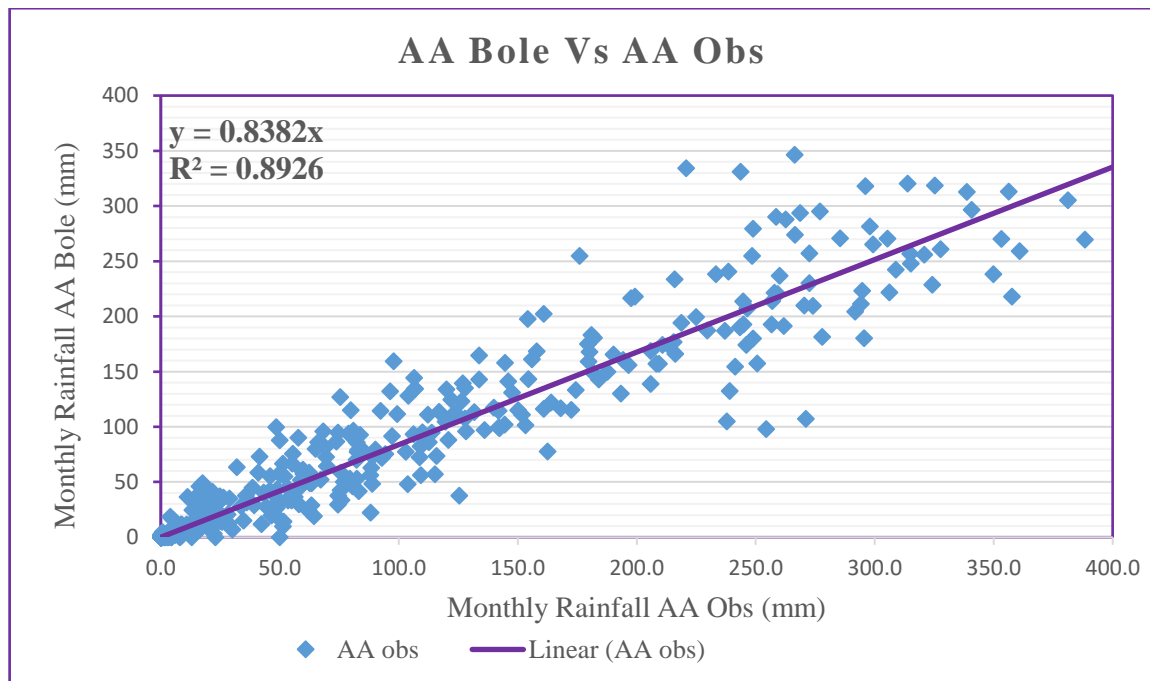


Figure 2 Regression analyses for Addis Ababa bole rain gauge using Addis Ababa observatory as a donor rain gauge

### 3.2.1.1.2 Homogeneity test

The homogeneity of the annual total rainfall was investigated by using statistical analysis tool, XLSTAT 2019, based on the four methods of homogeneity test; Pettitt's test, Standard normal homogeneity test (SNHT), Buishand's test (BRT) and Von Neumann's ratio (VNR) test. XLSTAT statistical software used hypothesis testing method to detect homogeneity of the rainfall data (Agha, Bağçacı, and Şarlak 2017). The results are categorized into three classes, which are useful, doubtful and suspect according to the

## Regional Regression Model of Mean Annual Streamflow for Ungauged Catchments

---

number of tests rejecting the null hypothesis. The classification of the test results, according to the methods proposed by is as follows.

- Useful: The series that rejects one or none null hypothesis under the four tests at 1% significance level are considered. Under this class, the series is grouped as homogeneous and can be used for further analysis.
- Doubtful: The series that reject two null hypotheses of the four tests at 1% significance level is placed in this class. In this class, the series have the inhomogeneous signal and should be critically inspected before further analysis.
- Suspect: When there are three or all of the tests reject the null hypothesis at 1% significance level, then the series is classified into this category. In this category, the series can be deleted or ignored before further analysis.

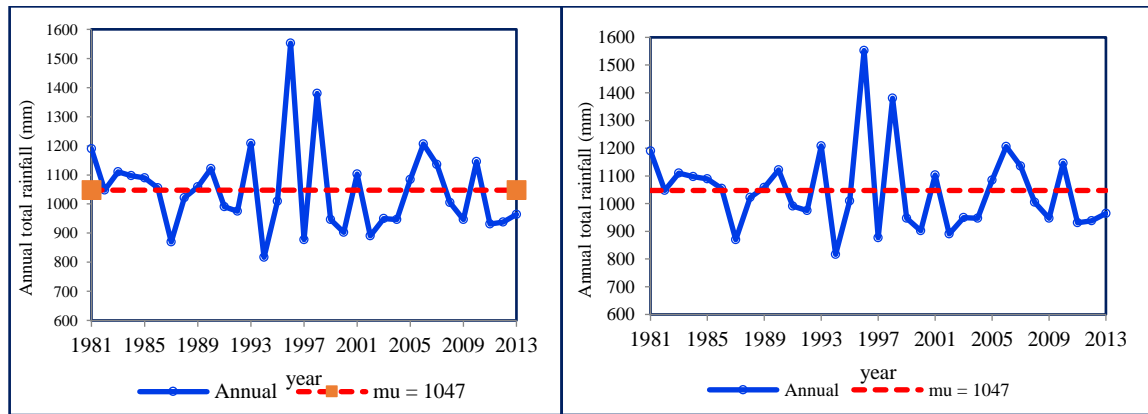
Based on alpha value of 0.05 (95% significance level), if p value is bigger than alpha value, therefore the series is homogeneous. Table 2 shows the summary result of the test. a total of nineteen (19) rainfall stations were found homogeneous (labeled as “useful”) and seven (7) stations was found as inhomogeneous (labeled as “suspect”).

## Regional Regression Model of Mean Annual Streamflow for Ungauged Catchments

Table 3 homogeneity test results using different methods at 95% of significance level

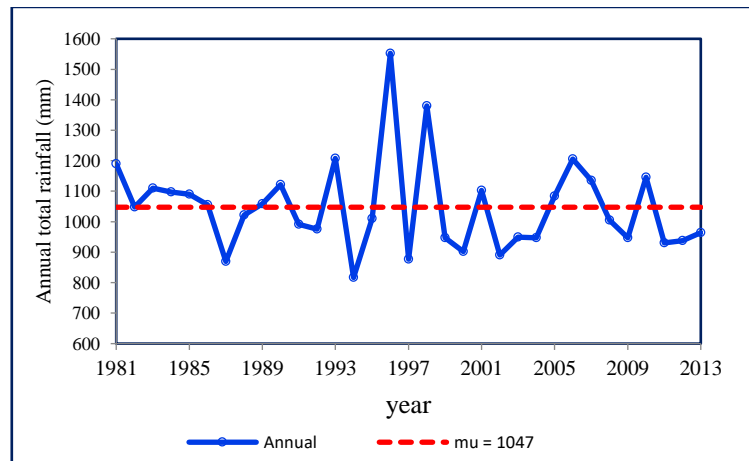
sr.no	station name	Pettitt's Test	SNHT test	Buishand	von Neumann	Result
1	Addis Ababa Bole	0.8376	0.7946	0.535	0.9317	useful
2	Addis Ababa Obs	0.4134	0.8407	0.9651	0.957	useful
3	Adaiytu	0.9236	0.0412	0.3872	0.3432	useful
4	Akaki	0.043	0.071	0.054	0.309	useful
5	Asgori	0.098	0.197	0.069	0.437	useful
6	Awra Melka	0.762	0.071	0.300	0.102	useful
7	Bologiorgis	0.950	0.763	0.477	0.910	useful
8	Boneya	0.022	0.019	0.008	0.004	suspect
9	Busa	0.005	0.008	0.005	0.001	suspect
10	Combolcha	0.219	0.771	0.494	0.233	useful
11	Dertuliben	0.068	0.001	0.019	0.047	suspect
12	Dessie Zuria Met	0.123	0.079	0.049	0.021	doubtful
13	Diredawa	0.519	0.616	0.321	0.368	useful
14	Ejere	0.358	0.236	0.093	0.454	useful
15	Ginager	0.878	0.231	0.224	0.016	useful
16	Ginchi	0.779	0.545	0.561	0.460	useful
17	Guder	0.288	0.424	0.241	0.001	useful
18	Guranda Meta	0.001	0.000	0.000	0.003	suspect
19	Haik	0.900	0.613	0.416	0.396	useful
20	Jara	0.012	0.034	0.005	0.028	suspect
21	Kimoye	0.974	0.018	0.201	0.061	useful
22	Nazret	0.595	0.760	0.819	0.990	useful
23	Kora	< 0.0001	0.000	< 0.0001	< 0.0001	suspect
24	Sebeta	0.083	0.138	0.013	0.001	Doubtful
25	Sendafa	0.129	0.703	0.988	0.970	useful
26	Teji	0.164	0.968	0.869	0.824	useful
27	Welenchit	0.774	0.700	0.707	0.859	useful
28	Welenkomi	0.005	0.049	0.031	0.184	suspect

# Regional Regression Model of Mean Annual Streamflow for Ungauged Catchments



a) Pettitt's test

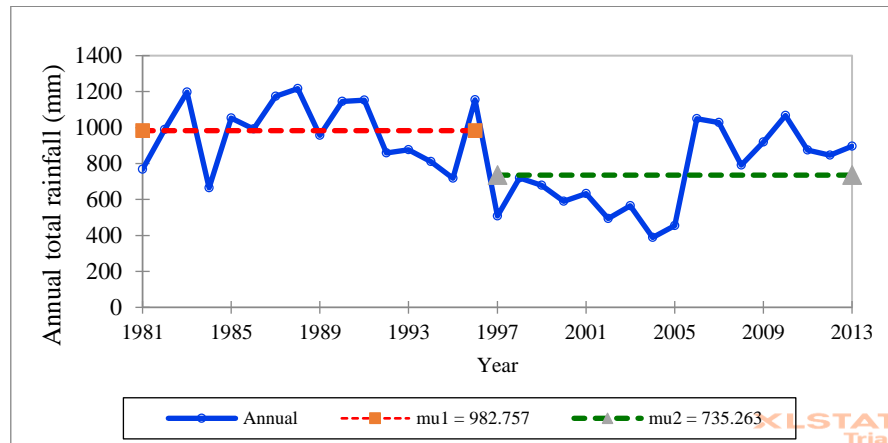
b) SNHT test



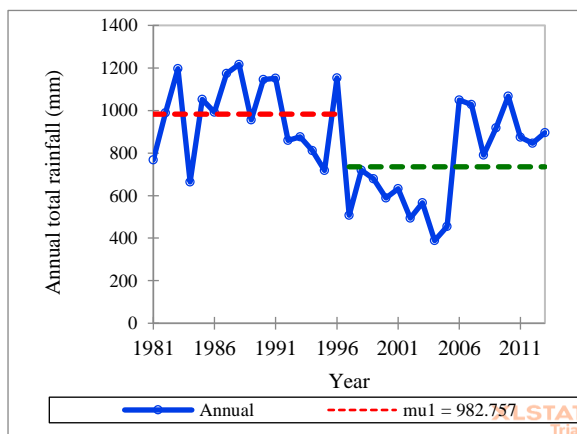
c) BRT test

Figure 3 Results for station Addis Ababa bole using different homogeneity test methods. (\*mu represent the mean rainfall)

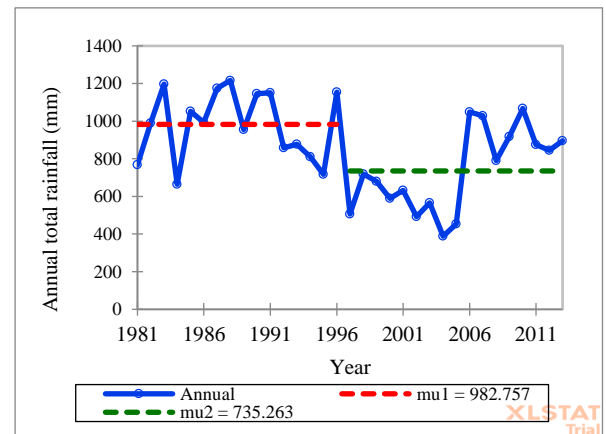
# Regional Regression Model of Mean Annual Streamflow for Ungauged Catchments



a) Pettitt's test



b) SNHT test



c) BRT test

Figure 4 change year in annual rainfall series/break year for station Boneya (\*mu1 and mu2 represent the mean rainfall before and after the change point)

### 3.2.1.1.3 Trend Analysis

A trend analysis of annual rainfall variability was conducted by means of the Mann–Kendall (MK) test using the homogenous time-series rainfall data. In MK test, the positive test statistic,  $S$  indicates an increasing trend, while a negative value indicates a decreasing trend in rainfall. When setting up a study, a risk threshold above which  $H_0$  should not be rejected must be specified. This threshold is referred to as the significance level  $\alpha$  and should lay between 0 and 1. Low alphas are more conservative. The choice of  $\alpha$  should depend on how dangerous it is to reject  $H_0$  while it is true.  $\alpha$  is normally set at 0.05 or 0.01 or 0.001. The statistical test produces a number called  $p$ -

# Regional Regression Model of Mean Annual Streamflow for Ungauged Catchments

---

value (bounded between 0 and 1), which is the probability of obtaining the data or more extreme data under the null hypothesis. More practically, the p-value will be compared to alpha.

- If p-value < alpha, reject Ho and accept alternative hypothesis (Ha) with a risk proportional to p-value of being wrong.
- If p-value > alpha, do not reject Ho, but this does not necessarily implying for it to be accepted. It either means that Ho is true, or that Ho is false but it shows that the experiment and statistical test were not “strong” enough to lead to a p-value lower than alpha.

The Mann–Kendall test statistics are given as follows (Bekele et al. 2017).

$$S = \sum_{k=1}^{n-1} \sum_{j=k+1}^n \text{sgn}(X_j - X_k)$$

S = any integer between  $-n(\frac{n-1}{2})$  and  $n(\frac{n-1}{2})$ ,  $X_j$  and  $X_k$  are sequential time series value, n is the number of data in the set,  $\text{sgn}(X_j - X_k)$  is the sign function and is given as:

$$\text{sgn}(X_j - X_k) = \begin{cases} 1 & \text{if } (X_j - X_k) > 0 \\ 0 & \text{if } (X_j - X_k) = 0 \\ -1 & \text{if } (X_j - X_k) < 0 \end{cases}$$

It is also assumed that for  $n > 8$ , the S test statistics are normally distributed with a mean value of zero and variance is calculated using the following Equation.

$$\delta^2 = \frac{n(n-1)(2n+5)}{18}$$

From this, the standardized test statistics, Z, is calculated using the following relations:

$$Z = \begin{cases} \frac{S-1}{\delta}, & \text{if } S > 0 \\ 0, & \text{if } S = 0 \\ \frac{S+1}{\delta}, & \text{if } S < 0 \end{cases}$$

## Regional Regression Model of Mean Annual Streamflow for Ungauged Catchments

The decision to either reject or accept the null hypothesis is then made by comparing the calculated Z with the critical value at a chosen level of significance. In this study, the critical value (alpha) is tested at 95% (0.05) confidence level. XLSTAT 2019 calculated the p-value and if the p-value is less than the critical value (alpha) at the chosen significance levels, Ho is rejected. Rejecting Ho indicates that “there is a trend” in the time series, while accepting Ho indicates “no trend was detected”. On rejecting the null hypothesis, the result is said to be statistically significant.

For the analysis of trends, twenty (20) rainfall stations indicate decreasing trends while only eight (8) stations showing increasing trends. The magnitude of trend was identified using Sen’s slope estimator indicating positive (increasing) and negative (decreasing) magnitudes corresponding to MK test value. Only seven (7) stations indicate a statistically significant trend with value of S and the magnitude of Sen's slope are higher than the others.

Table 4 Mann-Kendall and Sen’s slope estimator test results

Sr.no	station name	S	P-Value	MK Test trend	Test Interpretation	Sen's slope
1	Addis Ababa Bole	-82	0.209	Decrease	No trend	-3.41
2	Addis Ababa Obs	14	0.840	Increase	No trend	0.68
3	Adaiytu	-92	0.159	Decrease	No trend	-3.89
4	Akaki	-128	0.049	Decrease	There is a trend	-5.05
5	Asgori	-80	0.221	Decrease	No trend	-3.69
6	Awash melka	-46	0.486	Decrease	No trend	-1.39
7	Bologiorgis	-80	0.221	Decrease	No trend	-3.20
8	boneya	-126	0.053	Decrease	No trend	-9.42
9	busa	128	0.049	Increase	There is a trend	13.69
10	Combolcha	-52	0.429	Decrease	No trend	-2.18
11	dertuliben	-188	0.004	Decrease	There is a trend	-11.74
12	Dessie Zuria Met	34	0.609	Increase	No trend	1.65
13	Diredawa	-54	0.412	Decrease	No trend	-2.97
14	Ejere	50	0.448	Increase	No trend	1.99
15	Ginager	-40	0.546	Decrease	No trend	-3.35
16	Ginchi	-98	0.133	Decrease	No trend	-5.86
17	Guder	-22	0.745	Decrease	No trend	-2.75
18	Guranda meta	266	< 0.0001	Increase	There is a trend	24.76
19	Haik	-8	0.914	Decrease	No trend	-0.75

## Regional Regression Model of Mean Annual Streamflow for Ungauged Catchments

20	jara	174	0.007	Increase	There is a trend	12.93
21	Kimoye	2	0.988	Increase	No trend	0.58
22	Nazret	-256	< 0.0001	Decrease	There is a trend	-18.61
23	kora	60	0.361	Increase	No trend	2.99
24	Sebeta	-86	0.188	Decrease	No trend	-8.02
25	Sendafa	-22	0.745	Decrease	No trend	-1.06
26	Teji	0	1.000	Decrease	No trend	-0.02
27	Welenchit	-22	0.745	Decrease	No trend	-0.93
28	welenkomi	-204	0.002	Decrease	There is a trend	-9.65

Table 5 Selected Rainfall stations of Awash River Basin for the study

Sr.no	station name	Elevation	Longitude	Latitude	duration
1	Addis Ababa Bole	2354	38.79871	8.981081	1981-2013
2	Addis Ababa Obs	2386	38.7475	9.01891	1981-2013
3	Adaiytu	507	40.783	11.11	1981-2013
4	Akaki	2057	38.786	8.870	1981-2013
5	Asgori	2072	38.334	8.790	1981-2013
6	Awash melka	916	40.150	8.983	1981-2013
7	Bologiorgis	1963	39.359	8.836	1981-2013
8	Combolcha	1857	39.718	11.084	1981-2013
9	Diredawa	1180	42.533	9.967	1981-2013
10	Ejere	2254	39.257	8.773	1981-2013
11	Ginager	3100	39.567	9.317	1981-2013
12	Ginchi	2132	38.133	9.017	1981-2013
13	Guder	2011	39.755	8.958	1981-2013
14	Haik	1985	39.680	11.305	1981-2013
15	Kimoye	2150	38.338	9.008	1981-2013
16	Nazret	1622	39.283	8.550	1981-2013
17	Sendafa	2569	39.022	9.152	1981-2013
18	Teji	2091	38.367	8.833	1981-2013
19	Welenchit	1458	39.430	8.670	1981-2013

### 3.2.1.1.4 Monthly Rainfall Characteristics

Previous studies have attempted to characterize rainfall and streamflow in different parts of the country such as (Tadese et al. 2019) used to characterize the monthly and annual rainfall variability of a wash basin using a total of 27 metrological stations and they found that mean monthly rainfall of the stations in the Upper Awash basin varied from 6

# Regional Regression Model of Mean Annual Streamflow for Ungauged Catchments

to 223 mm in the period 1975-2004, 13 to 222 mm in the middle awash and 16 to 186 mm in the lower awash.

In this study a total of 19 rainfall stations was conducted and their monthly and annual variation was clearly explained. The mean monthly rainfall of Awash River basin for the 19 stations varied from 3.15 mm to 411.8 mm in the period of 1981-2013 G.C. comparatively the monthly rainfall was low starting from November to March and started to increase in June and reaches its peak value during July to September. The minimum rainfall was recorded in December (3.15 mm at Adaiytu) and the maximum in August (411.8 mm at Ginager).

