AFRICA CENTER OF EXCELLENCE FOR
WATER MANAGEMENT
ADDIS ABABA UNIVERSITY

Optimization of rain gauge network using Entropy, Geo-statistics and Remote Sensing: Case of Upper Awash River Basin, Ethiopia

By

Eyerusalem Berihun

A thesis submitted to Africa center of excellence for water management in partial

fulfillment of the requirements for the degree of Master of Science (M.Sc.)

Advisor: Fasikaw Atanaw (PhD)

September 2021

AFRICA CENTER OF EXCELLENCE FOR
WATER MANAGEMENT
ADDIS ABABA UNIVERSITY

Optimization of Rain Gauge Network Using Entropy, Geo-statistics and remote sensing: Case of Upper Awash River Basin, Ethiopia

BY

Eyerusalem Berihun

The thesis entitled “Optimization of Rain Gauge Network Using Entropy, Geo-statistics and Remote sensing: Case of Upper Awash River Basin,” Ethiopia” by Eyerusalem Berihun is approved for the degree of Master of Science in Hydrology and water resource management.

Approved by Board of Examiners

Dr. Fasikaw Atanaw	____________	______________
Advisor	Signature	Date
____________________	______________	_______________
Internal Examiner	Signature	Date
______________________	_______________	_______________
External Examiner	Signature	Date
_______________________	______________	_____________
Chairman	Signature	Date

Declaration

This thesis entitled "Optimization of Rain Gauge Network Using Entropy, Geo-statistics and Remote sensing: Case of Upper Awash River Basin, Ethiopia" is my original work and has been not presented or published for a degree in other universities and that all sources of materials used in this thesis work are duly acknowledged.

Eyerusalem BerihunTarekegn ________________________

Abstract

An adequate rain gauge network gives immediate and precise rainfall data that are crucial for the effective and economical water resource development and management. This study was aimed at assessing the adequacy of the currently existing rain gauge network and optimization of the required rain gauge network in the Upper Awash river basin, central-east part of Ethiopia. A proposed model composed of Multi-Criteria Decision Analysis (MCDA), Kriging-based Geo-statistics, remote sensing and entropy methods were used. These methods were applied to obtain an optimal rain gauge network in 22,664 square km. currently; there are 16 existing rain gauge stations in the study area. A monthly average rainfall data was used for preliminarily checking the adequacy of the existing rain gauge network. The MCDA method with the Analytical Hierarchical Approach (AHP) was used for selection of a suitable site for rain gauge stations considering land use, slope, road accessibility, drainage network and orographic effects. The results from the preliminary checking for adequacy revealed that the existing rain gauging stations are not adequate enough; with estimation error of 7% demanded 148 additional rain gauges. The rain gauge suitability mapping using Multi Criteria Decision Analysis (MCDA) was used for locating the new additional stations. Considering maximum redundant information to be 60%, an additional 136 rain gauge stations were proposed. Finally by using 136 RGS again checked maximum redundant information then out of those stations only 49 rain gauge stations gives priority to be installed and an existing rain gauges of 16 totally 65 stations were required in the Upper Awash river basin . The results of the study showed that a combined use of the MCDA, the geo-statistical, remote sensing and entropy methods can highly improve the optimal spatial distribution and number of rain gauges in a network.

Key words: Rainfall, Rain Gauge Network, Kriging, Entropy, Variogram, Awash

Acknowledgment

I want to confess that I never do anything entirely by myself. Many people contributed their knowledge and ideas to accomplish my paper. So I would like to take this opportunity to thank them for their generous help.

First of all, I would like to thank my **Almighty GOD** for his merciful and endless guidance throughout my life and for giving me a supportive and loving family.

Next to God I would like to express my sincere thankfulness and deep appreciation to my advisor, Dr. Fasikaw Atanaw, (for his guidance, precious advices, encouragement and corrections throughout the research work without which succeeding would be too challenging).

I would like to say thanks to both Addis Ababa University and Axum University for providing me all the necessary arrangements for my master's degree study. In addition, I appreciate the National Meteorological Agency and to my host company ministry of water irrigation and electricity for giving me their genuine hospitability.

Some special words of gratitude go to my friends Tarik, Mami, Redu, and Eshet who have always been a major source of support and funny when things get a bit discouraging. Thanks, guys, for always being there for me.

Lastly, I would like to say thank you to all colleagues, friends and relatives.

Table of Contents

List of tables

List of figures

Abbreviations and Acronyms

AHP	Analytical Hierarchal Process
ASE	Average Standardized Error
ASL	Above Sea Level
BCM	Billion Cubic Meter
CSA	Central Statistical Agency
DEM	Digital elevation model
DSS	Decision Support System
FAO	Food and Agricultural Organization
FDRE	Federal Democratic Republic of Ethiopia
GIS	Geographic Information Systems
GTP	Growth and Transformation Plan
ITCZ	Inter tropical Convergence Zone
Km	Kilometer
M	Meter
MCDA	Multi-Criteria Decision Analysis
MoWIE	Ministry of Water Irrigation and Electricity
NMA	National Meteorological Agency
PET	Potential evapo-transpiration
RGN	Rain Gauge Network
RGS	Rain Gauge Stations
RS	Remote sensing
UARB	Upper Awash River Basin
WMO	World Meteorological Organization
Yr	Year
^{O}C	Degree Celsius

1. Introduction

1.1 Background

Because the climate and weather of an area have a significant impact on most hydrologic processes, hydro meteorological data play an important role in hydrological research. Hydrological data are essential for a variety of tasks, including water resource assessment, climate change impact, and flood forecasting. It is also argued that without hydro meteorological measurements, comprehension of the rainfall-runoff phenomenon and statistical analysis of prior hydrological data, which are essential for preventing floods and disasters, is impossible (Hong et al., 2016). Data collection is critical in hydrology and water resources because it is an activity that creates information about the past and current conditions of water systems, which can then be used to make educated decisions despite the limited number of sensors available (www.mdpi.com/journal/entropy).

Rainfall is the most important type of meteorological data which is an important input for effective planning, designing, operating and managing of water resource projects. It is used in various water resources management tasks such as the assessment of potential areas for irrigation and water harvesting structures, flood frequency analysis and forecasting, water budget analysis, stream flow estimation, and design of hydraulic structures. Rainfall is without a doubt the most important component of the hydrological cycle, causing significant geographical and temporal variations. A rain gauge network is commonly used to measure rainfall at specific locations. In watershed-scale water resource planning and management, regional rainfall estimation is applied. A well-distributed rain gauge network with adequate rain gauges is required to produce an accurate and exact estimate of regional rainfall. A rain gauge network's main goal is to deliver needed rainfall data with the appropriate accuracy in a logical and cost-effective manner (Hackett 1966). A comprehensive analysis of the number and location of rainfall gauges is required to meet the rain gauge design objectives (accuracy in estimations at the lowest cost) (Bras1990). Various statistical theories are used in the literature for this aim. To estimate the ideal number of rain gauges in a watershed, Patra (2001) employed the coefficient of variance and the permissible percentage of error metrics. Rodriguez-Iturbe and Mejia (1974) used a random design technique to create design curves for estimating regional rainfall mean. Shih

(1982) proposed a multi-step approach based on rain gauge station covariance for rain gauge design.

Rain gauge networks are typically set up to aid in the direct measurement of rainfall data that can be used to illustrate a catchment's spatial and temporal rainfall patterns (Khairulet al., 2016). An appropriate rain gauge network provides timely and accurate rainfall data, which is essential for the efficient and cost-effective construction of hydraulic structures. This aids in reducing the hydrological and economic hazards that many water resource issues entail (Adhikary et al., 2015). Inadequate rainfall data in a catchment area has a significant influence on the ability to simulate, anticipate, and plan for catastrophic events such as floods and droughts, which have apparent detrimental health and socioeconomic consequences (Hong et al., 2016The hydrological cycle is the foundation for hydrologists' understanding of water sources at or beneath the Earth's surface, as well as its subsequent transit through numerous channels back to the oceans' primary storage. Quantifying the amount of water in different phases of the cycle and measuring the pace of water transfer from one phase to another within the cycle are two of the most difficult tasks for hydrologists. As a result, a key goal of engineering and scientific hydrologists is to measure the components of the cycle. For the development and management of a country's water resources, nationwide schemes to measure hydrological variables are now considered essential. As a result, responsibility for measurement stations is concentrated in central or regional government agencies, and hydrological measurements are given careful consideration. Cost-benefit analyses of data collection are also being conducted, and scientific planning is being advocated to ensure that the best networks are in place to give the necessary information (Elizabeth M. Shaw, Keith J. Bevin, Nick A. Chappell and Rob Lamb).

Hydrologists are frequently called upon to estimate the catchment's average rainfall and/or point rainfall at unknown locations based on observed sample measurements at nearby places. Without an appropriately planned rain gauge network, this task will not be representative. Most of the rain gauge networks used for hydrological studies is sparsely distributed and thus unable to provide adequate rainfall estimation for effective hydrological analysis.

The inclusion of erroneous rainfall data results in a serious design flaw, rendering the basins incapable of predicting changes in climate and land use/cover, resulting in incalculable human and property losses (Adhikary et al., 2015).Even though the rainfall monitoring networks at the country scale are sparsely distributed, many types of research have been conducted to understand the spatial and temporal variability of rainfall using the existing rain gauge stations in the country. The previous researches on the rainfall variability have been done on different spatial and temporal scales (Osman and Sauer born (2002)) Therefore, assessment of the adequacy of the locally existing rain gauge networks based on international standards for an optimum rain gauge network design, selection of the most suitable sites for installing the rain gauge stations, and evaluating the importance of each rain gauge in giving an accurate rainfall data is necessary. Moreover, designing an optimal rain gauge network that gives maximum information with a minimum error is also necessary for an effective operation of water resource development studies (Shafiei et al., 2014).

1.2 Statement of the problem

The lack of accurate rainfall data in developing country like Ethiopia is a major difficulty during the design and planning of water resource developments, resulting in overestimation and underestimating of hydraulic structures, as well as too many hydrological and economic risks. Ethiopia has twelve major river basins, one of which is the Awash Basin, which is also the most heavily used. The natural flow regime and the geographical and temporal variability in the basin are poorly understood and remain poorly defined in this data-scarce basin with a restricted number of gauging stations. Due to a lack of data and information, water resource planning and management decisions have been limited.

Optimal design of rain gauge networks in large drainage basins is very important in order to obtain rainfall data that has an adequate representation of the catchment rainfall characteristics. When a comprehensive rainfall dataset from a catchment area is thoroughly compiled and densely collected over a long period of time, water resource planning, designing, operating and management for the study area is more likely to be developed and accomplished (Yeh et al., 2011). Some of the problems in existing hydraulic structures within the river basin are described with evidence from different researcher's study. Haregeweyn et al. (2006) conducted a survey and review of the general characteristics and difficulties of 54 recently created reservoirs, as well as the characteristics of their individual catchments, to solve these issues. The majorities of the reservoirs are at risk of insufficient inflow, excessive seepage, and silt deposition, according to the report. These problems were mainly attributed to the use of inaccurate rainfall data, poor database on hydrology and sediment yield, and the lack of it was, therefore, the intention of this study to provide an optimal rain gauge network and their suitable location in the Awash River basin that could result in accurate and vital information for water resource developments. Adequate distribution of rain gauging stations held to predict the influence of the temporal and spatial variation of rainfall and the impact of the climate change within the river basin.

1.3 Objective of the study

1.3.1 General objective

The main objective of this study is to determine the optimum number and location of rain gauge stations within the Upper Awash River basin.

1.3.2 Specific objectives

The specific objectives of this study are:

- To check the adequacy of existing rain gauge network within the Upper Awash River Basin.
- To develop a set of multi criteria for location of rain gauge suitability using multi-criteria decision analysis.
- To optimize of the number of candidate rain gauge stations in the Upper Awash River basin using entropy approach.

1.4 Research questions

- Which location has inadequate rain gauge station?
- How is the spatial variation of the existing rain gauge stations?
- Where do we put the new proposed stations?

1.5 Research hypothesis

1.5.1 Null hypothesis

Upper Awash River Basin has enough rain gauge stations

1.5.2 Alternate hypothesis

Upper Awash River Basin doesn’t have enough rain gauge stations

1.6 Significance of the study

The results of the study would be of great importance to various sectors in which those a direct relevance to the use of rainfall data has. Those include water resource planners, Ethiopian National Meteorological Agency, research centers and communities living around the river basin. And this study will be a very important input to the Ethiopian NMA, for expanding the distribution of the rain gauge stations found in the river basin. In addition to, the rainfall data collected using of the optimized rain gauge networks would be of great importance for the various water resource development works such as the assessment of potential areas for irrigation and water harvesting structures, flood forecasting and prevention and design of hydraulic structures planned in the river basin. The optimized rain gauge networks are also important to research centers, in which rainfall data is the prime input variable for their research work.

1.7 Scope of the Study

The scope of the study is focused only on the determination of the optimum number of rainfall stations and their location in the Upper Awash Basin.

2 Literature review

The assessment of this invaluable resource, which takes into account the identification of sources and the evaluation of their capacity, reliability, and quality, implying the measurement and collection of data of interest, is one of the most important elements of effective water resource planning and management (Mishra and Coulibaly, 2009). Monitoring networks are built and occasionally optimized for decision-making based on water management objectives for this purpose (Loucks et al., 2005). "The objective of a monitoring network is to ensure a density and distribution of stations in a region such that, by spatial interpolation between datasets at different stations, it will be possible to determine the characteristics of the basic hydrological and meteorological elements anywhere in the region with sufficient accuracy," the World Meteorological Organization (WMO) stated in 1981. (Van der Made 1988, p.20).

Rain-gauge networks provide estimates of local rainfall, which is essential for hydrological applications. As a result, it's critical to measure a rain-gauge network's performance and assess the contribution of each rain-gauge to the overall accuracy of areal rainfall estimation at basin scale. The majority of hydrologic designs are based on rainfall. Many hydrological applications rely on the estimation of average rainfall over a basin region based on data collected at many rain gauges (Chua & Bras 1982; Bastin et al. 1984). The necessity to reliably capture the area average rainfall in basins drives the design of rain-gauge networks. Furthermore, good understanding of spatiotemporal rainfall is required for effectively predicting discharge and determining other hydrological processes in rainfall-runoff modeling (Beven 2001). A wide number of studies have found that the density and distribution of rain-gauge networks can have a considerable impact on simulated discharge, sediment, and other forms of watershed responses (Seed & Austin 1990; Duncan et al. 1993; St-Hilaire et al. 2003; Chaplotet al. 2005; Birdossy & Das 2008). For example, Anctil et al. (2006) found that when the area average rainfall is computed by a number of rain-gauges less than a minimum threshold value, model performance rapidly degrades. They also discovered that some rain gauge network combinations produce better estimates of area rainfall than using all of the basin's existing rain gauges.