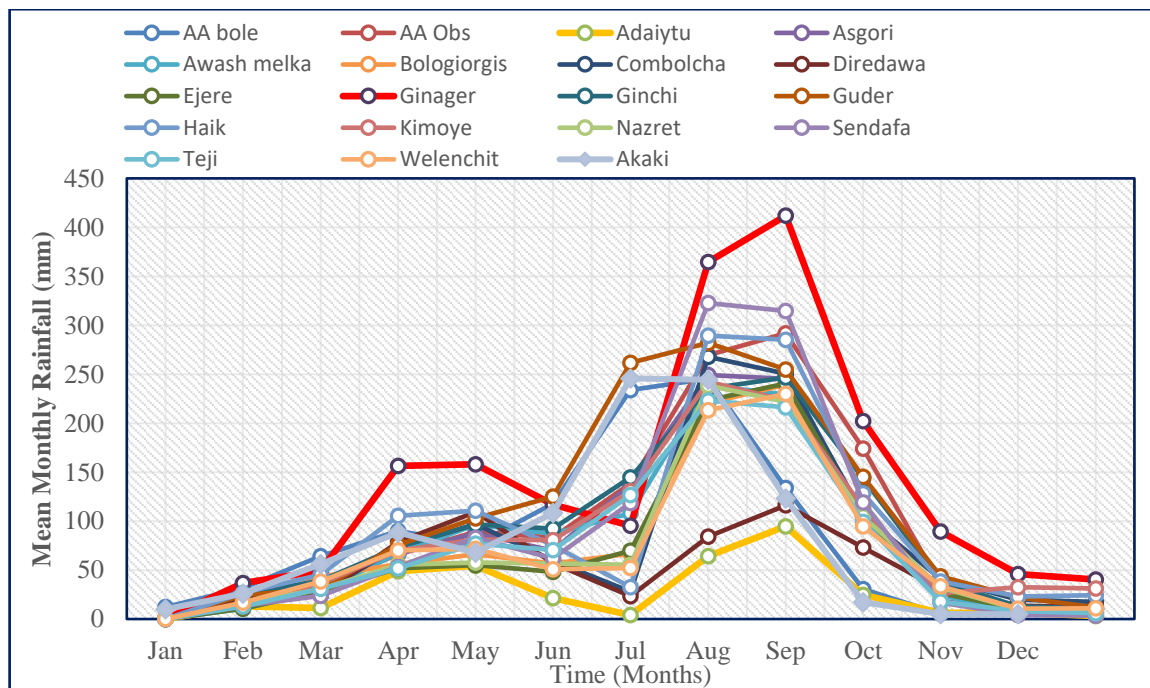


Figure 5 Monthly rainfall distribution of Awash River basin at the selected station

### 3.2.1.1.5 Annual Rainfall variability

The mean annual rainfall distribution across the 19 stations in Awash River Basin varies from 352.9 mm in Adaiytu to 1766.4 mm in Ginager. Figure 6 shows the annual rainfall in Awash River Basin of the 19 rainfall stations; the annual rainfall varied from 38.76 mm to 2650.1 mm from rainfall stations Adaiytu and Guder respectively. Diredawa and Guder rainfall stations recorded the lowest and highest annual rainfall respectively. However, many of the stations in the Basin receive less than 1100 mm rainfall per annum.

## Regional Regression Model of Mean Annual Streamflow for Ungauged Catchments

---

According Gebremichael (2014) CV is used to classify the degree of variability of rainfall events as less ( $CV < 20$ ), moderate ( $20 < CV < 30$ ), high ( $CV > 30$ ), very high  $CV > 40\%$  and  $CV > 70\%$  indicate extremely high inter-annual variability of rainfall. Based on this, from the observed data considered that all the stations having above 30 % of coefficient of variation (CV) highlighting the high variability of rainfall over the zone. The coefficient of variability (CV) of rainfall in Ethiopia varies from 10% to 70%, of which, areas with CV above 30% are vulnerable to drought (Gebremichael 2014). As shown the table below CV of the selected stations varies from 12.6 % in Addis Ababa observatory (i.e. the rainfall variability is less) to 34.5 % in Adaiytu (i.e. the rainfall variability is very high). The annual weighted average rainfall of the stations is 1028.4 mm with 9.5 % CV.

Table 6 Mean rainfall (mm) and CV (%)

Name	SD	Mean	CV
Addis Ababa Bole	149.6	1047.5	14.3
Addis Ababa Obs	153.0	1214.9	12.6
Akaki	128.0	997.1	12.8
Asgori	156.7	1043.3	15.0
Bologiorgis	152.9	913.1	16.7
Ginchi	256.2	1159.3	22.1
Ginager	292.7	1766.4	16.6
Guder	400.1	1380.1	29.0
Kimoye	286.1	1083.3	26.4
Nazeret	161.7	886.2	18.2
Sendafa (Add)	203.8	1134.2	18.0
Teji	127.8	940.2	13.6
Welenchiti	283.4	892.8	31.7
Ejere	150.1	875.2	17.1
Adaiytu	121.9	352.9	34.5
Combolcha	153.3	1013.9	15.1
Dire Dawa	151.4	642.8	23.6
Haik	212.6	1184.9	17.9
Awash Melka	294.2	1011.8	29.1
<b>Mean</b>	<b>97.6</b>	<b>1028.4</b>	<b>9.5</b>

# Regional Regression Model of Mean Annual Streamflow for Ungauged Catchments

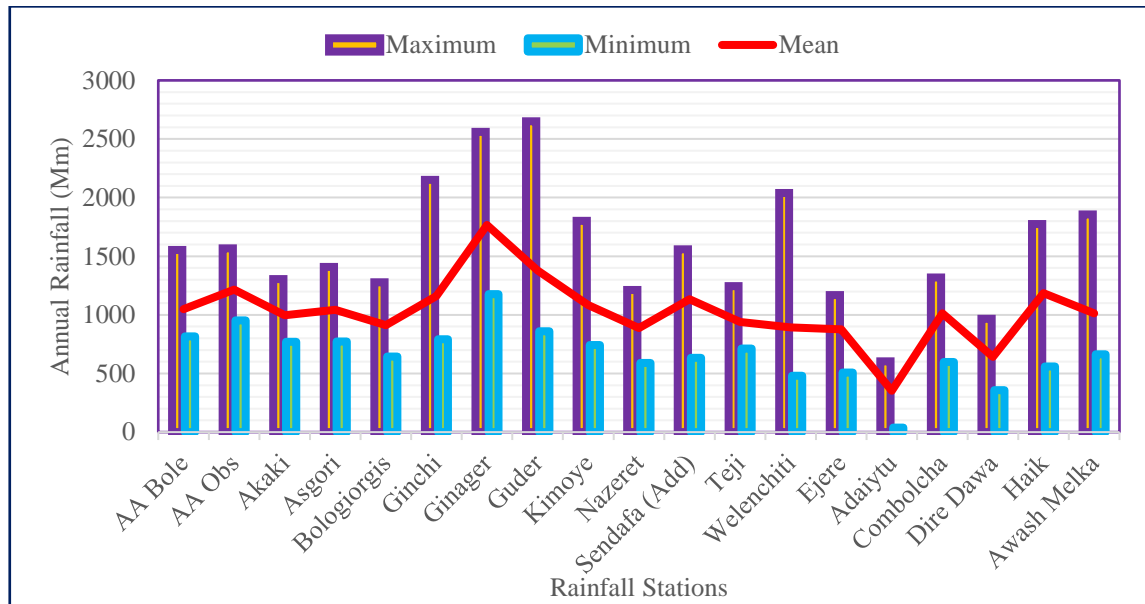


Figure 6 Annual Rainfall (mm) in Awash RB of the selected stations

### 3.2.1.1.6 Areal Rainfall

#### Ground Rainfall Data

Areal rainfall estimates are of great importance in mitigating the effects of rainfall excesses and deficits in many parts of the world (Grimes, Pardo-Igúzquiza, and Bonifacio 1999). In this study Areal rainfall estimation was used to interpolate the point rainfall values from 19 stations to areal rainfall of Awash River basin with a grid value of 4 x 4 km resolution using Geostatistical interpolation method (i.e. Ordinary Kriging).

# Regional Regression Model of Mean Annual Streamflow for Ungauged Catchments

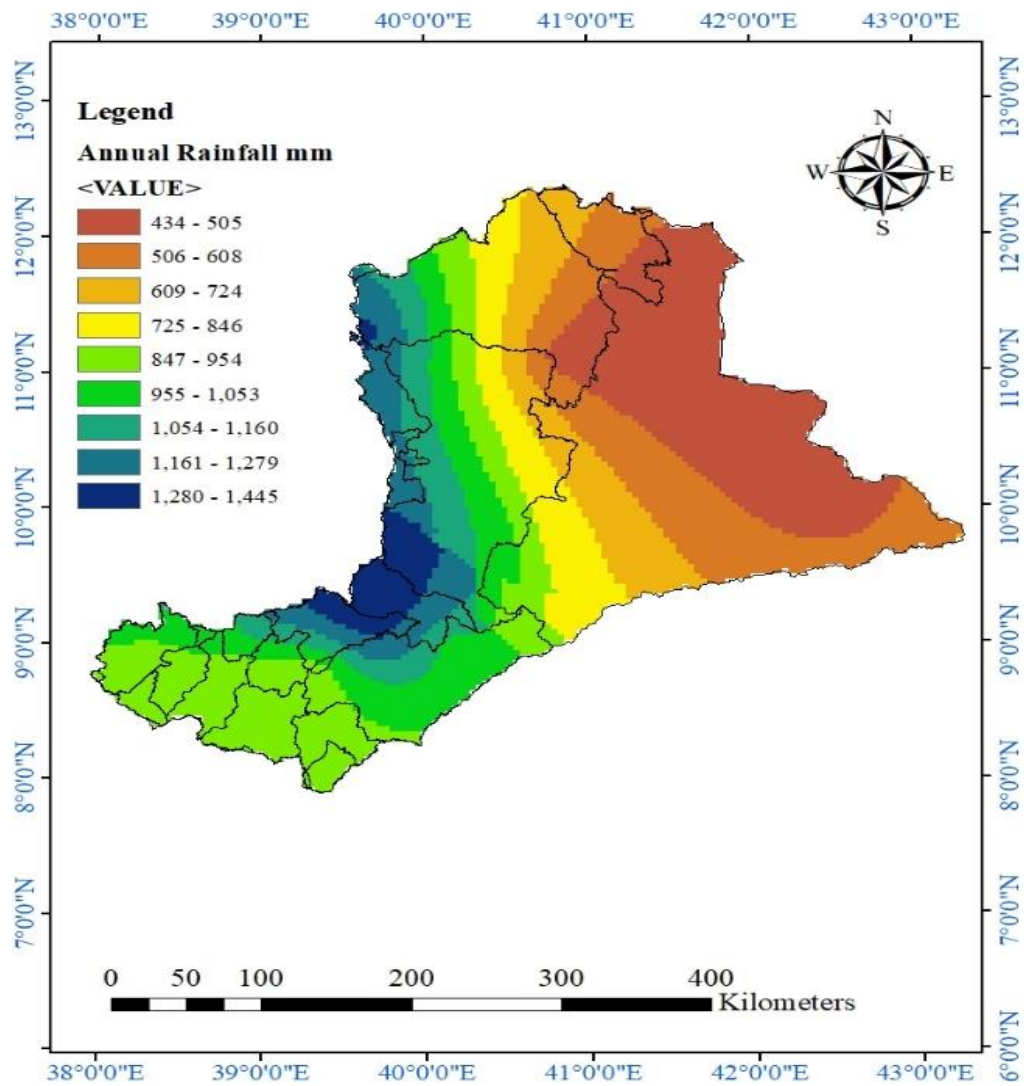


Figure 7 Kriging map showing annual areal rainfall of Awash River basin for the year 2000

# Regional Regression Model of Mean Annual Streamflow for Ungauged Catchments

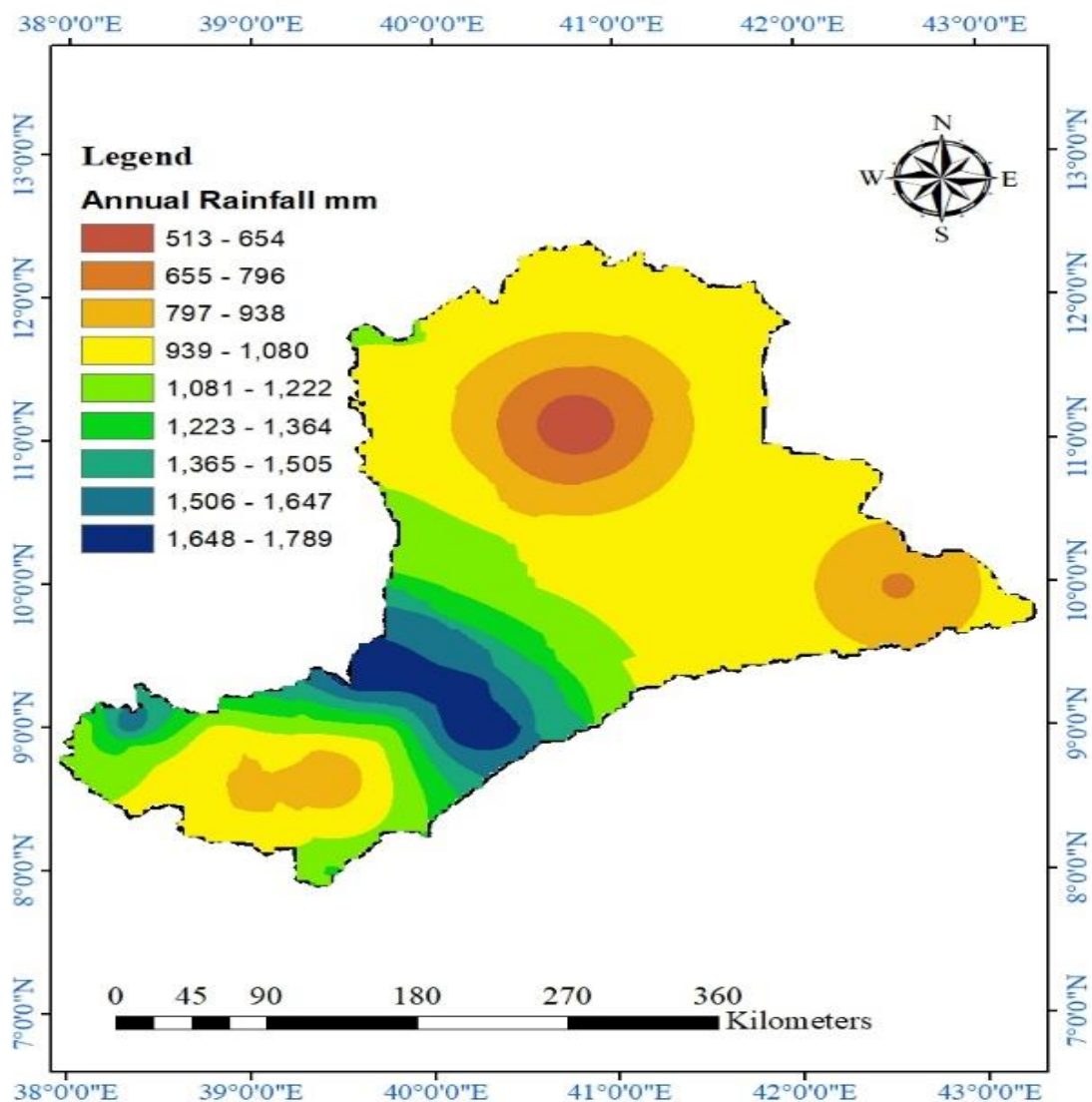


Figure 8 Kriging map showing annual areal rainfall of Awash River basin for the year 1981-2013

## 3.2.1.2 Stream flow data

### 3.2.1.2.1 Selection of hydrological gauging Stations

Long records of stream flow data are required and should be available In order to create regression models, The World Meteorological Organization recommends that climatic analysis be executed with data for 30 years (Barbarossa et al. 2017). Since there is a lack of continuously recorded river flow data in most of the river basins of Ethiopia, It was enforced to use the selected gauging stations having a recorded data of more than 10

# Regional Regression Model of Mean Annual Streamflow for Ungauged Catchments

years used for rainfall – runoff analysis of individual gauged catchments and for a basin level. Mean annual flow of gauged rivers of Awash basin was developed from the mean monthly flow obtained from MWIE for the selected gauging stations and more or less the number of available measured gauged rivers were considered and directly used as a response variable for the development of mean annual flow of ungauged catchments.

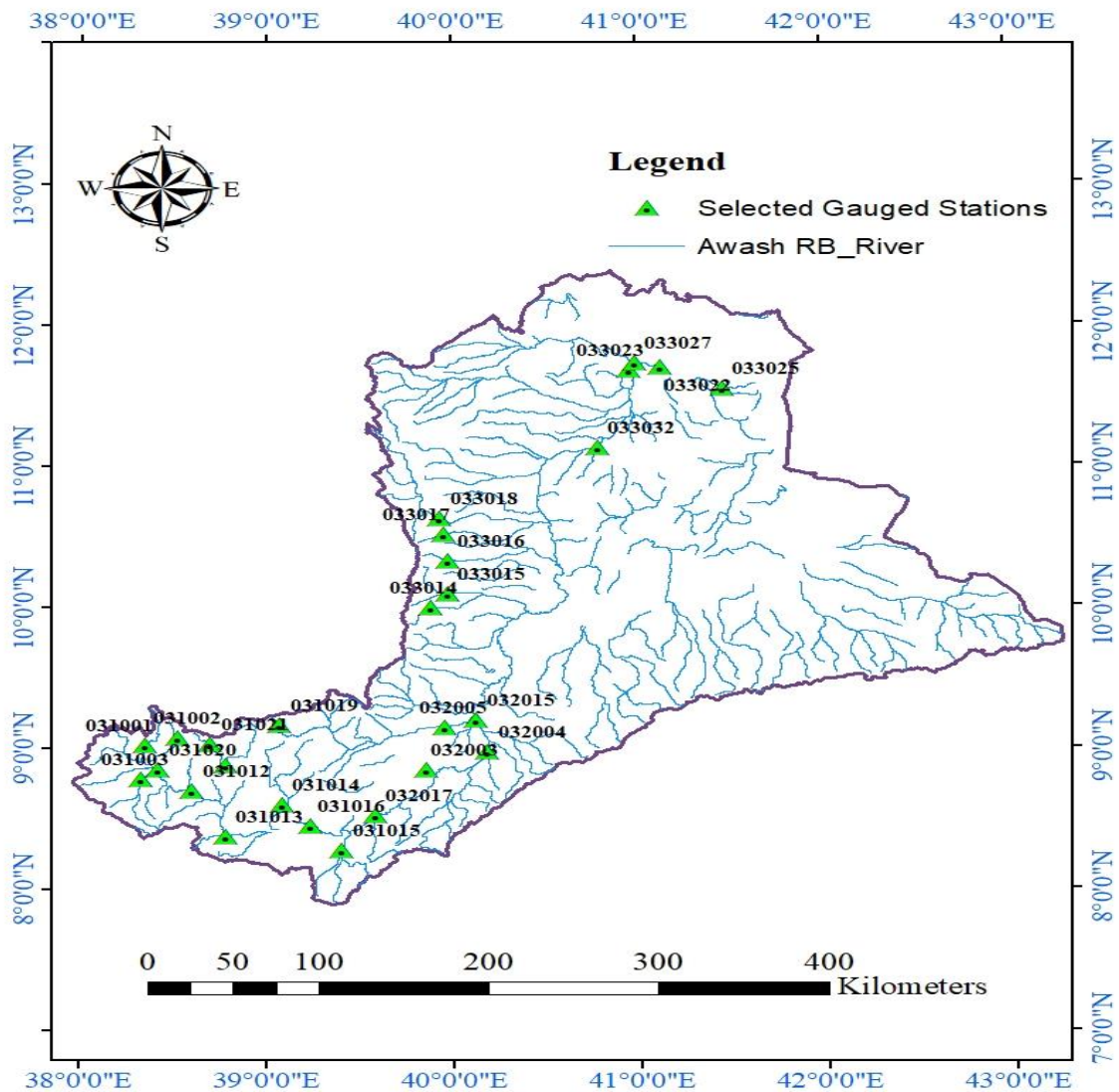


Figure 9 Map of selected gauged stations in Awash RB

From all stream flow gauging stations 13 are located on the main Awash River and the remaining 14 stations are located on awash tributaries, the following table summarizes the selected gauging stations of Awash River basin.