Rain-gauge network optimization is now considered outdated, as weather radars give greater spatial and temporal resolution rainfall data. Nonetheless, there are a number of potential reasons of mistake in radar rainfall measurements (Steiner ET al.1999; Jayakrishnan et.al.2004; Abdallah

& Alfred Sen 2010). The impact of radar rainfall estimating error on hydrological model outputs has been demonstrated in several studies (e.g. Borga ET al.2006). Estimates from radar can be skewed (because of a bright band, for example). Such biases might result in significant mistakes in hydrological simulation results (Berne & Krajewski 2013). The most pressing issue is that, with the exception of a few recent attempts in the Tana Basin, Ethiopia lacks radars for rainfall measurement.

2.1 Overview of rain gauge network design methods

Rain gauge networks have been developed and established all over the world for many years. Installation of new rain gauge stations is frequently required to meet the expanding demands for rainfall monitoring in particular sections of a catchment or watershed due to imbalanced regional growth and less effective design during the early phases (Wang et al., 2015). However, given the complicated environmental restrictions and limited resources, deploying an endless number of new rain gauge stations in a network is unfeasible in real-world applications. As a result of these challenges, optimal rain gauge network design is critical for achieving the best rainfall monitoring efficiency and optimizing resource use. Rain gauge station optimal configuration can be thought of as a facility location challenge (Wang et al., 2015). The goal of rain gauge network design is to place a particular number of rain gauge stations at specific places throughout the network to meet a catchment's or watershed's water management objectives.

The design and evaluation of rain gauge networks is associated with a number of key challenges, which range from establishing proper temporal and spatial scales to defining their scope at minimum costs. In theory, hydrologists usually take into account these challenges when developing new approaches for rain gauge network design. However, in practice, it has been reported that the rainfall data collected by the existing rain gauge networks remains, in general, inadequate for understanding and explaining the dynamics of natural water resources systems (Canadian Water Resources et al., 1994; IUCN, 1980). This could be because non-scientific factors such as political and social opinions determine the criteria for establishing the final rain gauge network in practice.

Existing rain gauge networks are frequently examined to ensure that the network's design objectives are met. The evaluation's findings include the reconfiguration of the existing rain gauge network, which may entail a redefinition of the network's size and scope. This can result in the addition of additional rain gauge stations through a network augmentation process in places where rainfall data cannot be adequately inferred from the existing network, or the elimination of redundant rain gauge stations through a network rationalization process (due to redundancy or uselessness of the collected data) (St-Hilaire et al., 2003; Mishra and Coulibaly, 2009). In general, the same procedures that are used to build monitoring networks are also used to evaluate them.

2.2 Fundamental aspects of geo-statistics

The statistical study of natural occurrences that can be generally described by a distribution of one or more variables in space is referred to as "geo-statistics." Regionalized variables are the name given to these variables (Journel and Huijbregts, 1978; Webster and Oliver, 2007). A regionalized variable has the distinct property of being able to take values based on its spatial position (Chebbi et al., 2011). Geostatistics is based on the notion of regionalized variables, which allows for the modeling of a variable's geographic variability based on the spatial dependence between surrounding data. In Geostatistics, the degree of spatial dependency is often stated by a Variogram (also known as semi Variogram), which contains the structural (spatial variability) information needed for a regionalized variable. A Variogram is a mathematical function that quantifies the spatial autocorrelation in regionalized data by measuring the distance and direction between two sites (Webster and Oliver, 2007).

2.3 Overview of entropy theory

Clasius was the first to adopt the term "entropy" as a scientific notion in thermodynamics in the 1850s. Boltzmann later developed a probabilistic interpretation of the notion in the framework of statistical mechanics in 1877. Planck identified the clear link between entropy and probability in the early 1900s (e.g., Harmancioglu and Singh, 1998; Singh, 2013). Shannon (1948a, b) extended the concept to a number of basic problems in coding theory and data transmission, providing an economic interpretation of the features of long sequences of symbols. Shannon finally developed, with his remarkable contributions in this area, the basis of modern information theory. Later, Jaynes (1957a, b) re-evaluated the method of maximum entropy and applied it to a variety of problems.

2.3.1 Concept and application of entropy method in rain gauge network design

When applying engineering concepts to the problem of data collection, it is necessary to receive a minimum number of signals in order to gather the largest amount of data. Redundant data isn't helpful, and it only adds to the confusion. As a result, the cost of getting data will only rise. A current monitoring network should be decreased or rationalized in the case of redundant information, and increased in the case of a lack of information (Mogheir and Singh, 2002). These principles are at the heart of the field of communications, and they apply equally to hydrologic data collection, which is fundamentally a form of communication with nature. Because the quantity of information obtained is equal to the amount of uncertainty reduced by making observations, the entropy criterion indirectly assesses the information richness of a set of data (Harmancioglu and Yevjevich, 1987).

In the realm of hydrology and water resources, entropy theory has been applied in a variety of ways (e.g., Singh 1997; Singh, 2013). The entropy concept, in particular, can be utilized to construct appropriate network design criteria in the rain gauge network based on quantitatively expressed information expectations and information availability. To construct the rain gauge network, the entropy theory can be employed to set up an information-based design strategy. This is justified by the fact that a network of rain gauges is essentially an information system (Krstanovic and Singh, 1992a, b). In fact, previous research into the use of the entropy principle in rain gauge network design has yielded promising results, particularly in terms of selecting technical design features like optimal rain gauge station locations and identifying redundant stations (e.g., Husain, 1989; Krstanovic and Singh, 1992a, b; Al-Zahrani and Husain, 1998; Yoo et al., 2008; Karimi-Hosseini et al., 2011; Ridolfi etal., 2011; Vivekananda and Jagtap, 2012; Li et al., 2012; Wei et al., 2014; Xu et al., 2015; Werstuck and Coulibaly, 2016; Xu et al., 2017). A number of studies have used the entropy theory in the design of monitoring networks in various fields, including water quality monitoring network design (e.g., Alpaslan et al., 1992; Harmancioglu and Alpaslan, 1992; Ozkul et al., 2000), and groundwater quality monitoring network design (e.g., Alpaslan et al., 1992; Harmancioglu and Alpaslan, 1992; Ozkul et al., 2000) (e.g., Mogheir and Singh, 2002; Mogheir et al., 2009).

2.3.2 Limitations of entropy method in rain gauge network design

Previous research has shown that the entropy technique is also a potential way for solving rain gauge network design challenges since it allows for the creation of a network based on quantitative information assessment. However, the entropy method has significant drawbacks that must be considered (Harmancioglu et al., 1994). The availability of accurate and sufficient data is required for a thorough evaluation of rain gauge network features using the entropy method. Numerical challenges and, as a result, inaccurate findings are common in applications with insufficient data. When examining spatial and temporal frequencies in the multivariate situation, for example, the covariance matrix's proprieties pose a significant numerical challenge. Entropy measures cannot be derived consistently when the determinant of the matrix is very small, as the matrix becomes ill-conditioned. This happens a lot when the sample size is very tiny (e.g., Harmancioglu and Alpaslan, 1992). On the other hand, when it comes to data availability, the question is how many data points are sufficient. Determining when a data record is sufficient is particularly difficult (Harmancioglu et al., 1994). The occurrence of gaps in data series limits entropy estimations, especially in the time domain, preventing temporal design after certain time lags (Harmancioglu et al., 1999). The similar problem exists in network space/time architecture, resulting in incorrect findings.

The mathematical definition of entropy ideas for continuous variables is another issue with the application of the entropy theory. Shannon's basic concept of entropy is derived for a discrete random variable, and the difficulty of picking discretizing class intervals to approximate probabilities with class frequencies arises when this definition is extended to the continuous case. Distinct measurements of entropy vary with class intervals, resulting in a different base level or scale for evaluating uncertainty for each set of class intervals. As a result, for each selected value of class interval, the same variable is assumed to have distinct entropy values. It may even take on negative values, which contradicts the entropy function's positive characteristic in principle. Defining confidence levels for entropy metrics, particularly for Tran's information, could be one way to solve the aforementioned flaw. This is still a problem that needs to be looked at. An a priori probability distribution function can fix this, however there are disagreements over which a priori distribution to use, therefore the problem remains unsolved.