## Regional Regression Model of Mean Annual Streamflow for Ungauged Catchments

Table 7 the selected gauging stations of Awash RB

s.no	Code	River	Station	Catchment Area	Lat	Long	Record length
Stations on Awash mainstream							
1	033032	Awash	Nurhera	14414.38	11.13	40.78	1981-1997
2	033033	Awash	wonji	11790.25	11.22	39.62	1981-2013
3	031012	Awash	Belo	2524.14	8.70	38.60	1981-2012
4	031013	Awash	Adaiytu	50654.51	8.38	38.78	1981-2001
5	031016	Awash	Berga	69789.66	8.45	39.23	1991-2005
6	031020	Awash	Dubti	65973.02	8.85	38.42	1983-2000
7	032015	Awash	Melkahomble	7579.62	9.20	40.12	1981-2011
8	032017	Awash	Melkasedi	23447.25	8.53	39.58	1987-2008
9	033022	Awash	Melkawerer	29061.00	11.70	41.12	1981-2010
10	033025	Awash	Melkakunture	4454.12	11.55	41.45	1981-2012
11	32003	Awash	Methara	18146.83	8.85	39.85	1981-2009
12	33023	Awash	Tendaho	60825.22	11.68	40.95	1981-2002
13	32004	Awash	Awash 7 kilo	21111.03	8.98	40.18	1981-2009
Stations on Tributaries							
14	31001	Berga	Addis alem	210.77	9.02	38.35	1981-2009
15	33027	Logia	logia vilage	3095.15	11.73	40.98	1988-2001
16	33014	Robi	shewarobit	258.35	10.00	39.88	1991-2003
17	31019	Kesem	beke	43.89	9.17	39.07	1986-2006
18	033016	Ataye	Afeso	160.06	10.33	39.97	1990-2001
19	031004	Akaki	Akaki	886.89	8.88	38.78	1981-2007
20	031002	Borkena	Boru	1769.68	9.08	38.52	1997-2013
21	033017	Jara	Jara	223.18	10.52	39.95	1990-2002
22	033015	Jeweha	Jeweha	566.87	10.10	39.97	1990-2002
23	031015	Keleta	Sirie	758.92	8.28	39.40	1990-2010
24	032005	kesem	Awramelka	2993.36	9.15	39.95	1990-2008
25	031021	Little	D/s of gefersa dam	133.41	9.03	38.70	1990-2004
26	031014	Mojo	Mojo	1459.49	8.60	39.08	1981-2013
27	031003	Teji	Asgor	648.57	8.78	38.33	1981-2008

# Regional Regression Model of Mean Annual Streamflow for Ungauged Catchments

---

## 3.2.1.2.2 Stream flow Data quality

Missed data may be resulting from lack of appropriate records, shifting of station location and processing should be given attention because they lead inconsistency and ambiguous results that may contradict to the actual situation.

Before using the required data for the regression analysis, it is crucial to check for the missing data and to check the quality of the observed data. In this thesis stream flow data obtained from MWIE had missing data, but filling missed monthly flow data may encounter error for the model development, especially those recording years having missed data during good flow seasons (i.e. from June-October) were dropped out and only observed monthly stream flow data were used for the development of mean annual flow for ungauged catchments of awash RB.

## Outlier test

An outlier is an unusually small or large value in a given dataset. Outliers can have one occurred due to a recording error, an observation from a population not similar to that of most of the data such as a flood caused by a dam break or release rather than precipitation and a rare event from a single observation that is quite skewed.

The Tukey fence was used to screen the outliers greater than a threshold value that can affect the detection of inhomogeneity. The data range is represented as:  $[Q_1 - 1.5 \cdot IQR, Q_3 + 1.5 \cdot IQR]$  where  $Q_1$  and  $Q_3$  are the lower and upper quartile points, respectively and IQR is the interquartile range (Bekele et al. 2017). Values outside the Tukey fence are considered as outliers. In this study, such outliers were set to a limit value corresponding to  $1.5 \times IQR$ .

## Regional Regression Model of Mean Annual Streamflow for Ungauged Catchments

---

Table 8 Outlier check of MAF for kesem at beke station

Year	MAF (MCM)	Outlier Check	AMR (mm)	Q1	Q3	IQR	Upper Limit	Lower Limit
1986	26.12	False	1089.06	27.39	42.80	15.40	65.91	4.29
1987	9.09	False	1070.93					
1988	40.11	False	1157.73					
1989	26.78	False	1208.14					
1990	44.62	False	1444.80					
1991	57.09	False	806.19					
1992	41.85	False	1061.02					
1993	41.08	False	1279.41					
1994	76.10	True	982.22					
1995	39.10	False	1130.05					
1996	53.50	False	1485.80					
1997	19.27	False	1029.77					
1998	66.87	True	1269.38					
1999	42.20	False	1250.85					
2000	33.54	False	1182.18					
2001	27.35	False	1218.83					
2002	32.77	False	1029.84					
2003	34.49	False	1280.82					
2005	33.86	False	1239.03					
2006	27.42	False	1224.14					

Once the lower and upper quartile points of flow record is estimated lower and upper limits of the data range were calculated as shown on the above table. Then using Microsoft excel command the possibilities of flow data being an outlier was listed as either False or True (False-the given time series data is not outlier, True- the data is an outlier). Based on this information flow data displayed as an outlier (i.e. 76.10 MM<sup>3</sup> and 66.87 MM<sup>3</sup> with their corresponding rainfall value 982.22 mm and 1269.38 mm respectively) were further examine for the occurrence during those record years using the independent variable (i.e. Rainfall), based on this information these two records were detected as outliers and removed from the MAF analysis.

# Regional Regression Model of Mean Annual Streamflow for Ungauged Catchments

## 3.2.1.2.3 Monthly Streamflow Characteristics

Most of the gauged rivers in Awash River basin selected in this study have good flow seasons starting from June to October and relatively the flow becomes normal flow or base flow from November to May, the maximum stream flow were recorded in the month of august almost in most of the stream flow gauged stations. The following figures shows the monthly flow pattern on good flow seasons of the selected gauged stations.

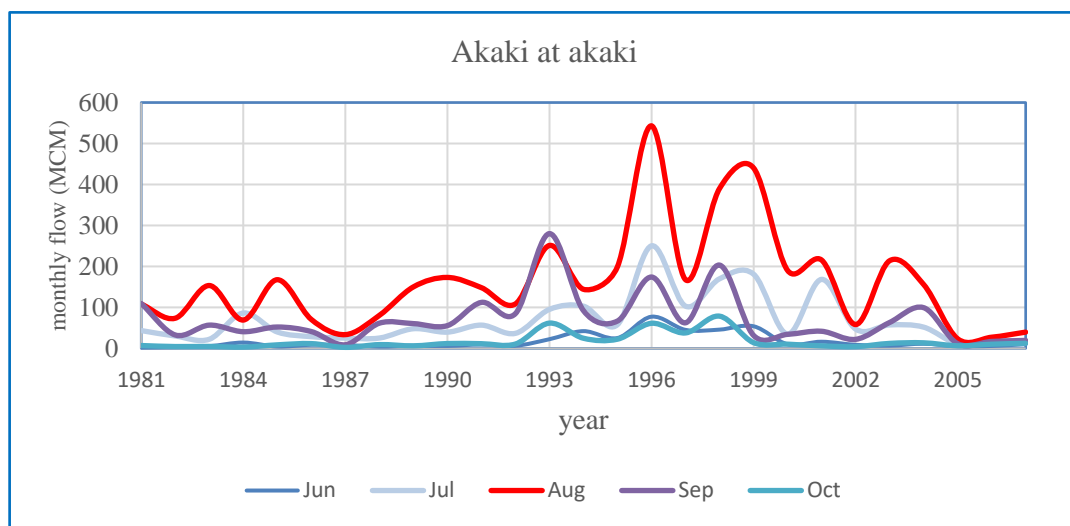


Figure 10 Monthly stream flow pattern of Akaki gauging station

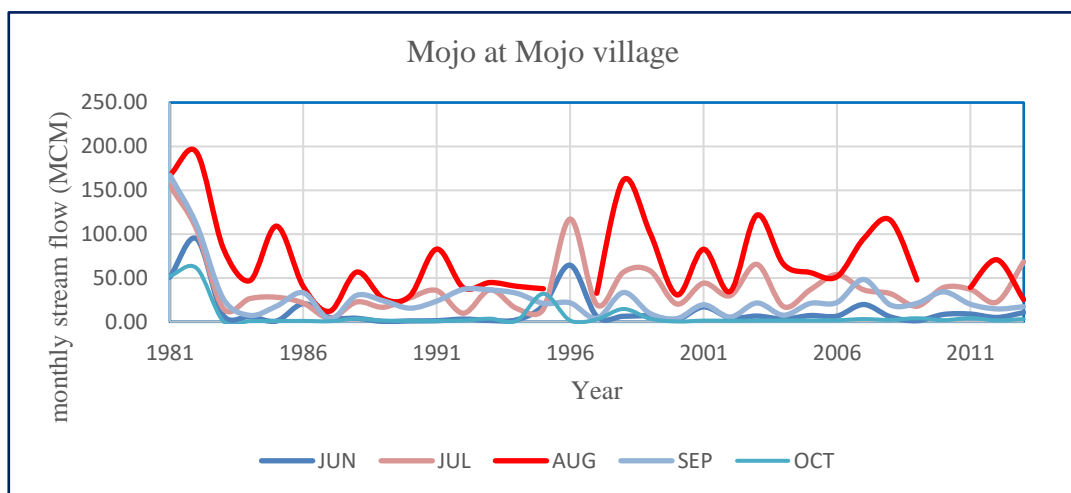


Figure 11 Monthly stream flow pattern of mojo gauging station

# Regional Regression Model of Mean Annual Streamflow for Ungauged Catchments

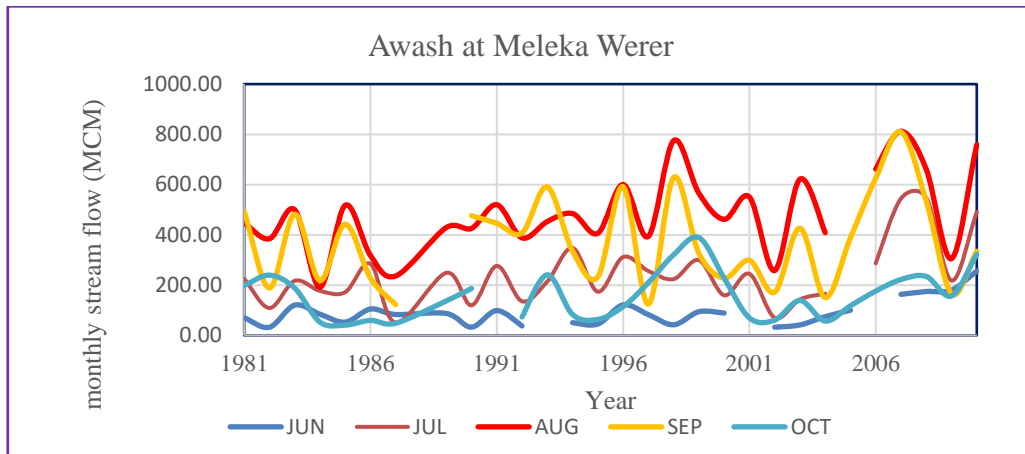


Figure 12 Monthly stream flow pattern of Awash at meleka werer gauging station

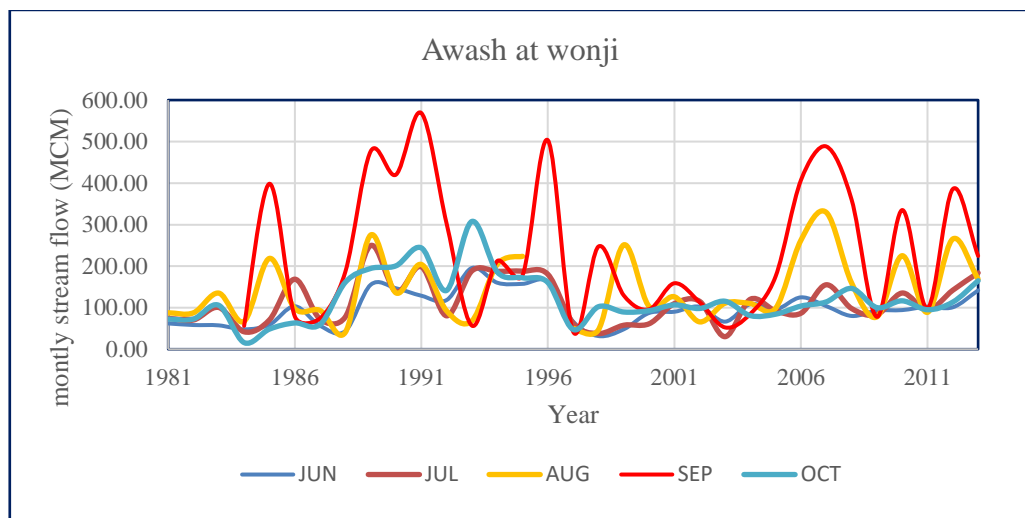


Figure 13 Monthly stream flow pattern of Awash at wonji gauging station

### 3.2.1.2.4 Annual Streamflow Characteristics

Figure 14 shows annual streamflow of 27 stations in the Awash RB. The annual flow varied from 4.592 MCM at logia to 3615.144 MCM at Awash@Melkawerer and the mean annual flow was varied from 36.68 MCM to 2026.82 MCM at kesem beke and tendaho respectively. The maximum and the minimum annual discharge was recorded at Awash@Melkawerer and Ataye Nr Afeso stations, respectively.

# Regional Regression Model of Mean Annual Streamflow for Ungauged Catchments

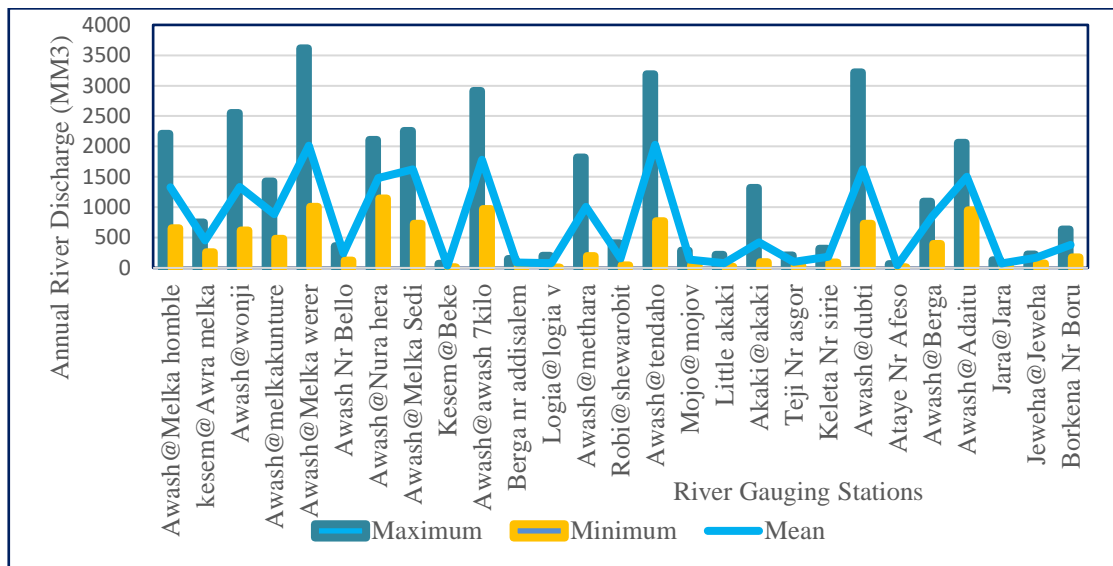


Figure 14 The annual river discharge of Awash RB at 27 stations

Table 9 Mean and CV of selected gauging stations

s.no	station name	MAF(MCM)	SD	CV (%)
1	Awash@Melka homble	1328.94	388.25	29.21
2	kesem@Awra melka	453.55	135.09	29.79
3	Awash@wonji	1333.99	532.99	39.95
4	Awash@melkakunture	882.63	246.64	27.94
5	Awash@Melka werer	2010.91	659.60	32.80
6	Awash Nr Bello	244.05	56.81	23.28
7	Awash@Nura hera	1481.23	258.72	17.47
8	Awash@Melka Sedi	1623.75	372.84	22.96
9	Mojo@mojov	135.86	56.29	41.43
10	Little akaki	79.16	57.81	73.03
11	Akaki@akaki	419.58	317.54	75.68
12	Teji Nr asgor	101.09	46.34	45.84
13	Keleta Nr sirie	181.63	72.86	40.11
14	Awash@dubti	1621.98	801.81	49.43
15	Ataye Nr Afeso	40.44	15.17	37.53
16	Awash@Berga	847.09	202.68	23.93
17	Awash@Adaitu	1499.15	373.64	24.92
18	Jara@Jara	73.82	32.22	43.64
19	Jeweha@Jeweha	168.86	40.50	23.98
20	Borkena@swampout let	379.89	159.65	42.03
21	Awash@awash 7 kilo	1779.53	511.82	28.76
22	Kesem@Beke	36.69	13.44	36.64
23	Berga Nr addis alem	91.52	28.70	31.36

## Regional Regression Model of Mean Annual Streamflow for Ungauged Catchments

---

24	Logia@logia v	75.76	60.91	80.39
25	Awash@methara	1003.83	386.76	38.53
26	Robi@shewarobit	155.47	110.36	70.99
27	Awash@tendaho	2026.82	583.47	28.79

Table 8 gives the summary of descriptive statistics of the 27 streamflow stations. The annual streamflow in the majority of the stations indicated about 3.7%, 37.03% and 59.26% of the stations showed normal ( $CV < 20$ ), moderate ( $CV$  of 20-30) and high ( $CV > 30$ ) variability of streamflow, respectively. This value of  $CV$  was seen as acceptable compared with the study conducted by (Tadese et al. 2019) in awash river basin using 18 stream flow stations, they found  $CV$  value of 5.6%, 27.8% and 66.7% of the stations showed normal ( $CV < 20$ ), moderate ( $CV$  of 20-30) and high ( $CV > 30$ ) variability of streamflow, respectively. These 27 gauging stations were collected for analysis of monthly and annual streamflow characteristics of the whole basin but for the regression analysis of ungauged catchments streamflow gauging stations located below koka dam were removed, since the release of water from koka reservoir during high rainfall seasons may directly disturb on the natural measurement of those gauges and this may intern results uncertainties on estimation of mean annual stream flow of ungauged catchments of awash river basin. Therefore a total of 9 gauging stations located on the upper awash and 8 stations found on middle and lower awash were used to formulate a regression analysis, this was discussed in detail on section 3.3.2.

### 3.2.1.3 Physiographical data

Physiographic data describes the physical characteristics of the earth's surface. In this thesis these datasets were derived from DEM (Digital Elevation Model), and derivatives of this are catchment area, mean elevation and mean slope. These datasets are illustrated with short explanation as shown below.

#### 3.2.1.3.1 Catchment area

The catchment area of the 27 gauged stations at Awash River basin were delineated by using the spatial analyst tools, hydrology in ArcMap 10.4 and the values were compared with the past recorded values at awash master plan obtained from MWRE. The delineated catchment area size ranges from 43 km<sup>2</sup> to 69789 km<sup>2</sup>.

# Regional Regression Model of Mean Annual Streamflow for Ungauged Catchments

## 3.2.1.3.2 Mean slope

Mean slope value of the selected gauged stream flow stations of Awash River basin was computed using 30 m resolution of DEM in ArcMap 10.4 and the value is in percentage as shown in the figure below. Mean slope of each gauged station was used as an independent/predictor variable in this study.

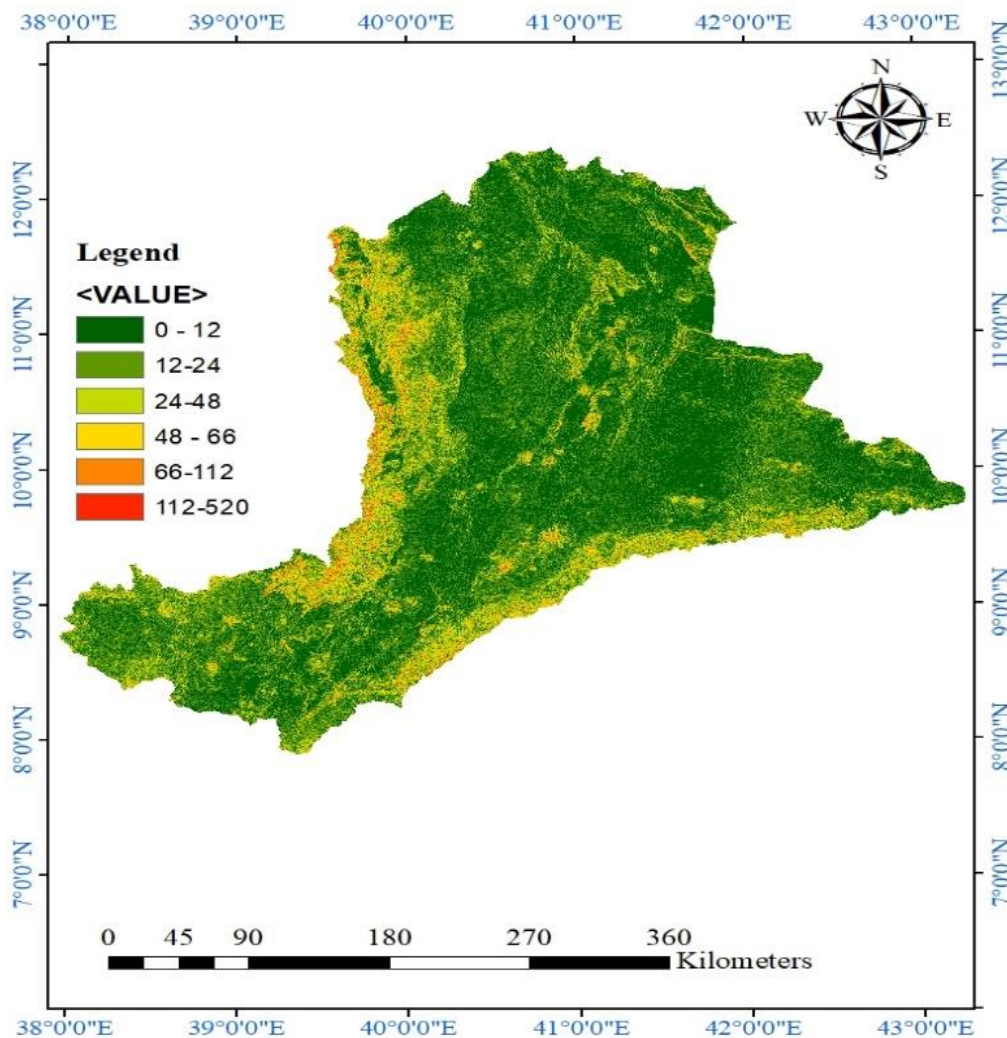


Figure 15 Slope map of Awash river basin

## 3.2.1.3.3 Mean elevation

Mean elevation for each delineated catchment area was simply estimated from DEM using ArcMap and used as an independent variable/predictor variable in this thesis.

# Regional Regression Model of Mean Annual Streamflow for Ungauged Catchments

---

## **3.3 Method**

### **3.3.1 Software used**

In this thesis most of the work was done by using ArcGIS version 10.4, Microsoft Excel, R version 3.5.1 and XLSTAT in order to process data, calculate necessary variable values and perform the regression analysis. These are introduced below with a short explanation.

#### **ArcGIS Version 10.4**

ArcGIS is an architecture geographic information system (GIS) for working with maps and geographic information. It is used for creating and using maps, compiling geographic data, analyzing mapped information, sharing and discovering geographic information, using maps and geographic information in a range of applications, and managing geographic information in a database. ArcGIS is a package of products such as ArcMap, ArcCatalog and ArcToolbox developed by ESRI (Environmental Systems Research Institute, Inc.). It is a software that allows for management of spatial data, GIS-data and map analysis and cartography. In this study ArcGIS version 10.4 was used for watershed delineation, Areal rainfall estimation, extraction of satellite rainfall data and preparation of maps.

#### **R version 3.5.1**

R is a free software environment for statistical computing and graphics. R provides a wide variety of statistical (linear and nonlinear modelling, classical statistical tests, time-series analysis, classification, clustering etc....) and graphical techniques, and is highly extensible. In this study version 3.5.1 was used for computation of regression model and to compute statistical tests. R-studio was the software used as development environment.

#### **XLSTAT version 2019**

XLSTAT is the leading data analysis and statistical solution for Microsoft Excel developed by Addinsoft Company, is a powerful yet flexible Excel data analysis add-on that allows users to analyze, customize and share results within Microsoft Excel. In this

# Regional Regression Model of Mean Annual Streamflow for Ungauged Catchments

---

thesis XLSTAT version 2019 was used for homogeneity and trend analysis test of ground recorded rainfall data.

## 3.3.2 Method of analysis

### 3.3.2.1 Selection of climatological and physiographic variables

Climatological and physiographic variables were selected according to what will be considered most likely to describe mean annual flow. It is reasonable that area have a major impact in the regression models to develop regionalization model for ungauged catchment. Other geomorphological variables like mean elevation and mean slope. Furthermore, climatic variables, such as mean precipitation. Mean elevation is considered as a predictor variable instead of adding mean annual temperature as explanatory variable because mean elevation can indicate the characteristics of temperature. Generally the following variables were included in the development of mean annual flow these are, Catchment Area, mean annual precipitation, mean elevation and mean basin slope are considered as explanatory or predictive.

Table 10 List of variables used as predictors in the model

Variable Name	Definition	Unit	Source	Record Length
<b>Physiographic variables</b>				
Area	Catchment Area	Km <sup>2</sup>	DEM	
S	Mean Slope	% (percentage)	DEM	
E	Mean Elevation	meters	DEM	
<b>Climatic Variables</b>				
P	Mean Annual Rainfall	mm	NMA	1981-2013

### 3.3.2.2 Classification of Regions

According to Belete B. et al, (2014) the unimodal and bimodal rainfall systems are the base for classifying the country into three rainfall regimes Commonly, these rainfall regimes are named as Regime A, Regime B, and Regime C. Based on this study most of the middle and lower awash parts are classified as regime A that have a bimodal rain fall distribution and the upper awash is unimodal. In this study awash river basin was classified in to two regions based on the similarity of rainfall distributions, region-I as shown on figure 17 including the upper awash part (i.e. Awash Uplands) and Region-II

# Regional Regression Model of Mean Annual Streamflow for Ungauged Catchments

including most of the middle and lower part of awash river basin covering mainly gauging stations located on awash western highlands (Robit at Robit Village, Jeweha at Jeweha, Ataye Near Afeso, Jara at Jara and Borkena Downstream of Swamp), Awash Upper Valley Sub basin (Keleta Near Sirie and Kesem at Awramelka) and Awash Lower Valley (Logia at Logia Village) but most of the gauging stations located on Awash middle Valley, Awash Lower Plains and Awash Eastern Catchments were not included in this region. Because most of the stations found in Awash middle valley are located on the main awash river, these gauging stations are affected by the release of water from koka reservoir and other stations found on Awash Lower Plains and Awash Eastern Catchments have either short period of records or not significant in the development of regression analysis since the area is not well developed and utilized compared to the upper and western highlands of the basin.

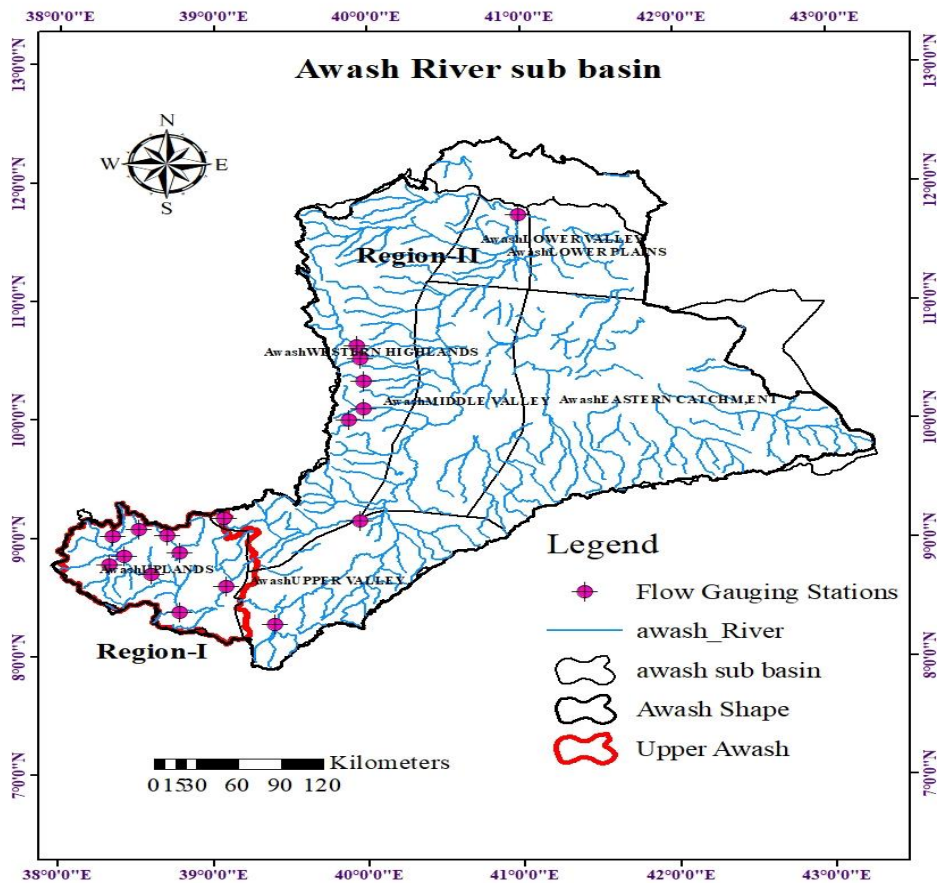


Figure 16 Map showing the regions of the study area

# Regional Regression Model of Mean Annual Streamflow for Ungauged Catchments

As shown on (figure 16) the upper awash represented as region-I and most of the middle and lower awash referred as region-II based on the rainfall distribution of the basin and this classification is used in this study for obtaining reasonable regression equation since these two regions have different rainfall characteristics.

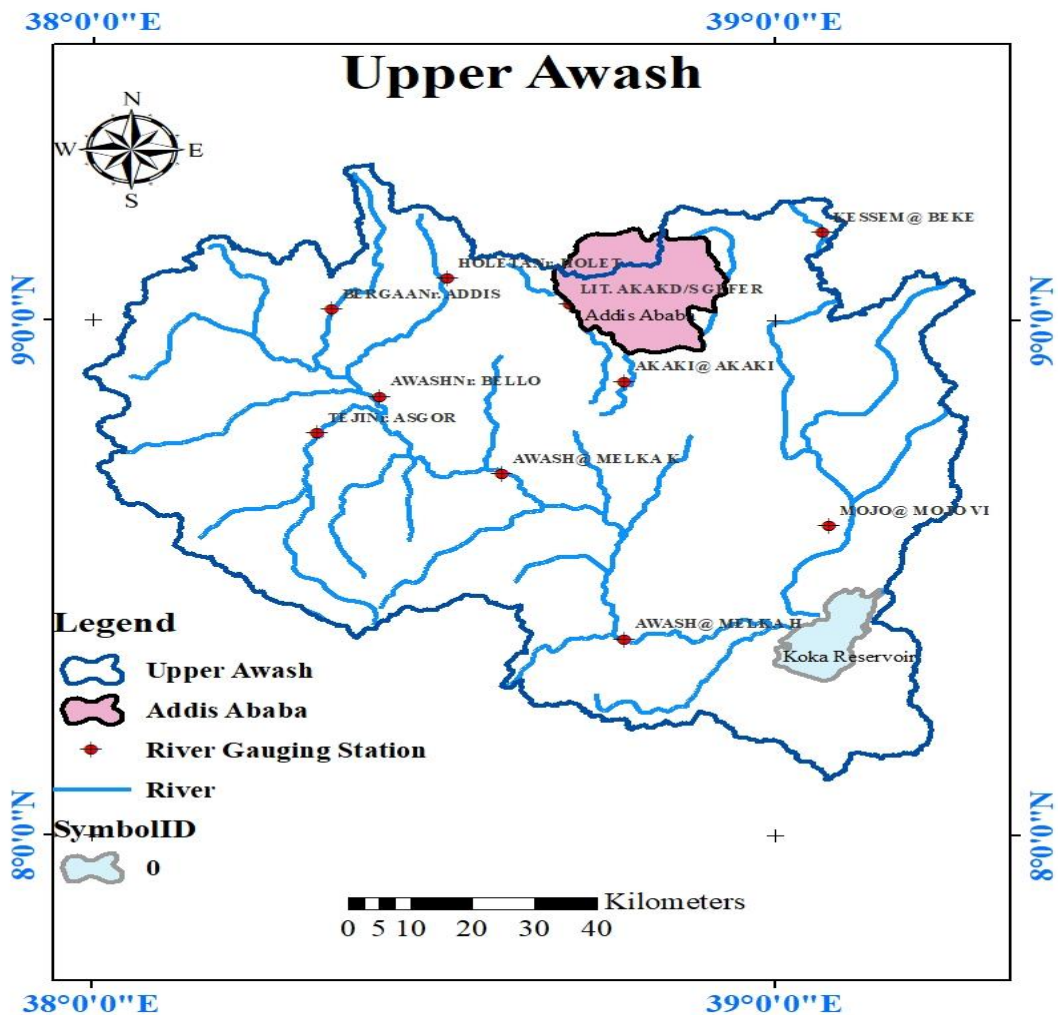


Figure 17 Figure Map of upper Awash River basin

As shown on (Figure 17) streamflow gauging stations located on awash uplands or upper awash grouped as Region-I consists of Awash at Melkahomble, Mojo at Mojo Town, Awash at Melkakunture, Teji Near Asgor, Awash Near Belo, Berga Near Addis Alem, Kesem at Beke, Little Akaki and Akaki at Akaki.

# Regional Regression Model of Mean Annual Streamflow for Ungauged Catchments

### 3.3.3 Computing the Regression Analysis

After calculating values of all the selected variables the regression analysis were performed using R programming language in the software R-Studio and Microsoft excel.

Table 11 Gauged stations used for training the model of Region-I

s.no	Station name	MAF(Mm3)	Area (km <sup>2</sup> )	MAP (mm)	elevation (m)	Slope (%)
1	Awash@Melka homble	1328.94	7579.62	1040.21	2293.16	13.73
2	Awash@melkakunture	882.63	4454.12	1025.27	2310.30	13.25
3	Awash Nr Bello	244.05	2524.14	1041.49	2357.17	13.70
4	Mojo@mojov	135.86	1459.49	1011.69	2146.76	10.86
5	Little akaki	79.16	133.41	1053.06	2653.92	18.60
6	Akaki@akaki	419.58	886.89	1094.68	2489.17	15.68
7	Teji Nr asgor	101.09	648.57	1024.50	2380.53	13.60
8	Kesem@Beke	36.69	43.89	1172.01	2860.18	20.63
9	Berga Nr addis alem	91.52	210.77	1063.20	2556.06	18.44

As shown on the above table the selected datasets for training the model of region-I that are located on upper awash sub basin which is upstream of koka dam. From the total of 9 observations having 5 variables (1 dependent variable and 4 independent variables).

Table 12 Gauged stations used for training the model of Region-II

s.no	station name	MAF (MCM)	Area (km <sup>2</sup> )	MAR (mm)	Elevation(m)	Slope (%)
1	kesem@Awra melka	453.55	2993.36	1245.68	2064.64	23.07
2	Keleta Near sirie	181.63	758.92	962.85	2702.16	15.89
3	Ataye Near Afeso	40.44	160.06	1107.30	1840.51	30.69
4	Jara@Jara	73.82	223.18	1091.54	2072.63	31.22
5	Borkena@swampout let	379.89	1769.68	1053.21	1996.73	25.89
6	Logia@logia v	75.76	3095.15	894.91	780.45	7.66
7	Robi@shewarobit	155.47	258.35	1266.20	1905.15	30.57
8	Jeweha@Jeweha	168.86	566.87	925.02	1949.67	31.86

Table 12 shows the selected gauging stations for Region-II that were computed from the total of 8 observations having 5 variables (1 dependent variable and 4 independent

## Regional Regression Model of Mean Annual Streamflow for Ungauged Catchments

---

variables). The selected gauging stations used in this Region were stations below koka dam and all stations located on the main awash river were removed because of the release of water from koka reservoir during rainy seasons may affect the natural flow , as a result the measurement from these stations were not be used for the development of MLR model. Therefore only 8 gauging stations from tributaries located in middle and lower awash sub basins were considered for the regression analysis of Region-II.

### 3.3.3.1 Test correlation between variables

After dividing the whole Awash River basin into Region-I and Region-II based on the rainfall distribution of the basin the correlation between the training data variables was tested. The test was computed using R programing language, that was used in order to identify which predictor variables were most likely appropriate to give a good fit of the regression model and to remove independent variables which showed correlation between each other and could therefore not be approved as predictor variables.

After performing the correlation between variables, identifying the more appropriate predictor variables was done by physically looking at the scatter plots of the response variable Q and each of the predictor variables. If there seemed to be a linear correlation between Q and any predictor variable, then that variable was used in training the regression model. In order to choose the given variable as an independent predictor variable the correlation coefficient between two variables should be lie between -1 and 1. Figure 18 below illustrates values of a correlation of the variables. It can be seen that there is a clear linear correlation between Q (MAF) and A (Area) (correlation coefficient = 0.90) and also between Q and mean slope (correlation coefficient = -0.44). It is also clear that rainfall (P) and mean Elevation are closely correlated (correlation coefficient = 0.85). This can be easily identified both from observing the scatter plots as shown on (figure 20) and also by their calculated correlation coefficients as shown on (figure 18).

## Regional Regression Model of Mean Annual Streamflow for Ungauged Catchments

```
> round(cor(upperawash),2)
      Q      A      P      E      S
Q  1.00  0.90 -0.44 -0.62 -0.53
A  0.90  1.00 -0.69 -0.86 -0.81
P -0.44 -0.69  1.00  0.85  0.79
E -0.62 -0.86  0.85  1.00  0.97
S -0.53 -0.81  0.79  0.97  1.00
> panel.cor<-function(x,y,digits=2,prefix="",cex.cor,...)
```

Figure 18 Test of correlation between variables for Region-I

As shown on the above figure the correlation coefficient of mean annual rainfall and mean elevation is very high which is 0.85 (>0.8), between mean slope and mean Elevation (0.97) therefore further test should be conducted in order to be more comfortable with the variables and variance inflation factor and tolerance were conducted using R statistical software using the following algorithm. One of the important assumptions of the classical linear regression model is that there is no exact collinearity among the predictor variables otherwise, the issue is referred to as multicollinearity (Imdadullah, Aslam, and Altaf 2016). According to this study VIF and Tolerance can be estimated as;  $VIF=1/1-R^2$  and  $TOL=1-R^2$ .

```
# Check multicollinerity using VIF and Tolerance
library(car)
vif(upperawash)
write.table(vif(upperawash), "vif.csv", col.names=TRUE, sep=",")
tolerance<-1/(vif(upperawash))
```

Figure 19 The R script used to check multicollinearity using VIF and Tolerance

Table 13 VIF and Tolerance values between predictor/Explanatory variables

Explanatory Variables	R <sup>2</sup>	VIF	TOL
A, P	0.48	1.91	0.52
A, E	0.74	3.84	0.26
A, S	0.66	2.91	0.34
P, E	0.72	3.60	0.28
P, S	0.62	2.66	0.38
E, S	0.94	16.92	0.06

# Regional Regression Model of Mean Annual Streamflow for Ungauged Catchments

According to Imdadullah et al., (2016 ) if the values of VIF is less than 10 and the value for tolerance between each predictor variables is greater than 0.1 there is no multicollinerty between independent variables. As shown on (table 13) these values were under the specified ranges except for mean Elevation, since mean elevation was highly correlated both with mean annual rainfall and mean slope, it should be discarded from the analysis and only catchment area, mean annual rainfall and mean slope were used as a predictor variable for the development of MLR for Region-I. On the other hand test of corelation for variables of Region-II is shown on (appendix D) and all the independent variables were used for the formulation of the MLR analysis.