2.4 Optimization method

In general, optimization refers to the study of issues in which the goal is to minimize or maximize a real function by selecting values for real or integer variables from a permitted set in a systematic manner. In network architecture, the basic application of optimization is to maximize information while decreasing cost (Mishra and Coulibaly, 2009). According to Langbein (1979), rain gauge network design does not have to be based on formal optimization strategies such the lowest cost of achieving data accuracy. To accommodate a combination of design criteria, a design can be built on judgmental analyses. According to the available literature, most previous studies used a sequential trial and error technique to reduce Kriging variance in rain gauge networks and optimize the network (e.g., Shamsi et al., 1988; Loof et al., 1994; Papamichail and Metaxa, 1996; Haggag et al., 2016). However, a few studies used optimization techniques such as simulated annealing, genetic algorithms, and particle swarm optimization to optimize rain gauge networks (e.g., Pardo-Igúzquiza, 1998; Barca et al., 2008; Chebbi et al., 2011; Adib et al., 2016; Aziz et al., 2014; Aziz et al., 2016; Aziz et al. Since the early work of Bras and Rodriguez-Iturbe (1976) and Delhomme (1978), who presented a methodology for rain gauge network design based on the minimization of the mean areal Kriging variance, other optimization strategies have been proposed in the literature. Pardo-Igzquiza (1998), Barca et al. (2011), Chebbi et al. (2011), Aziz et al. (2014), Adib et al. (2016), and Aziz et al. (2017) all used optimization techniques in tandem with the Kriging-based geostatistical method for rainfall network size and augmentation (2016).

Delhomme (1978) used a technique called the fictitious point approach to find the best position for rain gauges, whereas Pardo-Igzquiza (1998) and Aziz et al. used an automatic optimization technique called simulated annealing (2014). Barca et al. (2008) proposed a method for determining the best site for new rain gauge stations inside an existing network of rain gauges. Kriging and probabilistic approaches (simulated annealing) were integrated with a geographic information system in the methodology (GIS). Chebbi et al. (2011) used a simulated annealing technique to analyze mono-objective criteria, assuming 1-hour rainfall intensity interpolation and erosivity factor interpolation and a single extreme rainfall event. Rainfall amounts were mostly taken in a deterministic fashion in prior research. To produce the best new rain gauge placements, a single rainfall pattern was chosen for which the average Kriging variance was minimized (Delhomme, 1978; Pardo-Igzquiza, 1998; Chebbi et al., 2011). In two recent researches, evolutionary algorithm and particle swarm optimization approaches combined with Geo-statistics were employed for rain gauge network design to reduce Kriging variance for network optimization (e.g., Adib et al., 2016; Aziz et al., 2016). The main disadvantage of the optimization method is that it frequently produces a network configuration that is not possible to deploy as the optimal network. Because it is nearly impossible to re-locate and re-install all current rain gauge stations in their new determined locations (which may occur frequently as part of the solution) using the optimization method, which entails significant financial and logistical resources. As a result of these challenges, the sequential trial and error method was developed (e.g., Shamsi et al., 1988; Loof et al., 1994)

2.5 Data requirement for rain gauge network optimization

The temporal scale of the rainfall data to be used for rain gauge network optimization is often determined by the purpose for which the rain gauge network is being optimized. Some of the rain gauge networks are intended for water resource planning and management applications (e.g. water balance studies, storage dams, and groundwater recharge). The other sort of rain gauge network is one used for storm drainage and flood protection, such as hydrological studies required for hydraulic design of bridges, culverts, channels, and storm water networks, among other things. Long-term averaged rainfall data, such as mean monthly and mean annual rainfall depths, are the most important data from the first kind. The data required for the second type of rain gauge network are high-temporal resolution recordings (hourly and sub-hourly records), which are essential for defining the storm pattern and, as a result, the design rainfall intensity (Elsayed and Awadallah, 2016).and the methods to fill missing rainfall data are listed below

a) Normal ratio method

When the typical annual precipitation at any of the index stations differs by more than 10% from that of the interpolation station, the normal ratio method (NRM) is applied. The precipitation levels at the index stations are weighted by the ratios of their normal annual precipitation data in this method, which is expressed as (Maidment, 1993):

$$Px = \frac{1}{n}\sum_{i=1}^{n} \left(\frac{Nx}{Ni} Pi\right) \text{……………………………………………………………..Equation 2 - 1}$$

Where: Ni is the normal annual precipitation of stations at the neighboring stations

Nx is the normal annual precipitation of the index station and

P is the given precipitation of all the station

b) Inverse distance weighting method

The amount of missing precipitation in a given station can be estimated from nearby stations by giving a weighted distance value to the index stations. The precipitation in a given station is thought to be inversely proportional to the distance of nearby stations. The precipitation in a station which has missing data is calculated by (Maidment, 1993):

$$\mathrm{Px} = \frac{\sum_{i=1}^{n} \frac{1}{\mathrm{dim}} \mathrm{Pi}}{\sum_{i=1}^{n} \mathrm{dim}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{Equation 2 - 2}$$

Where: P_X - The amount of precipitation in a station with missing data

Pi - The amount of precipitation in the ith station

n - Number of stations

d - Distance of an ith station from the station with missing data

m - Distance weight

The inverse distance-square method is the most commonly used one and hence the value of m will be 2 (Maidment, 1993).

c) The simple linear regression method

The simple linear regression is employed to transfer hydrologic information between two cross-correlated stations (Maidment, 1993). This method is not only used to correlate precipitation or flow of one station to another, but it is also used to relate precipitation of one station to flow of another station.

Let y_t and x_t correspond to the short and long sequence of hydrological variables respectively. Similarly, let N1 and let N1+N2 be the length of the short and long period sequence. Hence, a simple linear regression is formulated to fill or extend missing data for the short sequence y_t. The simple mathematical linear regression model between variables y_t and x_t can be represented by the following equation (Maidment, 1993).

$$Yt = a + bxt + \alpha\theta(1 - e^2)½\, \sigma y \varepsilon t \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{Equation 2 - 3}$$

Where: Y_t is the dependent variable (short record)

Xt is the independent variable (a long record)

a and b are population parameters of the regression

α is a coefficient

θ =1 when noise εt is added; θ = 0 when εt is not added

ρ is the population cross-correlation coefficient between yt and xt

σy is the population standard deviation of yt

εt is a normal variable with mean zero and variance 1 and uncorrelated with xt

3. Material and methods

3.1 Description of the study area

3.1.1 Location

The Awash River Basin is the fourth largest catchment in Ethiopia by area. The river rises near Ginchi and flows in a north easterly direction through the northern extension of the Rift Valley to eventually discharge into Lake Abe near the Djibouti boarder, a distance of some 1,200km. The geographic location of the basin is between latitudes of $7^0$53’N and 12^0N and longitudes of $37^0$57’E and $43^0$25’E (Taddese et al., 2006). The uplands are further divided into the Upper Basin or the Koka catchment upstream of Koka Dam, the Eastern Catchment and the Western Catchment based on location in the basin.

The Upper Awash Basin is located in Ethiopia's highland plateau, at elevations of 1500 to 3000 meters above sea level. It's upstream from the Koka Dam. The Upper Awash catchment's land use consists primarily of farmed agricultural land, forest land, rural areas, and towns. The Upper Awash River covers the river section from its source up to Koka Reservoir. The Study area drains a catchment area close to 22,664 km2 and the length of the river up to Koka is around 220 km (Hal crow, 1989).

The major tributaries to the upper Awash are Akaki and Mojo rivers. Akaki River starts from the mountainous areas of the northern part of Addis Ababa and join the main Awash River between Melka-Kunture and Melka-Hombole gauging stations. Mojo River, the other main tributary to Awash, originates from the high lands northeast of Addis Ababa. It drains a catchment area close to 1,900 km2 and travels a total length of about 105 km before joining Awash. According to currently available Ethio-GIS five sub catchments are located within the Upper Awash River Basin

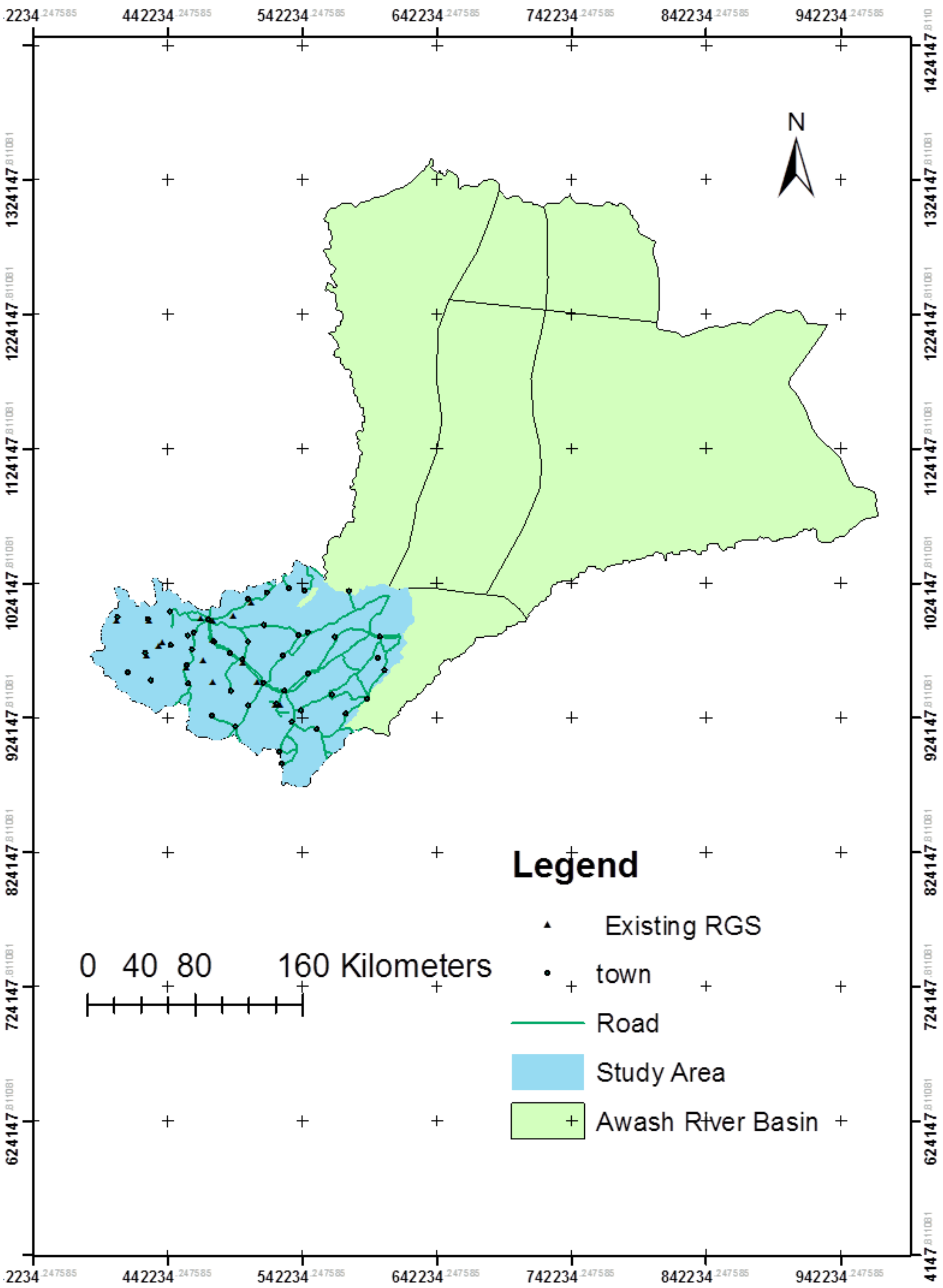


Fig 3- 1: Location of Upper Awash River Basin

3.1.2 Population

The Awash River Basin found in the Oromia region the total population of Oromia National Regional State was 35 million according to Central Statistical Agency, 2018. The 2011 census reported the population of Oromia as 35,000,000; this makes it the largest regional state covering

286,612square kilometers. The capital Addis Ababa is situated in the Western part of the Upper Awash River Basin. The urban population of Addis Ababa has grown from 0.5 to 3 million from 1950 to 2008, and is expected to rise to between 4.5 and 7.5 in 2030.