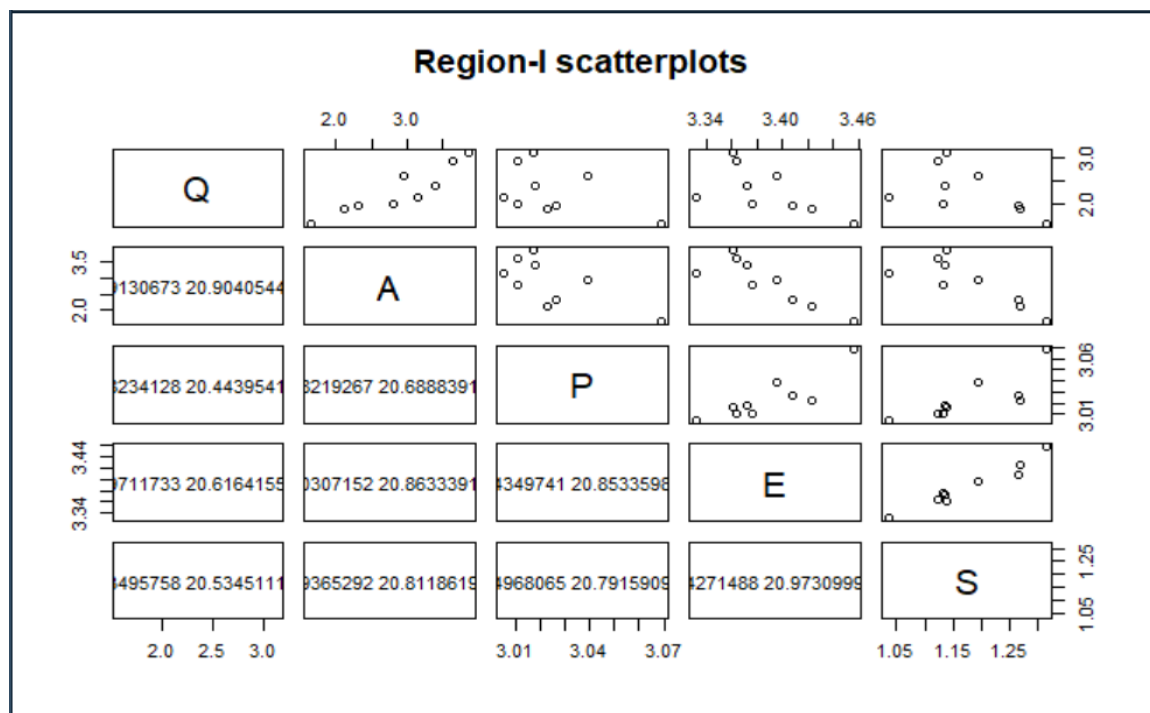


Figure 20 Scatter plots of the training variables for Region-I

As clearly shown from the above scatter plot catchment area, mean Elevation and mean slope are linearly related with mean annual flow but mean annual rainfall does not have linear relationship with mean annual streamflow.

### 3.3.3.2 Model specification

In this thesis work an assumption was made that the response variable (mean annual stream flow) could be related to the predictor variables by a linear model since the

# Regional Regression Model of Mean Annual Streamflow for Ungauged Catchments

---

predictor variables are more than one, the multiple linear regression (MLR) model were used to estimate the mean annual flow of ungauged catchments in the study area.

### 3.3.3.3 Estimation of model parameters

Regression analysis was performed for the study area with the method stepwise regression, several regressions were performed using different variables. They were performed in Microsoft excel and R using the function *lm* which fits a linear model to the data set using the OLS-method. The following are some of the different models performed by the regression analysis and the sample R script adopted to formulate these models is shown on appendix D.

- **Regression using all predictor variables.** Including all predictor variables was the first step to develop the regression analysis since the MLR predictability will increase as multiple independent variables are considered. The regression models would improve the model and analysis was performed using mean catchment area, mean annual rainfall, mean elevation and mean slope of each regions.
- **Regression using most suitable variables.** This was done by testing combinations of the different variables based on the assumptions that made the multiple linear regression to be applicable for prediction purpose. The process was continue until the best possible model was reached.

### 3.3.4 Model performance and Validation

There are three types of quantitative statistical model evaluation methods as stated on section 2.5.5 the first is Standard Regression Statistics that Determines strength of linear relationship (e.g.,  $R^2$ ). The second is Error Index that quantifies deviation in observation units (e.g., RMSE) and the third one is Dimensionless which indicates relative model evaluation (e.g., NSE). In this thesis the model performance was evaluated by using different statistical measures like  $R^2$ , NSE and RMSE for each model combinations as well as the p-value was assessed during training the MLR model and predictor variables having P-value below the specifications were carefully analyzed unless if the value is highly significant the model is appropriate to meet the objective of the study. On the other hand a plot of observed mean annual flow against the predicted values were illustrated the degree to which the points match the linear line (denoted as the 1:1 line).

## Regional Regression Model of Mean Annual Streamflow for Ungauged Catchments

---

This scatter plot allows for a visual inspection of model performance, such that the higher the agreement between estimated and observed values is the more the scatters tend to concentrate close to the 1:1 line (Ritter and Muñoz-Carpena 2013).

Model fit was done using the trained data sets obtained from each regions and different statistical measures was conducted as shown on the result and discussion part. RMSE evaluate how closely the predictions match the observations by judging the best model based on the relatively small RMSE values. It should be noted that  $R^2_{adj}$ , unlike  $R^2$ , does not show measure of fit. Rather,  $R^2_{adj}$  is useful for selecting the variables to be included in a MLR model (Maity, n.d.). These statistical model evaluation measures and acceptable model errors used in this study to validate the performance of trained MLR models are clearly stated on section 2.6.5 of the literature part.

# Regional Regression Model of Mean Annual Streamflow for Ungauged Catchments

## 4 RESULTS AND DISCUSSIONS

### 4.1 Estimation of Annual flow for the selected catchments

An equation was developed between annual flow and annual precipitation for the selected gauged stations using linear, power and polynomial functions and the coefficients of the predictor variable were optimized by excel solver data analysis. Form of equation that shows a better  $R^2$ , NSE and RMSE value was selected as an equation for annual flow of gauging stations, based on this the following 8 stations out of 17 stations satisfy the relationship between Q (annual flow in MCM) and P (annual rainfall in mm) showing that the mean annual flow has a nonlinear approach with mean annual rainfall. The remaining stations were not significantly approved as the model performance index  $R^2 < 0.5$  ( $R^2 > 0.5$ , the model is acceptable) and the result is put on appendix C.

Table 14 Summary of Annual Streamflow equation for the selected Gauged catchments

S.no	Station Name	Trained functions	Equation	Model Performance		
				$R^2$	NSE	RMSE
1	Awash at Melka homble	power	$Q=6.258*10^{-6}P^{2.07}$	0.64	0.63	18.40
2	kesem at Awramelka	power	$Q =0.13P^{1.15}$	0.54	0.56	64.74
3	Awash at melkakunture	polynomial	$Q=0.00073P^2+39.32$	0.54	0.54	146.41
4	Awash at Bello	power	$Q=2.18P^{0.7}$	0.67	0.67	30.60
5	Mojo at mojo v	power	$Q=2.45*10^{-10}P^{3.91}$	0.62	0.54	38.03
6	Little akaki	power	$Q=1.31*10^{-10}P^{3.87}$	0.52	0.56	34.89
7	Berga Near Adisalem	power	$Q=0.0001P^{1.97}$	0.72	0.70	14.80
8	Kesem at Beke	power	$Q=0.00012P^{1.77}$	0.56	0.55	4.35

# Regional Regression Model of Mean Annual Streamflow for Ungauged Catchments

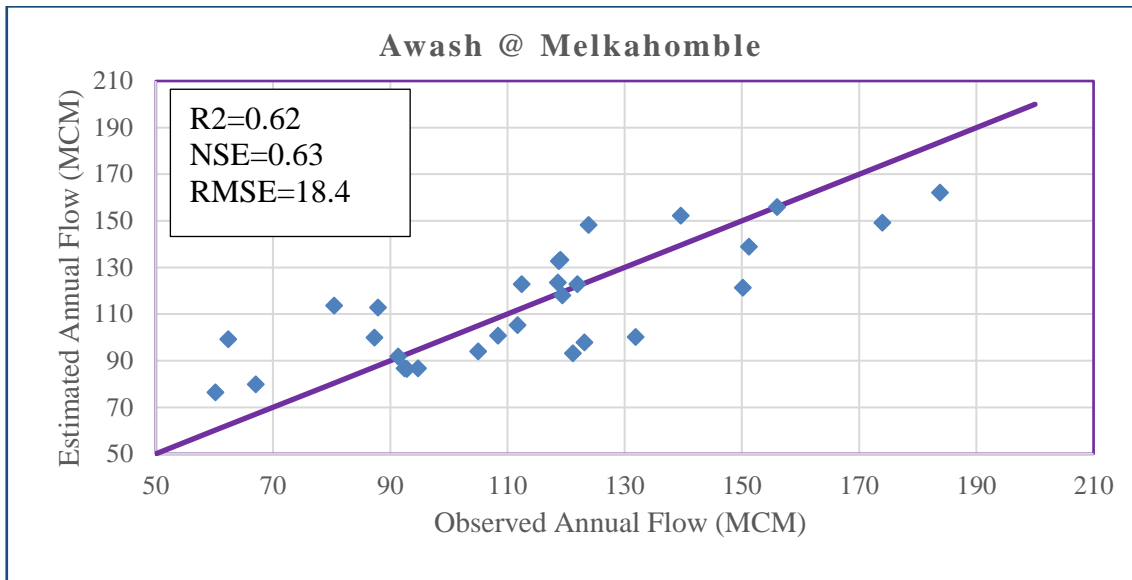


Figure 21 Plot of observed and estimated annual flow awash at melkahomble gauging station

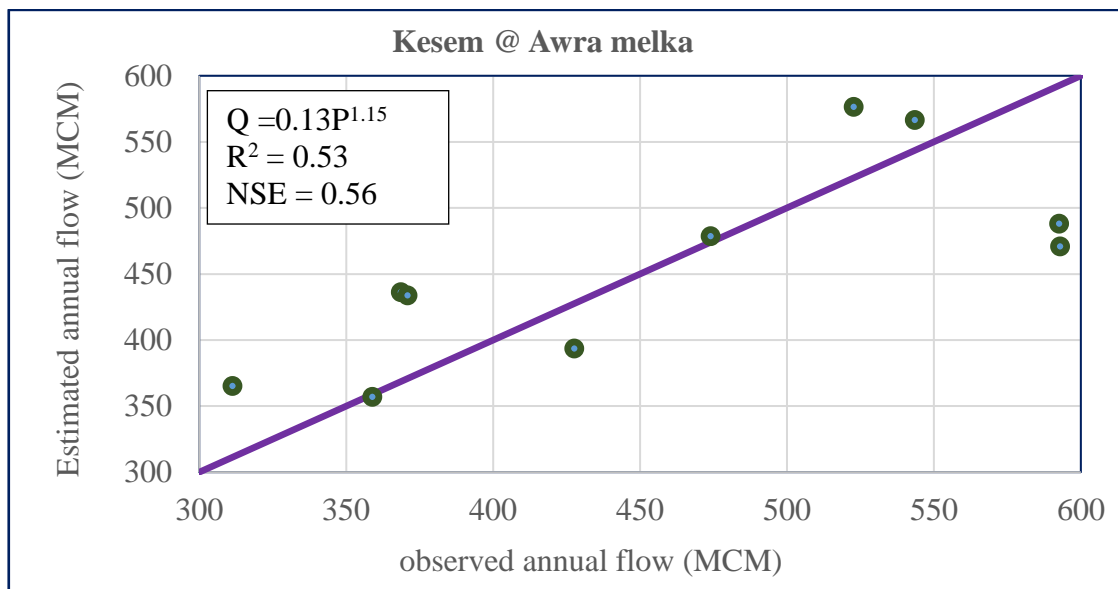


Figure 22 Plot of observed and estimated annual flow for kesem at awramelka gauging station

## Regional Regression Model of Mean Annual Streamflow for Ungauged Catchments

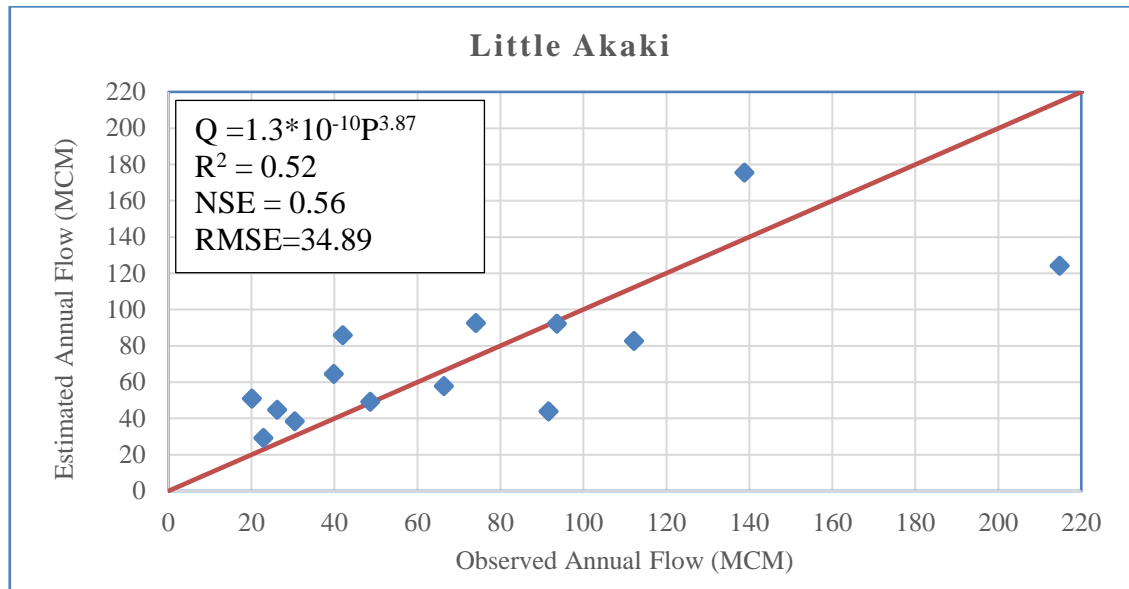


Figure 23 Plot of observed and estimated annual flow for little akaki gauging station

The above figures summarize how the points of observed and estimated annual flow fits to the 1:1 straight line, the more the points tend to this line the chosen equation is better represent the relationship between annual streamflow and annual rainfall variables.

# Regional Regression Model of Mean Annual Streamflow for Ungauged Catchments

## 4.2 Estimation of runoff coefficient for the gauged stations

Table 15 Value of runoff coefficients for the streamflow gauged stations

s.no	station name	MAF(MCM)	Area (km <sup>2</sup> )	MAP (mm)	C
1	Awash@Melka homble	1328.94	7579.62	1040.21	0.17
2	Awash@melkakunture	882.63	4454.12	1025.27	0.19
3	Awash Nr Bello	244.05	2524.14	1041.49	0.09
4	Mojo@mojov	135.86	1459.49	1011.69	0.09
5	Little akaki	79.16	133.41	1053.06	0.56
6	Akaki@akaki	419.58	886.89	1094.68	0.43
7	Teji Nr asgor	101.09	648.57	1024.50	0.15
8	Jeweha@Jeweha	168.86	566.87	925.02	0.32
9	Kesem@Beke	36.69	43.89	1172.01	0.71
10	Berga Nr addis alem	91.52	210.77	1063.20	0.41
11	kesem@Awra melka	453.55	2993.36	1245.68	0.12
12	Keleta Nr sirie	181.63	758.92	962.85	0.25
13	Ataye Nr Afeso	40.44	160.06	1107.30	0.23
14	Jara@Jara	73.82	223.18	1091.54	0.30
15	Borkena@swampout let	379.89	1769.68	1053.21	0.20
16	Logia@logia v	75.76	3095.15	894.91	0.03
17	Robi@shewarobit	155.47	258.35	1266.20	0.48

The runoff coefficient (C) is a dimensionless coefficient relating the amount of runoff to the amount of precipitation received. It is a larger value for areas with low infiltration and high runoff (pavement, steep gradient), and lower for permeable, well vegetated areas (forest, flat land). It is important for flood control, channel construction and for possible flood zone hazard delineation, As shown on the above table runoff coefficient of each station was computed using observed Mean annual flow, mean annual precipitation and catchment area, the result was in the acceptable range (<0.6) except the station kesem at beke. Catchment areas having relatively large runoff coefficient such as little akaki, akaki and shewarobit may generate a large volume of water during rainy season that may produce flooding, on the other hand catchment areas with small amount of runoff coefficient will generate relatively low flows, since this coefficient is depends on soil type, land use and land cover and other physiographic and climatological factors a detail knowledge about the catchment is needed to fully understood.

# Regional Regression Model of Mean Annual Streamflow for Ungauged Catchments

## 4.3 Regression Analysis of mean annual flow for Region-I

### 4.3.1 Results for Regression using all predictor variables

In this model, as shown on (figure 24) it can be seen that t-statistic of predictor variables mean annual rainfall is above 0.05, which is insignificant. This means that this predictor variable is not significantly related to the outcome variable or response variables statistically but those variables are highly significant in producing streamflow physically. Based on the associated t-statistic p-values Catchment area and mean slope will be used as the most appropriate predictor variables for estimation of mean annual flow for Region-I.

```
> summary(results)
Call:
lm(formula = Q ~ ., data = upperawash)

Residuals:
    1      2      3      4      5      6
-0.0001667  0.1122446 -0.2687348  0.0437753  0.0852112  0.1716435
    7      8      9
-0.0438879 -0.0524827 -0.0476026

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -13.3021    13.8601  -0.960  0.381264
A              0.9765     0.1335   7.313  0.000749 ***
P              3.0834     4.8159   0.640  0.550193
S              2.9309     1.3119   2.234  0.075772 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1615 on 5 degrees of freedom
Multiple R-squared:  0.939,    Adjusted R-squared:  0.9024
F-statistic: 25.66 on 3 and 5 DF,  p-value: 0.001829
```

Figure 24 Summary of results for the regression using all predictor variables for Region-I

In this model, it can be seen that the P-Value of predictor variables Catchment area and mean slope is below 0.1 and the p-value at the bottom of the summary is 0.001829, which is significant. This means that, out of the three predictor variables these two independent variables which is catchment area and mean slope are highly related to the outcome variable or response variables statistically but physically mean annual rainfall is the main driving force to generate mean annual stream flow of a given catchment. On the other hand slope also plays an important role in producing flow, since the steepest the watershed the more the surface runoff and the mildest the slope, the contribution of rainfall to runoff will be less so mean catchment slope has significances on the development of mean annual flow of ungauged catchments.

# Regional Regression Model of Mean Annual Streamflow for Ungauged Catchments

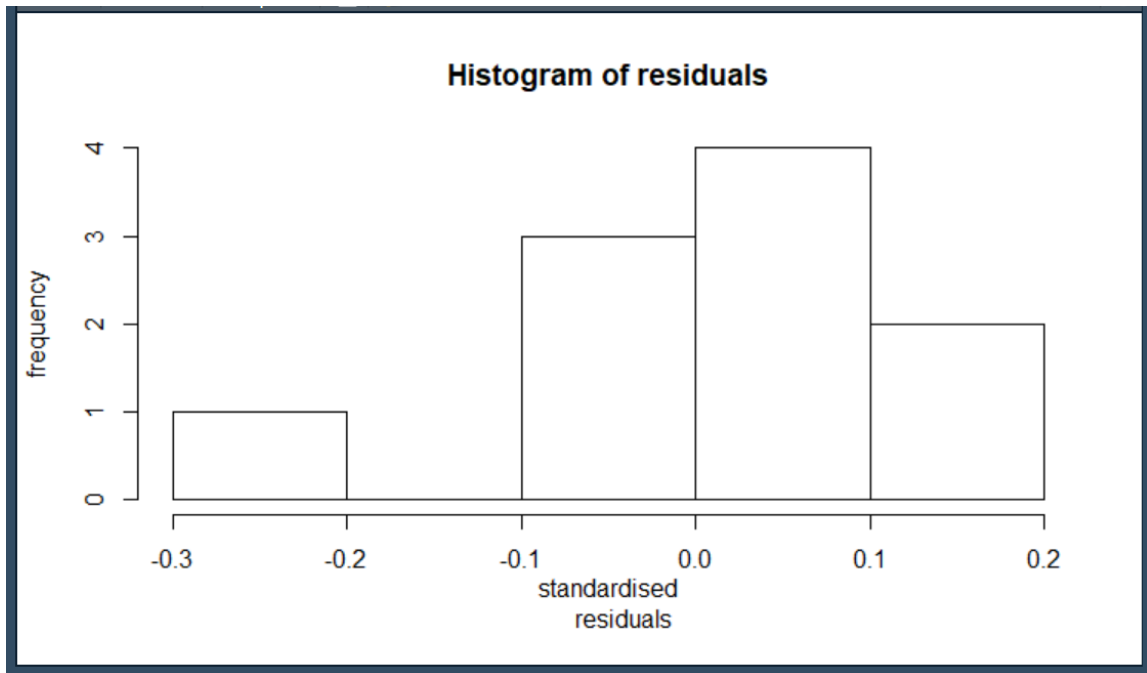


Figure 25 Residual plot of the model using all predictor variable

As shown on the above figure the residual seems normally distributed when the response variable was explained by all predictor variables that meets the assumption of a multiple linear regression model.

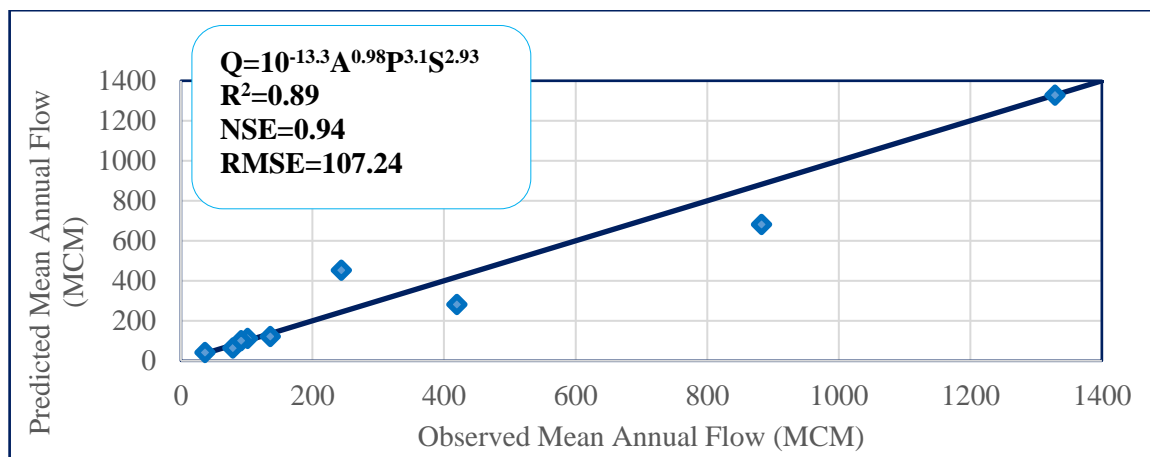


Figure 26 Plot of observed and estimated mean annual flow fit with 1:1 line

The above figure shows how well the graph of observed and estimated mean annual flow fit with a linear line (1:1 line), therefore the graphical representation of the regression using all predictor variables seems good and acceptable since most of the points between observed and predicted mean annual flow tends to the line 1:1.

# Regional Regression Model of Mean Annual Streamflow for Ungauged Catchments

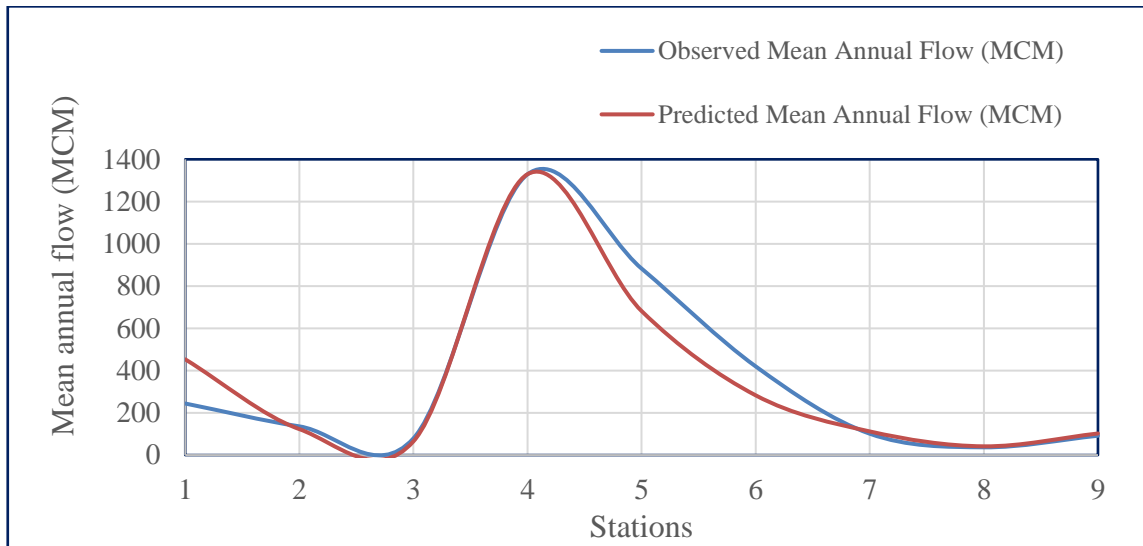


Figure 27 Plot of MLR model using all a predictor variable

```
87 #calculation of NSE and RMSE
88 Allvariable=read.csv(file.choose(),header=T)
89 library(hydroGOF)
90 sim<-Allvariable$QP
91 obs<-Allvariable$Q
92 NSE(sim, obs)
93 rmse(sim, obs)
94
```

88:45 (Top Level) R Script

Console F:/Desktop/NewR/AllAwash/

```
> NSE(sim, obs)
[1] 0.935351
> rmse(sim, obs)
[1] 107.2407
```

Figure 28 Estimation of NSE and RMSE for Region-I

Figure 27 shows the performance of MLR model of the selected datasets for the most suitable predictor variables of ten (9) gauged stations of Awash River basin for Region-I. As a result the following model of MLR was developed using the most appropriate independent variables and mean annual flow, which is the response variable.

# Regional Regression Model of Mean Annual Streamflow for Ungauged Catchments

## 4.3.2 Results for Regression using the most appropriate predictor variables

```
> summary(results)
Call:
lm(formula = Q ~ A + S, data = upperawash)

Residuals:
    Min       1Q   Median       3Q      Max
-0.263892 -0.066135  0.009013  0.053627  0.211120

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -4.4889    1.5374   -2.920 0.026637 *
A              0.9654    0.1257    7.678 0.000255 ***
S              3.3920    1.0412    3.258 0.017297 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1533 on 6 degrees of freedom
Multiple R-squared:  0.934,    Adjusted R-squared:  0.912
F-statistic: 42.47 on 2 and 6 DF,  p-value: 0.0002872
```

In this analysis it can be seen that t-statistic of predictor variables Catchment area is below 0.05 and the p-value at the bottom of the summary is 0.0002872, which is highly significant. This means that, the predictor variable catchment area and mean slope are highly related to the outcome variable or response variables statistically.

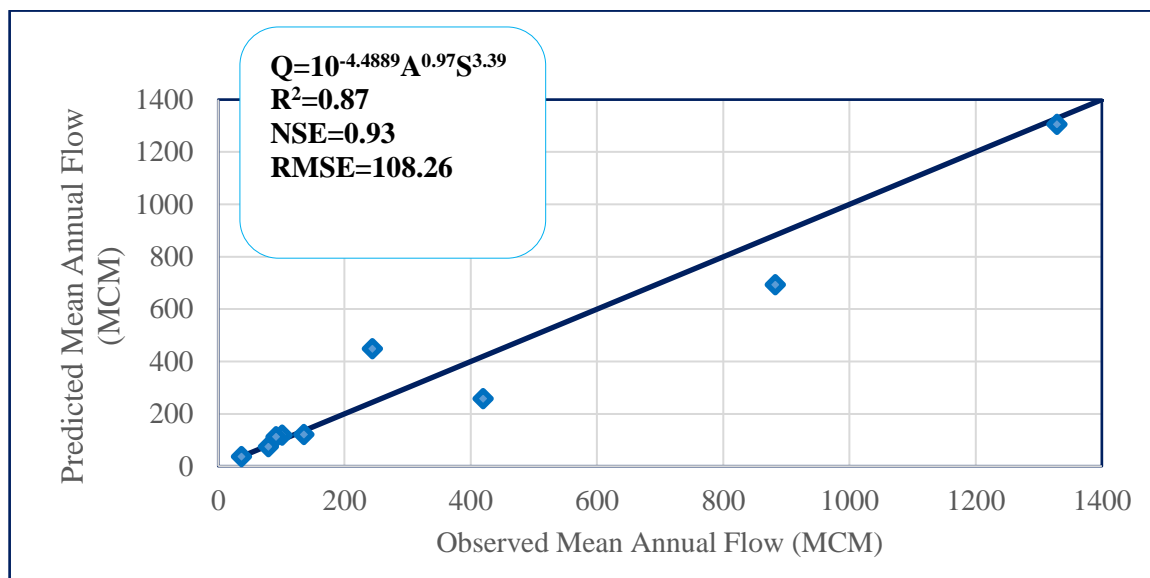


Figure 29 Plot of observed and estimated mean annual flow using most suitable variables

# Regional Regression Model of Mean Annual Streamflow for Ungauged Catchments

## 4.4 Regression Analysis of mean annual flow for Region-II

### 4.4.1 Results for Regression using all predictor variables

```
> summary(results)
Call:
lm(formula = Q ~ ., data = lowerawash)

Residuals:
    1     2     3     4     5     6
-0.09875  0.02484 -0.16185 -0.06809  0.05842  0.03718
    7     8
 0.00413  0.20412

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  -6.8010     4.1939  -1.622  0.2033
A              0.7671     0.1704   4.501  0.0205 *
P              0.7565     1.4039   0.539  0.6274
E              1.0987     0.5737   1.915  0.1513
S              0.6648     0.5680   1.171  0.3263
---
Signif. codes:
  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.171 on 3 degrees of freedom
Multiple R-squared:  0.9034,    Adjusted R-squared:  0.7747
F-statistic: 7.018 on 4 and 3 DF,  p-value: 0.07066
```

Figure 30 Summary results for the regression using all predictor variables of Region-II

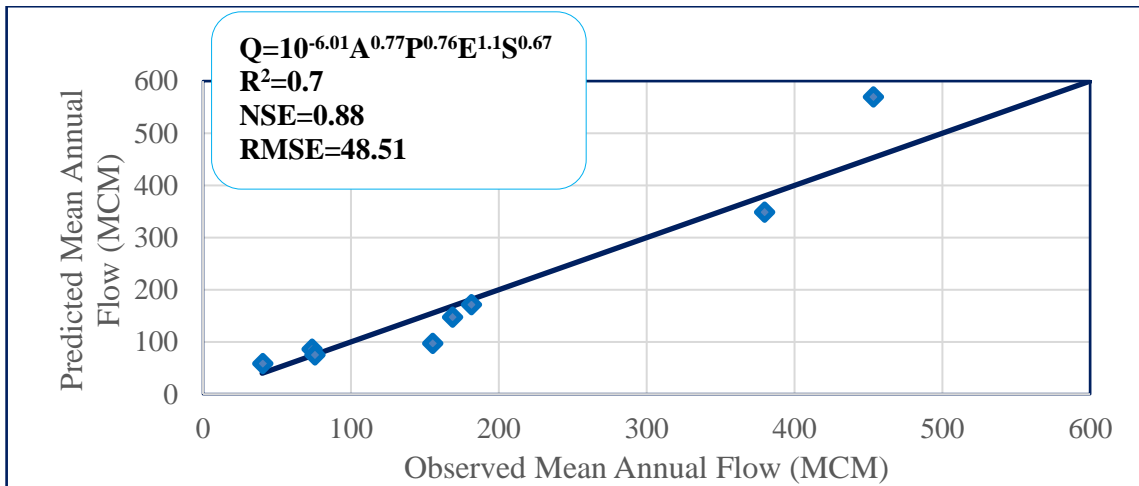


Figure 31 Plot of observed and estimated mean annual flow using all variables

# Regional Regression Model of Mean Annual Streamflow for Ungauged Catchments

## 4.4.2 Results for Regression using most appropriate variables

```
> summary(results)

Call:
lm(formula = Q ~ A, data = lowerawash)

Residuals:
    Min       1Q   Median       3Q      Max
-0.55291 -0.12484  0.09966  0.22357  0.25130

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   0.9370     0.6598   1.420   0.205
A              0.4284     0.2282   1.877   0.110

Residual standard error: 0.3088 on 6 degrees of freedom
Multiple R-squared:  0.37,    Adjusted R-squared:  0.265
F-statistic: 3.524 on 1 and 6 DF,  p-value: 0.1096
```

Figure 32 Summary results of the regression using area for Region-II

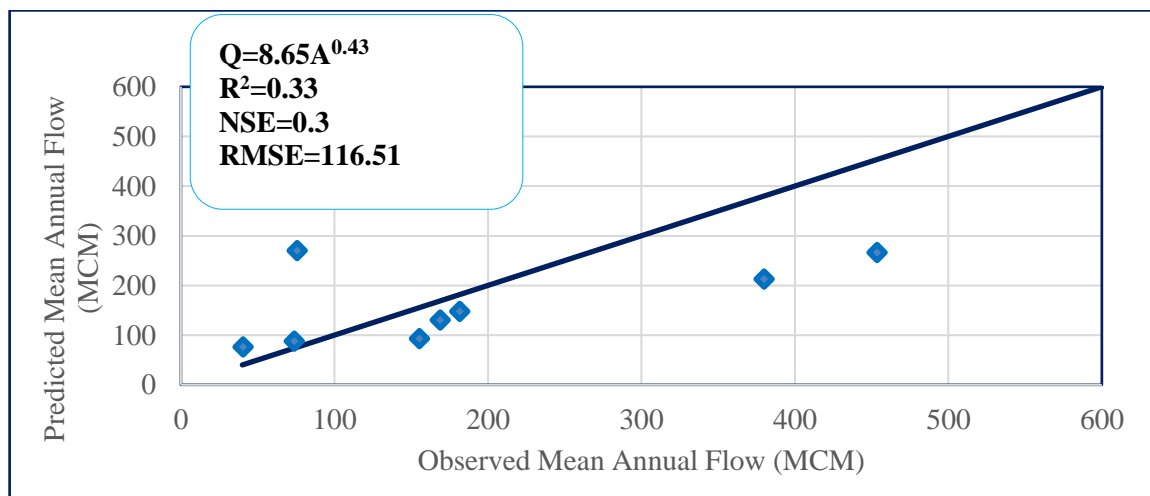


Figure 33 Mean Annual observed and predicted flow of Regression for Region-II

Summary of model performance indicators for the above trained models, the first MLR consisting all predictor variables and the second MLR was conducted using most appropriate variables identified by hypothesis test such as p-value of each regions were summarized on (Table 16). For Region-I the most significant predictor variables were catchment area, mean annual rainfall and mean slope but mean elevation is highly correlated with both mean annual rainfall and mean slope and it was removed from using as explanatory variable. This is due to the awash uplands (upper awash) is mountainous and the elevation decrease when moving from highlands of Ginchi to Koka dam, this implies that mean slope and mean annual rainfall would highly generate streamflow than

## Regional Regression Model of Mean Annual Streamflow for Ungauged Catchments

the mean elevation did. Therefore mean elevation was not significant to generate flow in ungauged catchments of Region-I compared to the other predictor variables.

Table 16 Summary of all models with model performance indicators

Regions	Predictor Variables used	Equation	Model performance Indicators		
			R <sup>2</sup>	NSE	RMSE
Region-I	A,P & S	$Q=10^{-13.3}A^{0.98}P^{3.1}S^{2.93}$	0.89	0.94	107.24
	A & S	$Q=10^{-4.49}A^{0.97}S^{3.39}$	0.87	0.93	108.26
Region-II	A ,P, E & S	$Q=10^{-6.01}A^{0.77}P^{0.76}E^{1.1}S^{0.67}$	0.70	0.88	48.51
	A	$Q=8.65A^{0.43}$	0.33	0.30	116.51

For each region two equations were developed and the most appropriate equation can be chosen by simply selecting an equation with high R<sup>2</sup> and NSE value and small RMSE value. Therefore the most appropriate equation for Region-I is  $Q=10^{-13.3}A^{0.98}P^{3.1}S^{2.93}$  and  $Q=10^{-6.01}A^{0.77}P^{0.76}E^{1.1}S^{0.67}$  was selected for Region-II. These equations can be used for mean annual streamflow estimation of ungauged catchments of each regions in Awash River basin.

Table 17 Criteria for the goodness-of-fit evaluation

Performance rating	Model efficiency interpretation	NSE
Very Good	SD >= 3.2RMSE	>= 0.9
Good	SD = 2.2RMSE-3.2RMSE	0.8-0.9
Acceptable	SD = 1.2RMSE-2.2RMSE	0.65-0.8
Unsatisfactory	SD < 1.7RMSE	<0.65

Table 16 shows the Criteria for the goodness-of-fit evaluation adopted from (Ritter and Muñoz-Carpena 2013) based on this criteria for the Region-I an equation with predictor variables A, P and S satisfies a Very good RMSE value ( SD = 447.36 , which is greater than (3.2\*107.24 = 343.17)), that is a very good performance and Very good NSE value

## Regional Regression Model of Mean Annual Streamflow for Ungauged Catchments

---

(NSE > 0.9). Note that SD stands for the standard deviation for the observed mean annual flow in MCM. On the other hand the goodness of fit evaluation for Region-II, an equation with four predictor variables satisfies a good RMSE value ( SD = 149.35 , which is greater than (2.2\*48.51 = 106.722)), that is a good performance and good NSE value (NSE > 0.8). Several regression were performed with the intention of optimizing the coefficients and the optimal multiple regression equation was found to be:

For Region-I

$$Q_{MAF} = 10^{-13.3} A^{0.98} P^{3.1} S^{2.93}$$

For Region-II

$$Q_{MAF} = 10^{-6.01} A^{0.77} P^{0.76} E^{1.1} S^{0.67}$$

Where  $Q_{MAF}$  is the mean annual flow in MCM

A is the catchment area in  $Km^2$

E is mean elevation of the catchment in meters, m

Based on Awash River basin master plan Regional Regression analysis of mean annual stream flow was developed for Awash River basin using very limited number of stations as;

$$MAF = 0.000845 \text{ AREA}^{0.80118} (\text{RAIN}-365)^{1.0048}$$

Where MAF= Mean Annual Flow, in MCM

Area = Catchment area, in  $Km^2$

RAIN= Mean annual rainfall, in mm

On the other hand a study conducted by USGS on Regional Regression Equations for Estimation of Natural Streamflow Statistics in Colorado in 2009 estimated the mean annual streamflow for mountain hydrologic region as;  $Q=10^{-2.31} A^{0.88} P^{1.71}$  where A in square miles and P in inches. These equations were used to estimate mean annual stream flow for ungauged catchments using only catchment area and mean annual precipitation but the equation developed in this thesis work consists of mean slope and mean elevation in addition to catchment area and mean annual precipitation and the result shows more

## Regional Regression Model of Mean Annual Streamflow for Ungauged Catchments

---

appropriate since the more catchment attributes consider the equation could be used to represent the physical characteristics of the basin. Generally The above equations will be used for mean annual streamflow estimation of ungauged catchments located on Region-I (awash uplands) and Region-II (most of western highlands of Awash River basin except ungauged rivers found on awash middle Valley, Awash Lower Plains and Awash Eastern Catchments) respectively.

## 5 CONCLUSIONS AND RECCOMENDATIONS

### 5.1 Conclusions

In this study homogeneity test and trend analysis of annual rainfall was analyzed using XLSTST 2019, out of 28 stations 19 stations were found as homogeneous and these stations were used for the development of mean annual rainfall of the selected gauging stations. Finally, the results of the present study would be a reference for future to researchers who studied on the water resources, hydrological processes, and climate change fields in Awash River basin.

Using annual rainfall and annual stream flow data an attempt was also made to formulate rainfall runoff model of each gauging stations and the model performance of the stations were performed using graphical and statistical measures, these rainfall runoff model for 8 gauging stations were developed using linear, power and polynomial functions. On the other hand runoff coefficient (C) was estimated using mean annual stream flow, mean annual rainfall and catchment area of each gauging stations and the result is important for flood control channel construction and for possible flood zone hazard delineation of Awash River basin.