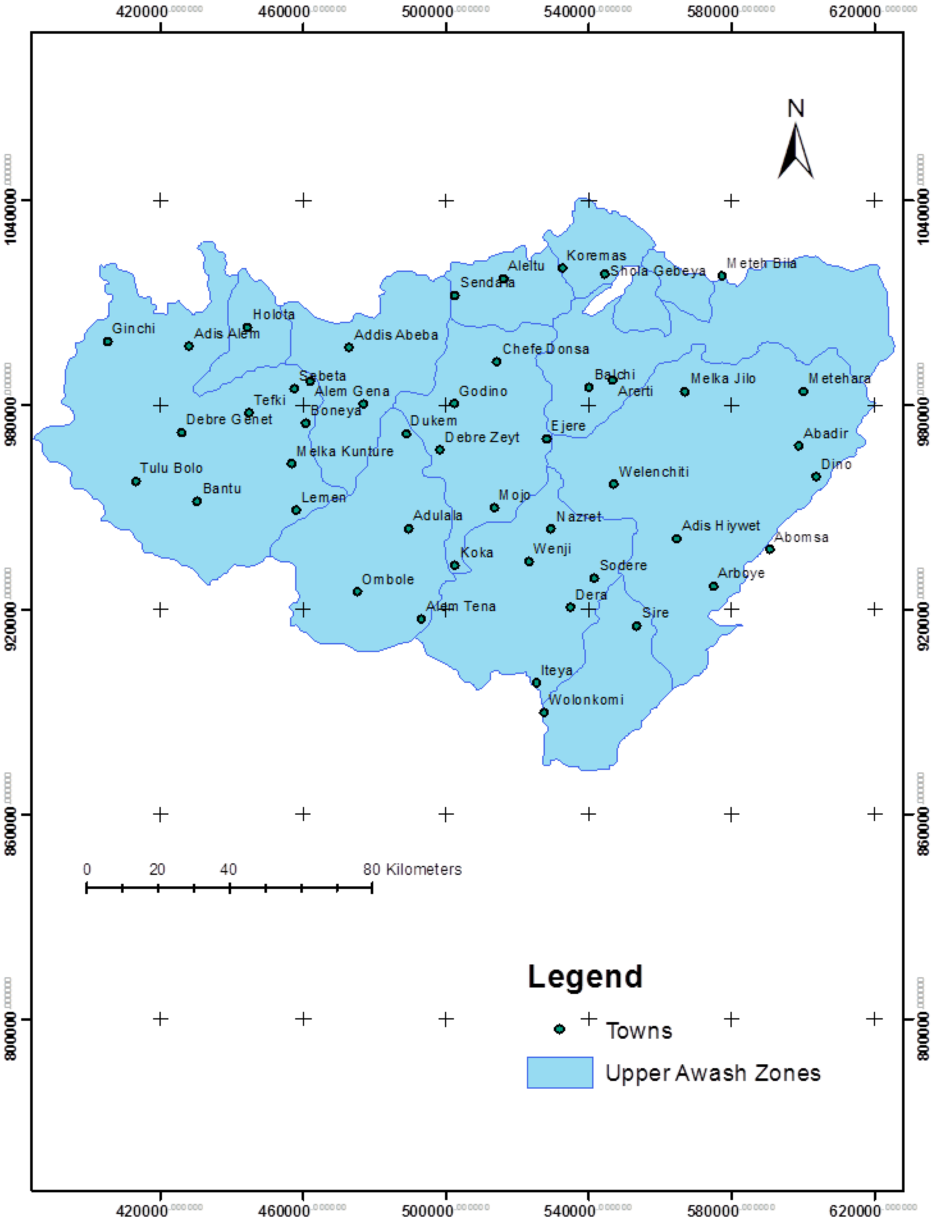


Fig 3- 2: Map of zones and towns in the Upper Awash river basin

3.1.3 Hydrological and meteorological conditions

A. Hydrological conditions

The Awash basin is divided into three sub catchments upper, (upstream from Koka Dam station) middle, (between Koka and Awash station) and lower basins (between Awash and Tendaho station) based on climatological, physical, socio-economic, agricultural, and water resources characteristics (Awash Basin Authority. Awash River Basin Strategic Plan Main Report; Awash Basin Authority: Amara, Ethiopia, 2017.) The rainfall pattern in the basin is predominantly uni modal with the main rains occurring from June to September (locally known as Kirmet), the rest of the year is mainly dry except small rains during March to May (locally known as Belg). According to the Awash Basin Authority the months of April to June are critical for irrigators in the basin as it is peak irrigation period for both small-scale farmers and large-scale

The Upper Valley of the Awash River rises roughly 3 kilometers above sea level on the high plateau west of Addis Ababa. From the western and northern parts of the Upper Awash River Basin to its southern side, the elevation lowers dramatically. In a 100-kilometer stretch from north to south, the height drops by more than 1400 meters. The Awash River drains the Becho plains to the east, where it is joined by a number of tributaries before entering Lake Koka. The Kebena, Great and Little Akaki, and Mojo Rivers are all key tributaries upstream of the Koka reservoir.

Generally, plateaus between 3000m and 2,500m receive 1,400 - 1,800 mm yr-1 and regions with altitudes ranges from 1600 to 2500m receive 1 000-1 400mm yr-1. The rainfall distribution is bimodal in this region, with a main rainy season from June to September and the short rainy period in March and April. Although the rainfall intensity is high in the region, the potential evapo-transpiration (PET) in the Upper Valley is higher, for instance at Koka is 1810 mm almost twice of the annual rainfall.

B. Meteorological conditions

The Awash River Basin's climate ranges from humid subtropical in central Ethiopia to desert in the Afar lowlands (Daniel 1977, Lemma 1996). The Awash River Basin is also heavily influenced by the intertropical convergence zone's movement (ITCZ). ITCZ causes two rainy seasons, a shorter one around March ('Belg') and a longer one between June and September ('Kiremt'), which partly overlap during its advance northwards in March/April and its retreat southwards. The rainy season in Ethiopia is bimodal in eastern Ethiopia and nearly unimodal in western Ethiopia. The dry season, known as 'Bega,' lasts from October to March (Seleshi, Yilma; Zanke, Ulrich (2004-06-30)). In the Rift Valley, semiarid to arid climates prevail. The highlands, on the other hand, receive more than 1600 mm of rain for about six months of the year. Malte Knoche, Christian Fischer, Eric Pohl, Peter Krause, and Ralf Merz (2014)

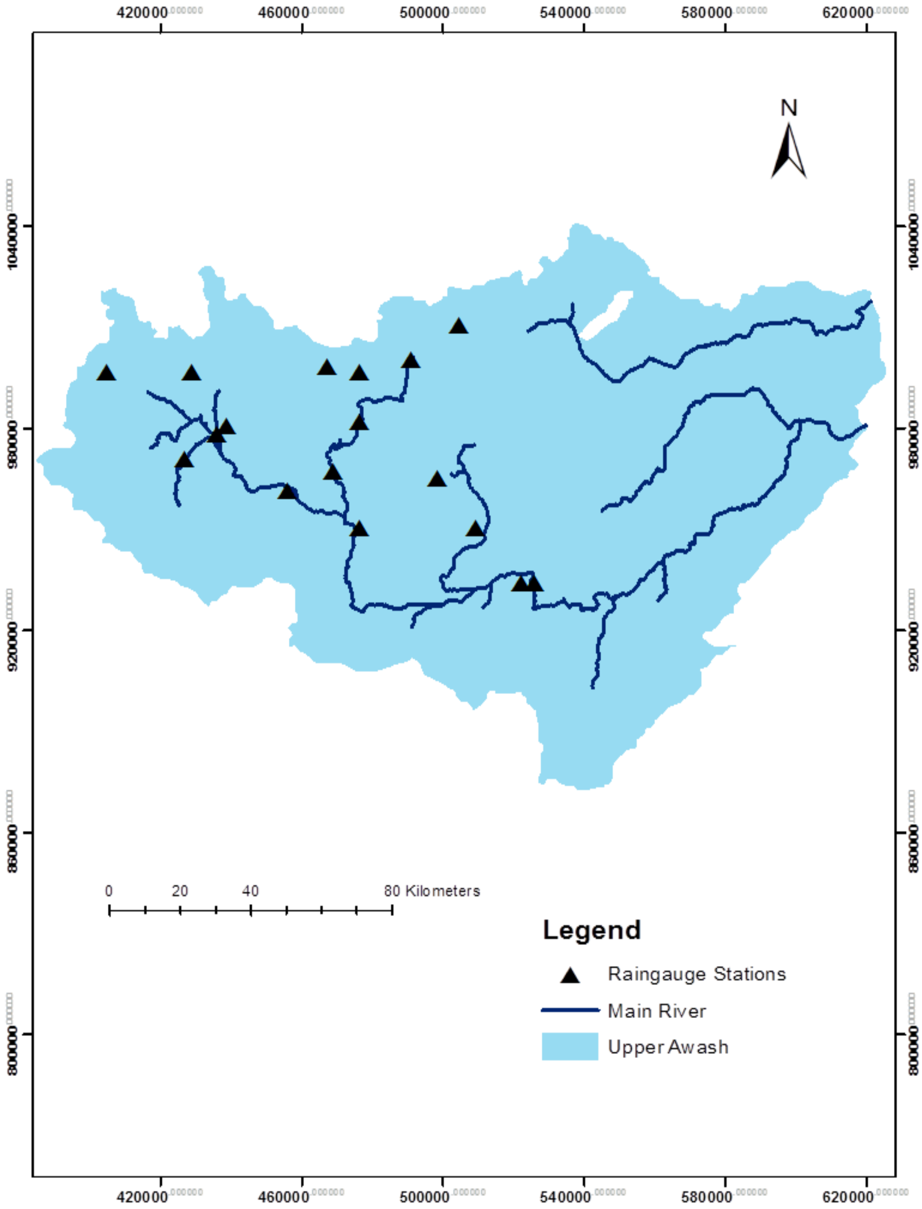


Fig 3- 3: The upper awash river basin and rain gauge stations

3.1.4 Topography

The Awash River begins in the middle Ethiopian highlands, at a height of around 3000 meters above sea level. The river enters the Great Rift Valley at a height of 1500 m after running southeast for about 250 km, and then follows the valley for the rest of its length to Lake Abe on the Djibouti Republic's border, at an altitude of about 250 m. The total length of the river is about 1200 km and its catchment area is 113,700 km^2 but my study area is 22,664km^2.