There are many physiographical and hydrological characteristics exist at ungauged site that used in estimating Mean annual flow and the choice of these characteristics depends on the availability of data and scope of the study. In this study regression equation using multiple linear regression model was developed to estimate mean annual stream flow of any ungauged catchment in Awash River basin, to formulate this model two regions were formed based on similarity of rainfall distributions namely Region-I (upper awash sub basin) and Region-II (most of middle and lower Awash basins). Four physiographical and hydrological characteristics were implemented. From those four variables catchment area, mean annual rainfall and mean slope were found significant to develop MLR for Region-I and all the four variables were used to develop MLR for Region-II of Awash River basin.

# Regional Regression Model of Mean Annual Streamflow for Ungauged Catchments

---

## 5.2 Recommendations

In using MLR for computation of flow of ungauged catchments the most important thing is availability of multiple predictor variables with many samples, but most of the rivers in the country are either of ungauged or the gauging station is not working properly due to this, analyses becomes difficult specially obtaining continuously recorded observed flow data was difficult. Therefore by considering more physiographic and climatological variables like soil type, temperature, land use land cover and other suitable predictor variables will increase the performance of MLR models as well as the result would have the ability to indicate approximately the actual flow values at any ungauged catchments of the basin. It is also recommended to use the result obtained in this thesis only for the specified regions of Awash River basin.

# Regional Regression Model of Mean Annual Streamflow for Ungauged Catchments

---

## REFERENCES

- Agha, Omar M. A. Mahmood, S. Çağatay Bağçacı, and Nermin Şarlak. 2017. "Homogeneity Analysis of Precipitation Series in North Iraq." *IOSR Journal of Applied Geology and Geophysics* 05 (03): 57–63. <https://doi.org/10.9790/0990-0503025763>.
- Assessment, Goodness-of-fit. 2007. "Model Performance Analysis and Model Validation in Logistic Regression." *Statistica* 63 (2): 375–96. <https://doi.org/10.6092/issn.1973-2201/358>.
- Badyalina, Basri, and Ani Shabri. 2013. "Streamflow Forecasting at Ungauged Sites Using Multiple Linear Regression." *Matematika* 29 (1): 67–75.
- Barbarossa, Valerio, Mark A.J. Huijbregts, A. Jan Hendriks, Arthur H.W. Beusen, Julie Clavreul, Henry King, and Aafke M. Schipper. 2017. "Developing and Testing a Global-Scale Regression Model to Quantify Mean Annual Streamflow." *Journal of Hydrology* 544: 479–87. <https://doi.org/10.1016/j.jhydrol.2016.11.053>.
- Bekele, Daniel, Tena Alamirew, Asfaw Kebede, Gete Zeleke, and Assefa M. Melese. 2017. "Analysis of Rainfall Trend and Variability for Agricultural Water Management in Awash River Basin, Ethiopia." *Journal of Water and Climate Change* 8 (1): 127–41. <https://doi.org/10.2166/wcc.2016.044>.
- Benoit, Kenneth. 2011. "Linear Regression Models with Logarithmic Transformations." London School of Economics, 1–8.
- Casella, George, Stephen Fienberg, and Ingram Olkin. 2013. *An Introduction to Statistical Learning*. Springer Texts in Statistics.
- Dar, Lateef Ahmad. 2017. "Rainfall-Runoff Modeling Using Multiple Linear Regression Technique" 5 (Vii): 214–18.
- Engeland, Kolbjorn, and H. Hisdal. 2009. "A Comparison of Low Flow Estimates in Ungauged Catchments Using Regional Regression and the HBV-Model." *Water Resources Management* 23 (12): 2567–86. <https://doi.org/10.1007/s11269-008-9397-7>.
- Gebregiorgis, Abebe S, and Semu Moges. 2013. "Basin Regionalization for the Purpose of Water Resource Development in a Limited Data Situation: Case of Blue Nile River Basin, Ethiopia," no. October. [https://doi.org/10.1061/\(ASCE\)HE.1943-5584](https://doi.org/10.1061/(ASCE)HE.1943-5584).

# Regional Regression Model of Mean Annual Streamflow for Ungauged Catchments

---

- Gebremichael, Abiy. 2014. "Analysis of Seasonal Rainfall Variability for Agricultural Water Resource Analysis of Seasonal Rainfall Variability for Agricultural Water Resource Management in Southern Region , Ethiopia Coefficient of Variation Inter Tropical Convergence Zone Length Of." *Journal of Natural Sciences Research* 4 (11): 1–26.
- Ghumman, A. R., Q. U. Khan, H. N. Hashmi, and M. M. Ahmad. 2014. "Comparison of Clark, Nash Geographical Instantaneous Unit Hydrograph Models for Semi Arid Regions." *Water Resources* 364–71. <https://doi.org/10.1134/S0097807814040071>.
- Grimes, D. I.F., E. Pardo-Igúzquiza, and R. Bonifacio. 1999. "Optimal Areal Rainfall Estimation Using Raingauges and Satellite Data." *Journal of Hydrology* 222 (1–4): 93–108. [https://doi.org/10.1016/S0022-1694\(99\)00092-X](https://doi.org/10.1016/S0022-1694(99)00092-X).
- Imdadullah, Muhammad, Muhammad Aslam, and Saima Altaf. 2016. "Mctest : An R Package for Detection of Collinearity among Regressors" 8 (December): 495–505.
- Islam, Zahidul. 2015. "A Review on Physically Based Hydrologic Modeling Submitted By," no. October: 46. <https://doi.org/10.13140/2.1.4544.5924>.
- Maity, Rajib. n.d. *Statistical Methods in Hydrology and Hydroclimatology*.
- Merz, Ralf, Günter Blöschl, and Juraj Parajka. 2006. "Regionalization Methods in Rainfall-Runoff Modelling Using Large Catchment Samples." *IAHS-AISH Publication*, no. 307: 117–25.
- Ritter, Axel, and Rafael Muñoz-Carpena. 2013. "Performance Evaluation of Hydrological Models: Statistical Significance for Reducing Subjectivity in Goodness-of-Fit Assessments." *Journal of Hydrology*
- Safarina, Ariani Budi, Hang Tuah Salim, Iwan K Hadihardaja, and M Syahril Bk. 2011. "Clusterization of Synthetic Unit Hydrograph Methods Based on Watershed Characteristics." *International Journal of Civil & Environmental Engineering IJCEE-IJENS* 11 (06): 76–85.
- Singh, P. K., S. K. Mishra, and M. K. Jain. 2014. "Revue de l'hydrogramme Unitaire Synthétique: Des HU Empiriques Aux Méthodes Géomorphologiques Avancées." *Hydrological Sciences Journal*. <https://doi.org/10.1080/02626667.2013.870664>.
- Streamstats, Montana. 2015. "Methods for Estimating Streamflow Characteristics at Ungauged Sites in Western Montana Based on Data through Water Year 2009 Montana StreamStats."

# Regional Regression Model of Mean Annual Streamflow for Ungauged Catchments

---

- Sudheer, K. P., A. K. Gosain, and K. S. Ramasastri. 2002. "A Data-Driven Algorithm for Constructing Artificial Neural Network Rainfall-Runoff Models." *Hydrological Processes* 16 (6): 1325–30. <https://doi.org/10.1002/hyp.554>.
- Swain, J.B., R. Jha, and K.C. Patra. 2015. "Stream Flow Prediction in a Typical Ungauged Catchment Using GIUH Approach." *Aquatic Procedia* 4 (Icwrcoe): 993–1000. <https://doi.org/10.1016/j.aqpro.2015.02.125>.
- Tadese, Mahtsente Tibebe, Lalit Kumar, Richard Koech, and Birhanu Zemadim. 2019. "Hydro-Climatic Variability: A Characterisation and Trend Study of the Awash River Basin, Ethiopia." *Hydrology* 6 (2). <https://doi.org/10.3390/hydrology6020035>.
- Vogel, Richard M., Ian Wilson, and Chris Daly. 1999. "Regional Regression Models of Annual Streamflow for the United States." *Journal of Irrigation and Drainage Engineering* 148–57. [https://doi.org/10.1061/\(ASCE\)0733-9437\(1999\)125:3\(148\)](https://doi.org/10.1061/(ASCE)0733-9437(1999)125:3(148)).
- Wagener, T., M. Sivapalan, J. McDonnell, R. Hooper, V. Lakshmi, X. Liang, and P. Kumar. 2004. "Predictions in Ungauged Basins as a Catalyst for Multidisciplinary Hydrology." *Eos* 85 (44): 440003. <https://doi.org/10.1029/2004EO440003>.
- Wilkinson, Scott N., William J. Young, and Ron C. DeRose. 2006. "Regionalizing Mean Annual Flow and Daily Flow Variability for Basin-Scale Sediment and Nutrient Modelling." *Hydrological Processes* 2769–86. <https://doi.org/10.1002/hyp.6070>.

# Regional Regression Model of Mean Annual Streamflow for Ungauged Catchments

## APPENDICES

**Appendix A:** Table Summarizes rain gauge missing data analyses

Rain gauge	Missing Data (1981-2013)	Reg. eq. $y = ax$	Rain gauge priority for infilling			
			1	2	3	4
Asgori	3.54%	Donor Gauges	Teji	Busa		
		slope	1.0671	0.6582		
		R2	0.8093	0.5427		
Akaki	1.01%	Donor Gauges	AA Bole	Sebeta		
		slope	0.9424	0.4438		
		R2	0.8595	0.635		
Koka Dam	12.63%	Donor Gauges	Nazret	Welenchit	Ejere	Bologiorgis
		slope	0.8322	0.7442	0.8421	0.8421
		R2	0.5541	0.3579	0.5196	0.4331
Adaitu	0%	Donor Gauges				
		slope				
		R2				
Bologiorgis	3.79%	Donor Gauges	Ejere	Welenchit	Nazret	
		slope	0.9429	0.8575	0.9327	
		R2	0.76	0.6326	0.7358	
Kora	14.39%	Donor Gauges	Awash melka	Guder		
		slope	0.4784	0.3613		
		R2	0.2374	0.1314		
Boneya	3.03%	Donor Gauges	Guranda Meta	sebeta	Akaki	Teji
		slope	0.7027	0.3967	0.8306	0.8799
		R2	0.5888	0.3081	0.7707	0.7127
Welenchit	6.82%	Donor Gauges	Nazret	Ejere	Bologiorgis	Koka dam
		slope	0.9479	0.9209	0.9158	0.7596
		R2	0.6918	0.605	0.6552	0.297
Guder	0.76%	Donor Gauges	bologiorgis	welenchit	Ejere	
		slope	1.1087	1.0036	1.1283	
		R2	0.2896	0.0939	0.2991	
Combolcha	1.26%	Donor Gauges	Dessie zuria	Haik	Jara	
		slope	0.8223	1.0786	1.1252	
		R2	0.887	0.7932	0.6532	
Diredawa	2.02%	Donor Gauges	Gursum	Jijiga	Kulubi	
		slope	0.6223	0.9461	0.5574	
		R2	0.378	0.3976	0.5414	
AA Bole	1.77%	Donor Gauges	AA obs	Akaki	Sebeta	
		slope	0.8382	0.9806	1.2841	
		R2	0.8915	0.8518	0.372	
AA Obs	3.54%	Donor Gauges	AA bole	Sebeta	Akaki	
		slope	1.1264	0.4384	1.1303	
		R2	0.8912	0.1464	0.8413	
Nazret	5.81%	Donor Gauges	Koka dam	Welenchit	Ejere	Bologiorgis
		slope	0.8469	0.8541	0.9215	0.9327
		R2	0.4938	0.6752	0.742	0.7358
Welenkomi	11.11%	Donor Gauges	Kimoye	Ginchi	Dertuliben	
		slope	0.7708	0.8879	1.098	
		R2	0.4779	0.7791	0.6545	
Haik	3.79%	Donor Gauges	Kutaber	combolcha		
		slope	0.8684	1.0786		
		R2	0.739	0.7932		
Busa	8.84%	Donor Gauges	Asgori	Teji		
		slope	1.1522	1.2786		
		R2	0.6074	0.5939		
Kimoye	3.79%	Donor Gauges	Welenkomi	Ginchi	Dertuliben	
		slope	0.9558	0.8292	1.0727	
		R2	0.5307	0.4357	0.3224	
Teji	1.26%	Donor Gauges	Asgori	Boneya		

## Regional Regression Model of Mean Annual Streamflow for Ungauged Catchments

		slope,a	0.8421	0.9471		
		R2	0.806	0.679		
Sendafa	8.33%	Donor Gauges	AA Bole	AA Obs		
		slope	1.1332	0.9713		
		R2	0.7823	0.7743		
Sebeta	5.05%	Donor Gauges	Guranda Meta	Akaki	AA Obs	AA Bole
		slope	1.1184	1.3548	1.1247	1.2841
		R2	0.2762	0.4288	0.4687	0.372
Jara	2.02%	Donor Gauges	Debresina	combolcha		
		slope	0.6294	1.1252		
		R2	0.5793	0.6532		
Guranda Meta	1.77%	Donor Gauges	sebeta	Boneya	Akaki	
		slope	0.4451	1.0875	1.0165	
		R2	0.1297	0.6012	0.7058	
Ginager	4.04%	Donor Gauges	Guder	Awra melka	Sendafa	Bologiorgis
		slope	0.986	2.7435	1.201	1.5975
		R2	0.2085	0.3723	0.5364	0.6102
Ejere	7.83%	Donor Gauges	bologiorgis	welenchit		
		slope	0.9143	0.8193		
		R2	0.789	0.5818		
Dertuliben	6.82%	Donor Gauges	Ginchi	Kimoye		
		slope	0.6393	0.5797		
		R2	0.5766	0.3181		
Ginchi	1.52%	Donor Gauges	Dertuliben	welenkomi	kimoye	
		slope	0.6393	0.99985	0.8292	
		R2	0.5766	0.7594	0.4357	
Dessie Zuria	21.20%	Donor Gauges	Combolcha			
		slope	1.142			
		R2	0.9114			
Awash Melka	121.20%	Donor Gauges	Kora			
		slope	1.176			
		R2	0.3738			

# Regional Regression Model of Mean Annual Streamflow for Ungauged Catchments

## Appendix B: Sample rain gauge stations with filled missing data

Addis Ababa bole monthly rainfall data

Year	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1981	0	42.6	254.7	79	18.4	56.9	273.9	256.1	180.5	24.7	0	2.7
1982	26.6	96.4	90.2	48.1	73.5	63.6	220.3	221.6	142.8	19	40.7	4.9
1983	12.4	41.2	28.9	113.7	186.9	56.1	217.9	213.7	202.2	35.9	0	1.5
1984	0	0.4	11.6	11.6	135	334.2	317.9	180.4	98.8	0	0	7
1985	35.1	0	49.1	132.3	92.8	110.9	209.8	260.8	168.6	29.8	0	0.4
1986	0	37.6	56.2	216.6	37.7	175.2	167.9	222.3	107.4	31.6	0	2.5
1987	0	49.1	180.1	85.7	154.6	71.9	155.9	98.1	57	16.6	0	0.4
1988	4.7	33.4	6.7	157.9	34.7	93.2	181.4	265.3	187.3	57.3	0	0
1989	3.4	33.7	58.4	143.3	0	88.1	218.1	318.6	150	36.8	0	7.9
1990	3.2	161.1	60.4	144.5	25.2	48.3	194.2	293.6	143	46.1	2.1	0
1991	0.2	29.6	134.1	15	7.7	107.5	279.4	287.9	123.1	4.4	2.1	0
1992	14.5	28	35	58.6	55	82.2	254.8	223.3	157	64.4	2.2	0.4
1993	11.7	52.1	11.6	168.3	91.5	157.2	209.5	291.7	190.1	24.1	0	0
1994	0	0	52.9	70	29	112.4	242.3	199.3	99.4	0.5	11	0
1995	0	81.3	73.1	133.3	95.9	77.4	165.5	256.9	97	0	0	29.3
1996	20.5	15.8	134.4	96	124.6	290.2	346.3	312.7	211.4	0.2	0.4	0
1997	29.1	0	22.1	66.8	44.8	128	257	160.7	94.7	58.6	15.3	0
1998	66.6	40	43.8	99.8	197.7	111.6	270.7	236.8	173.4	139.4	0	0
1999	4.4	0	35	17.8	30.5	104.6	294	270.5	62.8	127.1	0	0
2000	0	0	17.6	87.8	95.2	102.1	192.9	221.9	157.5	19.6	7.5	0
2001	0	10.3	174.3	14.8	116.7	166	289.4	207.3	113.3	10.6	0	0
2002	30.6	25.9	79.4	36.6	49.6	115.5	213.9	233.6	72.6	0.5	0	32.8
2003	4.8	44.7	48.9	111.5	18	111	204.3	238.4	130.2	4.6	0	33.3
2004	20.8	11.7	32.4	117.3	7	114.5	240.6	230.1	122.1	50	0.6	0
2005	55.4	14.1	41.8	116.2	164.6	159.1	174.3	248	77.6	25.8	7.2	0
2006	2	36.6	107.8	93.9	37.8	115.1	313.2	331.1	132.5	35.9	0	0
2007	9.9	21.3	61.1	86.8	134	157.6	191.3	305.4	130.9	37.2	0.1	0
2008	0	0	0	34	75.3	73.1	295.1	259.1	192.7	22.2	53.1	0
2009	40.9	0	12.4	46.1	52	77.5	238.2	269.5	86.1	42.4	2	79.9
2010	0.4	115.2	75.6	159.5	94.7	107.2	320.2	138.8	105	0	13.8	15.7
2011	3.4	13.6	27.9	19.1	86	148	183.1	296.5	141.3	0	11.9	0
2012	0	0	34.5	75.1	58.5	72.8	228.8	281.6	176.9	1.2	0	8.2
2013	3.9	0	63.5	114.4	78.5	101.4	157.6	270.2	126.7	45.3	3.2	0

# Regional Regression Model of Mean Annual Streamflow for Ungauged Catchments

## Addis Ababa observatory monthly rainfall data

Year	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1981	0.0	75.5	176.0	82.9	3.9	50.1	266.5	320.9	182.1	13.3	0.0	5.2
1982	48.7	80.9	57.8	103.7	115.9	31.9	259.3	257.9	133.8	64.4	43.2	12.8
1983	18.3	21.7	48.7	117.0	237.0	109.3	199.3	244.7	160.9	26.3	0.0	8.8
1984	0.0	8.0	8.8	8.4	127.8	220.8	296.1	295.6	142.4	0.0	4.4	16.3
1985	14.2	0.0	17.5	96.3	83.7	112.2	270.4	327.7	205.9	58.0	3.3	1.2
1986	0.0	35.7	88.0	197.6	125.4	179.5	180.1	250.4	127.8	36.1	0.0	2.8
1987	0.5	63.4	248.9	82.4	241.3	92.9	196.5	254.4	115.2	21.3	0.8	0.3
1988	9.7	53.4	5.3	144.6	16.6	106.2	277.9	299.3	229.7	59.9	0.0	0.0
1989	0.8	75.9	75.7	154.4	0.5	120.9	357.7	325.3	187.7	14.8	0.0	7.6
1990	0.8	155.9	59.2	106.4	20.0	88.8	218.7	268.6	184.0	16.2	6.0	0.0
1991	0.0	74.5	106.6	34.7	8.7	121.1	248.9	262.6	126.4	3.4	0.0	50.0
1992	20.2	33.7	20.2	41.0	52.0	109.1	248.5	294.7	209.4	69.7	0.0	2.9
1993	10.8	67.2	16.1	157.9	97.2	208.3	274.0	426.5	243.3	62.1	0.0	4.5
1994	0.0	0.0	82.4	82.3	63.3	123.4	308.9	225.0	142.0	0.5	14.7	0.0
1995	0.0	69.0	41.5	174.4	68.2	102.9	190.2	314.9	136.1	0.0	0.0	33.0
1996	28.1	5.2	106.8	128.2	122.0	258.5	266.4	338.7	294.2	0.2	0.2	0.0
1997	39.2	0.0	24.5	51.3	38.5	104.0	272.6	194.3	113.8	62.4	50.3	1.5
1998	55.2	20.5	49.0	48.5	154.2	124.4	285.4	260.0	213.6	126.9	0.0	0.0
1999	2.9	0.3	28.8	16.3	23.8	119.6	331.2	305.3	88.4	75.4	0.0	0.0
2000	0.0	0.0	17.6	49.9	110.0	144.5	244.8	306.2	250.6	46.4	21.1	0.0
2001	0.0	12.2	210.8	25.0	168.0	216.2	428.0	246.4	131.7	13.7	0.0	0.0
2002	14.7	21.0	90.2	56.3	63.1	172.5	256.9	215.9	108.8	0.2	0.0	16.5
2003	10.5	53.3	62.6	99.3	20.2	151.8	291.8	233.3	193.3	0.8	1.5	54.9
2004	24.8	20.3	49.5	139.9	30.1	141.9	238.5	272.6	164.0	76.9	0.0	0.0
2005	45.9	51.6	83.2	160.9	133.7	179.8	246.0	315.2	162.5	29.1	4.4	0.0
2006	0.7	11.2	121.4	78.9	74.6	150.1	356.3	243.6	239.1	54.0	0.3	8.0
2007	51.3	19.1	59.8	73.8	120.1	177.5	261.8	381.2	147.6	24.8	0.0	0.0
2008	0.0	13.0	0.0	49.4	94.3	88.9	277.0	360.9	256.7	88.2	79.4	22.9
2009	21.3	2.7	28.4	80.6	58.9	82.6	349.9	388.3	112.7	45.8	4.4	65.0
2010	2.6	79.8	55.5	97.8	74.4	271.1	313.9	205.8	237.8	1.8	25.7	15.0
2011	14.1	13.1	44.3	22.8	66.1	182.0	180.9	340.8	146.0	0.0	42.3	0.0
2012	0.0	0.0	15.8	84.6	50.2	69.4	324.2	298.0	215.5	2.3	0.0	9.8
2013	4.4	0.0	71.5	92.3	85.0	153.2	177.5	353.2	142.7	58.4	22.3	0.0