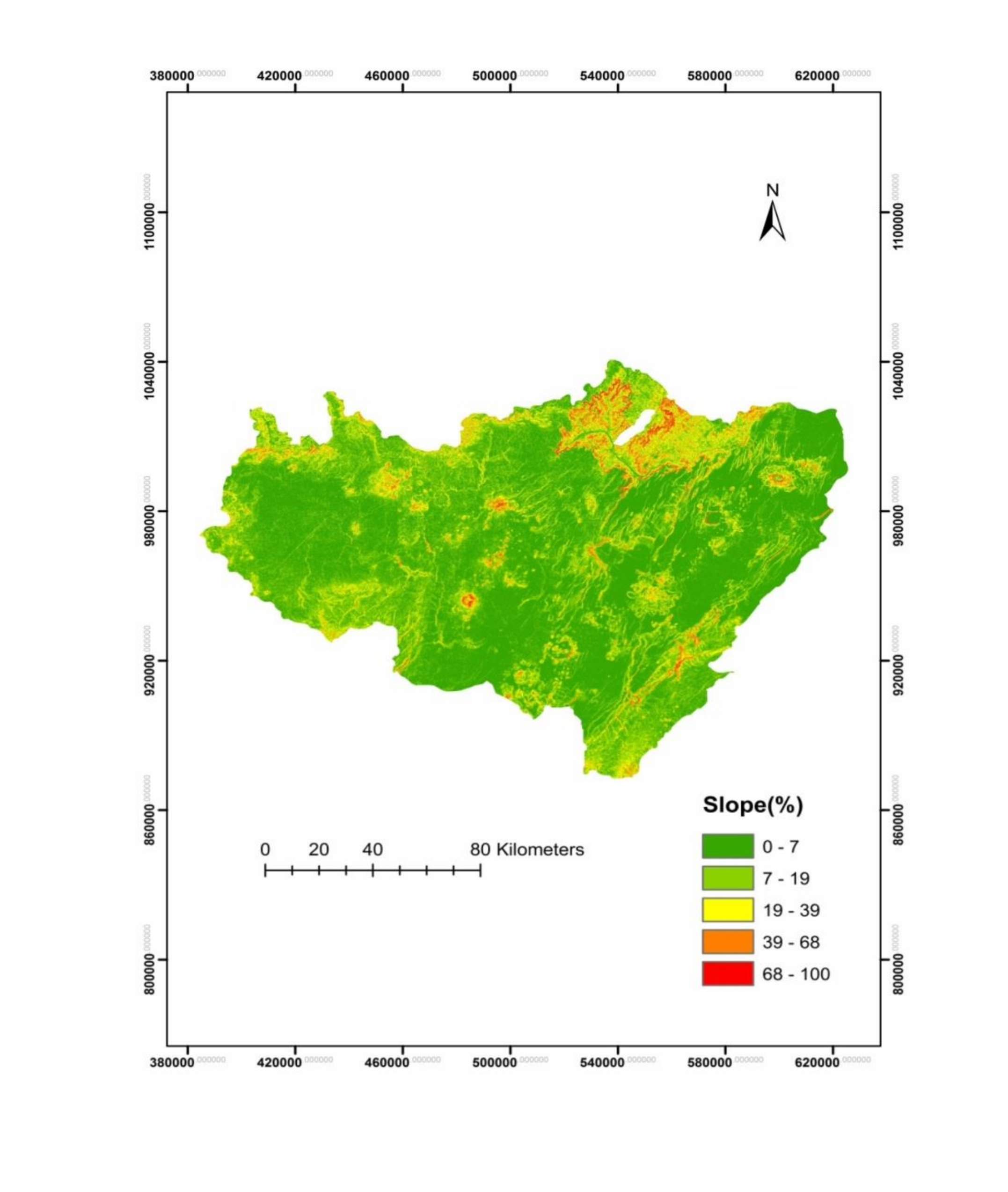


Fig 3- 4: Map of slopes in the Upper Awash River Basin

3.1.5 Land use and soil type

The Upper Awash catchments have mostly farmed agricultural area, grassland, and forestland, as well as rural and urban communities. 67 percent is highly cultivated, 25.5 percent is moderately cultivated, 4.5 percent is bush land, shrub land, or forested grassland, and 3 percent is urban area and alpine vegetation, according to estimates. Even within the upper Awash, there is a wide range of land usage.

In the upper most part where there is high rainfall, land use is complete in May with barley and Teff. Steeper slopes are heavily wooded with natural acacia and eucalyptus. On the lower most part, however, rainfall is too unreliable and the sparse dry acacia scrub gives way to wide stretches of bare ground with clumps of coarse grass and occasional thickets of acacia. The soil type in the upper awash sub-basin is diverse. The most common soil types are Clay, Sand, Clay-Loam, Silt-Clay -Loam, Sand-Clay, Silt-Clay (Paulos 1989). Land use and soil type have a direct impact on the flood amount, speed and potential to create damage and at the same time LULC has great effect on stabilization of rain gauges that the study should give attention for land use and land cover of the sub basin.