# Regional Regression Model of Mean Annual Streamflow for Ungauged Catchments

## Akaki gauging station monthly rainfall

Year	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1981	0.0	13.3	179.8	132.0	1.3	49.2	402.6	165.0	223.1	5.0	0.0	0.0
1982	12.1	35.4	39.5	94.6	75.2	63.5	238.8	255.9	125.2	25.8	11.0	8.1
1983	1.8	33.3	20.2	142.1	175.0	83.0	237.7	275.3	146.7	0.0	0.0	0.0
1984	0.0	0.0	40.4	5.1	130.0	215.3	277.9	227.1	71.7	0.0	0.0	1.9
1985	3.6	0.0	32.4	71.8	96.6	96.5	303.0	324.1	164.3	1.6	0.0	0.0
1986	0.0	95.4	67.7	148.7	83.2	143.4	189.4	216.5	86.1	9.4	0.0	0.0
1987	0.0	65.6	181.9	81.2	187.7	69.3	202.0	246.9	82.5	4.4	0.0	0.0
1988	0.0	44.5	0.0	96.0	23.8	124.6	255.9	278.1	253.5	35.4	0.0	0.0
1989	2.1	63.8	53.8	226.3	7.1	58.6	264.2	301.0	170.9	37.9	0.0	0.0
1990	7.7	120.6	48.4	159.1	37.3	78.9	280.7	222.9	107.3	5.8	1.2	0.0
1991	0.0	37.6	62.4	11.6	45.6	90.4	263.7	308.5	113.3	4.4	0.0	56.5
1992	33.5	24.2	30.5	15.5	25.6	100.4	218.4	276.0	86.7	43.3	0.2	0.0
1993	1.2	53.9	5.6	118.4	54.6	116.5	218.0	251.5	118.3	20.5	0.0	0.0
1994	0.0	0.0	62.7	72.2	20.2	131.2	219.3	181.0	94.5	0.0	11.0	0.0
1995	0.0	25.4	62.7	102.1	20.9	95.7	279.0	242.3	79.5	0.0	0.0	4.8
1996	15.3	0.3	79.7	38.8	90.5	240.1	292.5	234.1	119.0	1.9	0.0	0.0
1997	27.6	0.0	29.5	102.7	25.2	57.0	203.6	203.4	89.2	114.9	10.3	0.0
1998	32.7	30.2	19.6	69.3	159.9	116.9	207.8	280.0	118.5	36.9	0.0	0.0
1999	1.3	1.8	91.8	12.1	45.4	92.8	282.6	300.7	61.7	65.0	0.0	0.0
2000	0.0	0.0	29.1	93.0	64.9	100.1	188.9	210.0	124.1	17.2	23.4	3.8
2001	0.0	20.7	121.2	23.6	118.0	142.6	257.5	145.0	64.9	2.2	0.0	0.0
2002	31.1	10.5	87.8	53.9	76.6	108.0	167.1	166.3	52.3	0.0	0.0	17.7
2003	19.6	24.8	23.9	114.0	1.4	125.4	325.1	307.4	113.4	0.0	0.0	40.8
2004	15.6	15.8	61.4	154.5	15.4	95.2	150.3	189.1	80.9	4.8	3.4	0.7
2005	28.8	7.3	47.9	112.2	140.7	139.9	218.7	231.4	152.7	9.1	15.2	0.0
2006	2.6	44.2	56.3	79.7	22.0	84.3	276.4	262.6	148.1	38.0	0.0	3.2
2007	34.2	24.7	25.6	96.8	64.6	132.7	254.2	221.8	148.5	14.3	1.3	0.0
2008	0.0	0.0	0.6	34.2	62.4	140.2	253.5	252.3	181.6	7.2	64.8	0.0
2009	60.2	0.0	10.0	118.7	47.7	63.5	235.3	322.4	71.3	32.8	4.0	16.8
2010	0.0	63.8	126.2	170.0	95.2	164.8	334.4	169.8	154.1	5.2	14.8	7.8
2011	0.0	2.5	45.2	20.7	128.7	60.0	204.3	304.0	194.5	0.0	4.7	0.0
2012	0.0	0.0	32.5	61.0	26.1	80.6	228.0	243.9	122.9	0.0	0.0	0.0
2013	0.0	0.0	77.0	89.1	73.4	111.5	179.6	242.4	142.5	20.6	0.0	0.2

# Regional Regression Model of Mean Annual Streamflow for Ungauged Catchments

Kora rain gauge station monthly rainfall data

Year	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1981	0.0	5.2	150.4	231.1	28.3	21.4	170.3	213.8	175.6	52.2	7.7	0.0
1982	18.0	46.5	188.4	115.1	148.9	12.9	104.9	254.7	50.5	163.6	65.1	47.6
1983	0.0	90.5	45.3	165.1	193.5	63.3	207.6	67.6	89.7	37.6	15.9	17.6
1984	10.2	16.0	27.6	16.3	253.2	108.3	83.2	62.5	135.4	7.9	23.9	29.1
1985	39.9	5.6	32.9	292.4	234.7	15.7	88.2	108.6	55.8	27.1	12.0	12.6
1986	0.0	42.0	114.4	16.9	0.0	87.4	77.4	112.6	73.8	34.5	0.0	13.4
1987	0.0	14.7	114.6	133.7	190.3	22.1	17.7	96.7	67.2	0.0	0.0	3.2
1988	23.2	32.5	10.1	112.8	13.1	31.1	61.7	134.0	211.5	38.9	0.0	13.4
1989	0.0	44.2	63.0	205.1	23.3	39.2	73.4	96.3	114.9	15.1	0.0	89.3
1990	0.0	74.2	138.0	106.7	41.0	5.2	68.0	139.4	130.0	20.4	1.2	6.3
1991	2.4	36.6	137.5	57.1	3.9	34.1	113.0	123.9	98.4	0.0	0.0	27.3
1992	0.0	0.0	0.0	0.0	0.0	0.0	63.0	87.9	112.1	170.5	101.9	0.0
1993	0.0	34.9	14.3	174.6	90.0	24.7	243.9	207.9	129.3	24.2	0.0	0.0
1994	0.0	0.0	58.5	72.7	48.1	55.4	313.6	255.2	333.1	16.7	31.4	0.0
1995	0.0	16.6	129.7	180.5	51.3	27.0	167.8	223.4	185.8	0.0	3.8	39.7
1996	135.2	79.9	153.3	150.5	109.5	79.9	143.9	208.5	113.9	0.0	46.3	0.0
1997	5.0	0.0	5.5	142.1	62.9	90.4	135.5	51.1	15.9	281.2	146.2	0.0
1998	15.7	43.2	31.5	52.2	104.5	32.0	107.1	149.1	93.6	82.0	7.8	0.0
1999	15.9	27.1	105.1	19.0	18.2	55.5	239.1	159.2	109.0	35.5	0.0	0.0
2000	0.0	0.0	2.4	41.5	186.0	53.4	64.1	138.5	148.8	76.0	19.8	1.0
2001	0.0	19.5	108.5	28.0	24.0	52.0	115.5	112.4	7.0	0.0	0.0	16.8
2002	42.4	3.6	62.5	71.1	10.3	21.0	23.0	14.6	4.3	0.0	0.0	14.0
2003	11.6	23.0	93.0	98.4	0.0	43.8	76.8	104.0	99.6	0.0	51.5	8.1
2004	9.3	0.0	86.3	294.8	0.0	20.6	35.8	54.5	120.9	22.8	18.0	20.9
2005	13.8	0.0	19.9	39.7	88.8	60.0	104.1	37.3	25.2	0.0	50.0	0.0
2006	4.5	40.8	89.5	44.4	72.1	37.8	129.9	115.4	26.8	5.4	0.0	1.7
2007	6.0	5.1	29.0	21.2	22.6	15.8	29.9	187.1	68.8	2.6	0.4	0.0
2008	19.1	0.0	0.0	102.8	44.8	55.6	96.2	94.2	6.3	21.7	0.0	0.0
2009	0.0	0.0	0.0	48.3	4.3	20.6	39.2	82.0	35.9	25.2	0.4	19.2
2010	0.0	42.8	34.6	108.8	11.8	22.3	181.0	109.8	38.6	0.0	9.2	8.4
2011	0.0	0.0	33.9	22.9	29.9	9.4	37.9	62.9	14.0	0.0	6.7	0.0
2012	0.0	0.0	3.8	6.0	10.4	76.5	196.0	31.0	101.5	0.0	0.0	0.0
2013	54.7	0.0	91.4	93.7	70.1	31.2	119.4	80.1	38.6	57.6	17.9	0.0

# Regional Regression Model of Mean Annual Streamflow for Ungauged Catchments

**Appendix C:** Table showing Developed equations of Annual flow for each Gauging stations

s.no	station name	Trained models	Equation	model performance			
				R <sup>2</sup>	R <sup>2</sup> adj	NSE	RMSE
1	Awash at Melkahomble	power function	$Q=6.258*10^{-5}*P^{2.07}$	0.64	0.62	0.63	18.40
		polynomial	$Q=0.0001*P^2$	0.59	0.67	0.63	18.42
		Linear	$Q=0.11*P$	0.15	1.01	0.46	22.19
2	kesem at Awramelka	power function	$Q=0.13*P^{1.15}$	0.54	0.80	0.56	64.74
		polynomial	$Q=6.42E^{-07}P^2+0.36P+0.99$	0.41	0.94	0.55	65.46
		Linear	$Q=0.36*P$	0.41	0.94	0.55	65.45
3	Awash at melkakunture	power function	$Q=0.0014*P^{1.92}$	0.54	0.73	0.54	146.46
		polynomial	$Q=0.00073*P^2+39.32$	0.54	0.74	0.54	146.41
		Linear	$Q=0.81*P$	0.15	1.02	0.41	164.53
4	Awash Near Bello	power function	$Q=2.18*P^{0.7}$	0.67	0.58	0.67	30.60
		polynomial	$Q=0.000046*P^2+193.61$	0.04	1.03	0.04	54.63
		Linear	$Q=0.21*P+76.81$	0.67	0.58	0.67	30.59
5	Mojo at mojob	power function	$Q=2.45*10^{-10}*P^{3.91}$	0.62	0.65	0.54	38.03
		polynomial	$Q=0.00014P^2$	0.20	1.01	0.43	42.23
		Linear	$Q=0.14*P$	0.05	1.04	0.26	47.96
6	Little akaki	power function	$Q=1.31*10^{-10}*P^{3.87}$	0.52	0.79	0.56	34.89
		polynomial	$Q=6.87*10^{-5}*P^2$	0.15	1.06	0.42	40.10
		Linear	$Q=0.073*P$	0.04	1.08	0.24	46.01
7	Akaki at akaki	power function	$Q=1*10^{-9}*P^{3.81}$	0.28708	0.95429	0.13	210.363
		polynomial	$Q=0.00035P^2$	0.08	1.03	0.11	294.30
		Linear	$Q=0.39*P$	0.02	1.04	0.06	223.85
8	Teji Nr asgor	power function	$Q=9.01*10^{-7}P^{2.66}$	0.42	0.88	0.63	24.30
		polynomial	$Q=9.6*10^{-5}*P^2+0.015*P$	0.49	0.82	19.46	0.64
		Linear	$Q=0.093*P$	0.06	1.07	0.34	32.40
9	Keleta Nr sirie	power function	$Q=0.017*P^{1.34}$	0.19	1.02	0.21	63.16
		polynomial	$Q=0.00021*P^2+82$	0.30	0.99	0.30	54.20
		Linear	$Q=0.19*P$	0.12	1.04	0.19	63.67
10	Berga Nr	power	$Q=0.0001*P^{1.97}$	0.72	0.51	0.70	14.80

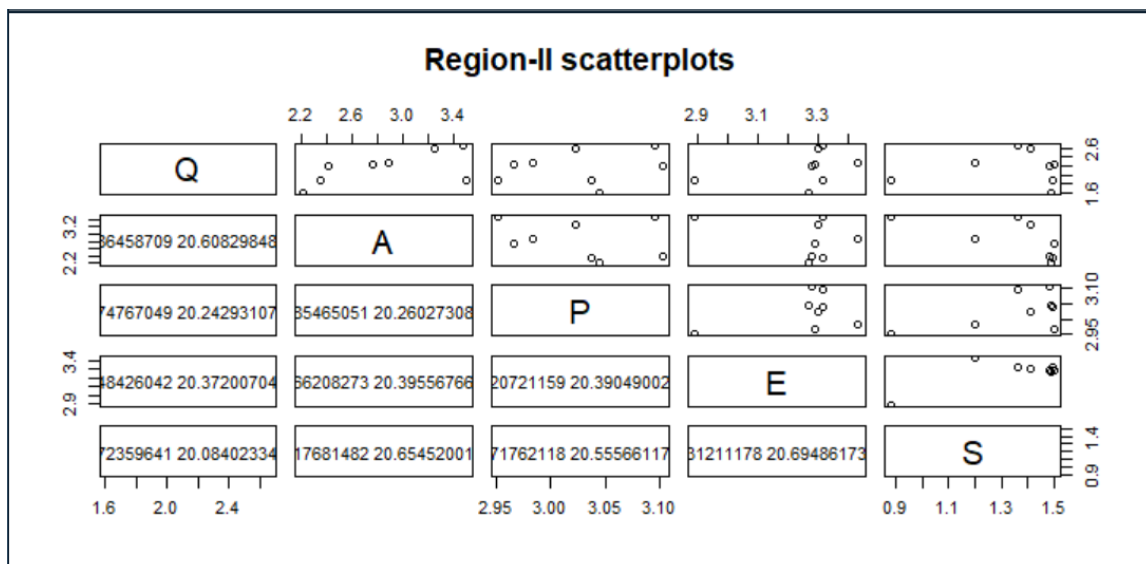
## Regional Regression Model of Mean Annual Streamflow for Ungauged Catchments

	Adisalem	function					
		polynomial	$Q=0.000077 * P^2 + 2.5$	0.70	0.53	0.70	14.80
		Linear	$Q=0.088 * P + 125.54$	0.19	1.01	0.53	18.68
11	Robi at Shewarobit	power function	$Q=6.58 * 10^{-07} * P^{2.69}$	0.29	1.02	0.28	89.39
		polynomial	$Q=0.000097 * P^2$	0.17	1.08	0.26	90.46
		Linear	$Q=0.13 * P$	0.05	1.11	0.17	95.89
12	Kesem at Beke	power function	$Q=0.00012 * P^{1.77}$	0.56	0.73	0.55	4.35
		polynomial	$Q=0.00002 * P^2 + 4.65$	0.55	0.75	0.55	5.72
		Linear	$Q=0.029 * P$	0.18	1.04	0.44	6.37
13	Logia at Logia v	power function	$Q=2.12 * P^{0.21}$	0.02	1.10	0.03	29.13
		polynomial	$Q=9.02 * 10^{-7} P^2 + 0.0078 P + 58.5$	0.00	1.10	0.01	47.71
		Linear	$Q=0.037 P + 31.91$	0.02	1.10	0.02	47.43

# Regional Regression Model of Mean Annual Streamflow for Ungauged Catchments

## Appendix: D Correlation coefficient and scatterplot of Region-II

```
> #Investigate possible multicollinearity: values must be <0.8
> round(cor(lowerwash),2)
      Q      A      P      E      S
Q 1.00  0.61  0.24  0.37  0.08
A 0.61  1.00 -0.26 -0.40 -0.65
P 0.24 -0.26  1.00  0.39  0.56
E 0.37 -0.40  0.39  1.00  0.69
S 0.08 -0.65  0.56  0.69  1.00
> panel.cor<-function(x,y,digits=2,prefix="",cex.cor,...)
```



# Regional Regression Model of Mean Annual Streamflow for Ungauged Catchments

---

## Appendix E: Sample R code used in development of regression model

```
1 #Importing the file of interest
2 upperawash=read.csv(file.choose(),header=T)
3
4 #Investigate possible multicollinearity: values must be <0.8
5 round(cor(upperawash),2)
6 panel.cor<-function(x,y,digits=2,prefix="",cex.cor,...)
7 {
8   txt=upperawash
9   usr<-par("usr");on.exit(par(usr))
10  par(usr=c(0,1,0,1))
11  r<-abs(cor(x,y))
12  txt<-paste(c(r,0.123456789),digits=digits)[1]
13  txt<-paste(prefix,txt,txt,sep="")
14  text(0.5,0.5,txt)
15 }
16 #scatterplots: how the variables are related to each other
17 plot(upperawash)
18 pairs(~Q+A+P+E+S, data=upperawash, main="Region-I scatterplots",
19       lower.panel=panel.cor)
20 pairs(~Q+A+P+E+S, data=upperawash, main="Region-I scatterplots",
21       lower.panel=panel.cor)
22 pairs(~Q+A+P+E+S, data=upperawash, main="Region-I scatterplots",
23       lower.panel=panel.cor)
24 pairs(~Q+A+P+E+S, data=upperawash, main="Region-I scatterplots",
25       lower.panel=panel.cor)
26 pairs(~Q+A+P+E+S, data=upperawash, main="Region-I scatterplots",
27       lower.panel=panel.cor)
28 pairs(~Q+A+P+E+S, data=upperawash, main="Region-I scatterplots",
29       lower.panel=panel.cor)
30
```

## Regional Regression Model of Mean Annual Streamflow for Ungauged Catchments

---

```
31 # Check multicollinearity using VIF and Tolerance
32 library(car)
33 vif(upperawash)
34 write.table(vif(upperawash), "vif.csv", col.names=TRUE, sep=",")
35 tolerance<-1/(vif(upperawash))
36
37
38 #Regression model with A
39 results=lm(Q~., data=upperawash)
40 results
41 summary(results)
42
43 #checking the residuals are normally distributed
44 hist(resid(results),main='Histogram of residuals'
45 |,xlab='standardised
46     residuals',ylab='frequency')
47 plot(results,which=1)
48
49 #plotting the confidence bounds on a scatterplot
50 plot(upperawash$A, upperawash$Q, pch=18)
51 resultsP=lm((Q)~(A), data=upperawash)
52 abline(resultsP,col="red")
53
54 #plot of residuals
55 abline(results)
56 plot(results)
57 par(mfrow=c(2,2))
58
```

## Regional Regression Model of Mean Annual Streamflow for Ungauged Catchments

---

```
67
68 #Regression model with A and S
69 results2=lm(Q~A+S, data=upperawash)
70 results2
71 summary(results2)
72
73 #checking the residuals are normally distributed
74 hist(resid(results2),main='Histogram of residuals',
75      xlab='standardised
76          residuals',ylab='frequency', type="l",col="grey")
77 plot(results2,which=1)
78
79 #plot of residuals
80 abline(results2)
81 plot(results2)
82 par(mfrow=c(2,2))
83
84 #calculation of NSE and RMSE
85
86 validation=read.csv(file.choose(),header=T)
87 library(hydroGOF)
88 sim<-validation$Qp
89 obs<-validation$Q
90 NSE(sim, obs)
91 rmse(sim, obs)
```