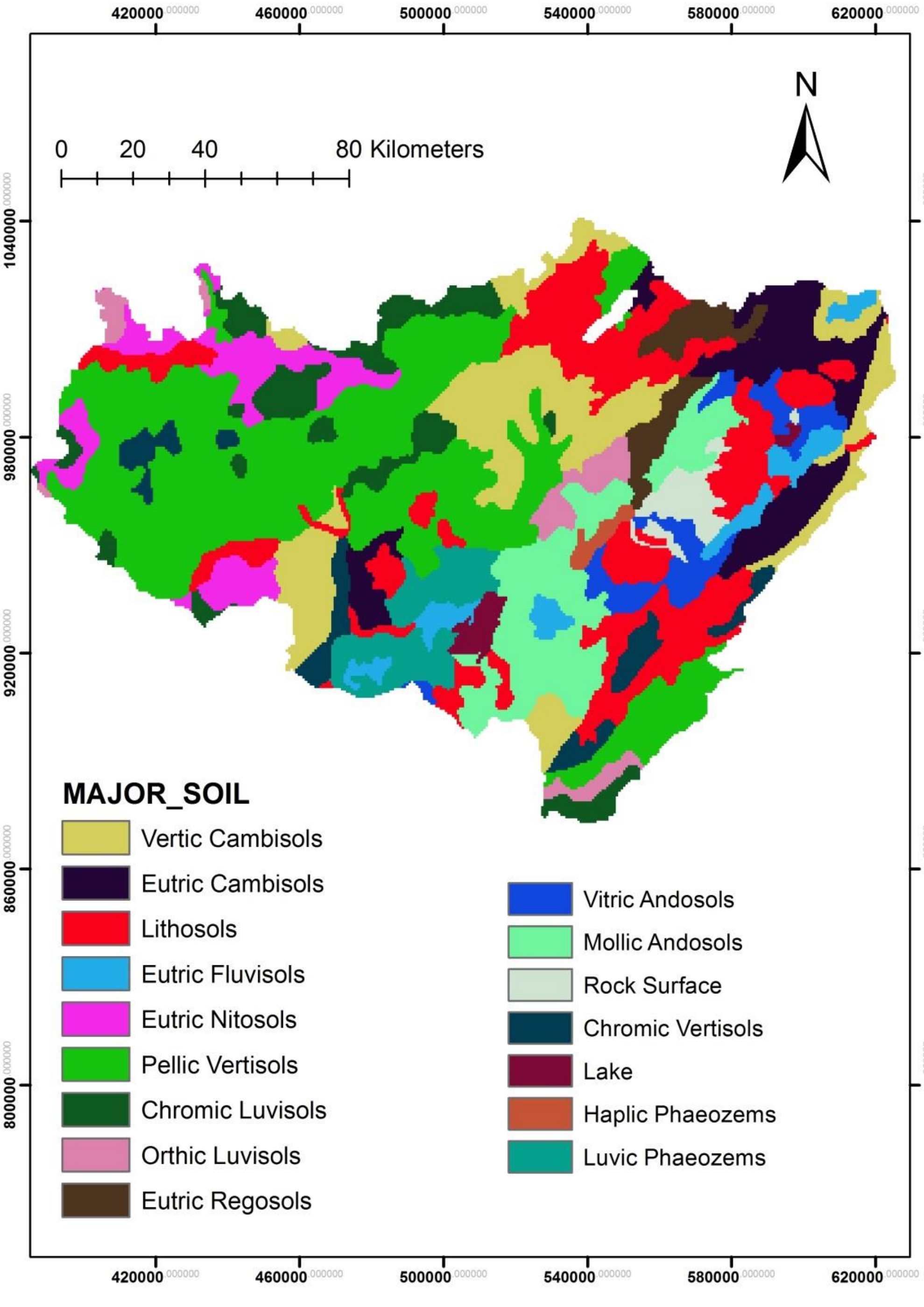


Fig 3- 5: Map of soil types in the Awash River Basin

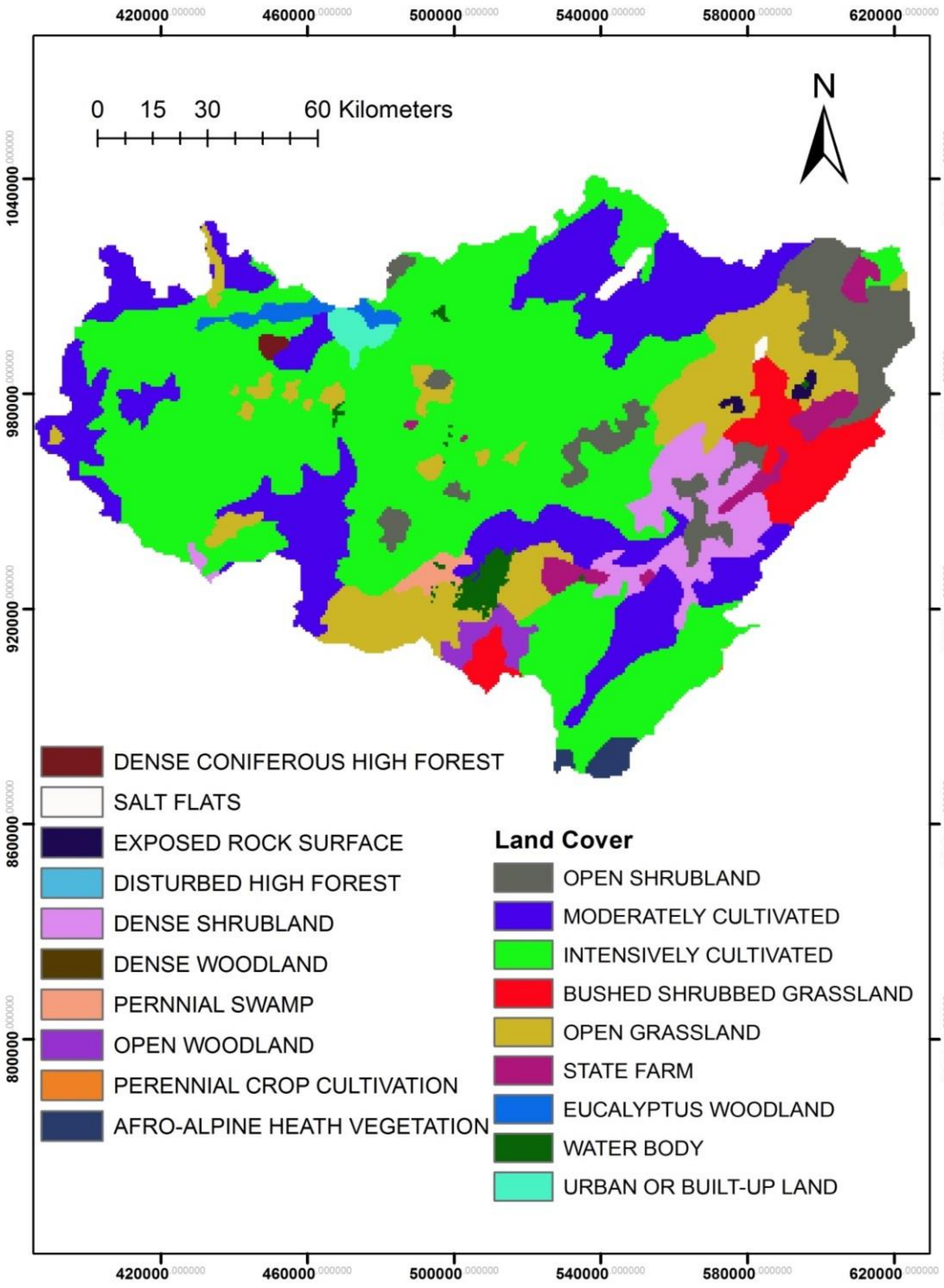


Fig 3- 6: Land use/cover map of the Awash River Basin

3.2 Data collection

For the execution of this thesis work, secondary data of rainfall collected from those of stations which are in operation till 2020. The data will be collected from the Ethiopian National Meteorological Agency (NMA), Addis Ababa main office. The basic data sets that required developing an input database for the model are climatic, stream flow, topographic and land use data. The climatic data such as rainfall, max and min Temperature, wind speed has obtained from the meteorological stations of the area. The daily stream flow data will be obtained from the Ministry of Water Irrigation and Energy. The topographic, soil, land use and other data collected from the concerning bodies. Digital elevation model (DEM) used to delineate the watershed and drainage patterns of the surface area using appropriate tools in Arc-GIS software. Sub-basin parameters such as slope gradients, elevations and stream network characteristics were derived from DEM.

Table 3- 1: Data collected for the study and their sources

Data	**Type**	**Characteristics**	**Source**	**Use**
30*30 DEM	Raster dataset	DEM 30m*30m	Ethiopian MoWIE GIS and Remote Sensing Department	Watershed delineation, slope and elevation extraction
Land use / cover	Raster dataset	30*30m resolution	Ethiopian MoWIE GIS and Remote Sensing Department	Input into rain gauge site suitability analysis
Soil data	Shape file format	1:200,000 scale	Ethiopian MoWIE GIS and Remote Sensing Department	Study area description
Roads and towns	Shape file format	Geometric type of line and point respectively	Ethiopian MoWIE GIS and Remote Sensing Department	Input into rain gauge site suitability analysis
Rainfall (mm) - 18 stations	Point measurement	Varies from 5 to 50 yr. record period	National Meteorological Agency of Ethiopia	Input in entropy analysis for determining an optimal number of rain gauges

3.3 Materials

The materials used for this research work included:

- Rainfall data of all the existing rain gauge stations within the river basin and stations.
- A Digital Elevation Model of the river basin to generate slope gradients and elevations.
- Arc GIS 10.5 for delineation of the watershed and by using satellite rainfall data to generate the rainfall data of the candidate rain gauge stations.
- R programming was used for computing Missing rain fall data and statistical parameter

3.4 Data quality control works

3.4.1 Infilling missing rainfall data

In most times meteorological stations are subjected to a short break in records, those may be due to an absence of the observer or instrumental failures. Other causes may be due to natural hazards such as landslides or due to external factors such as wars, mishandling of observed data by field personnel or accidental loss of data. Therefore, it is necessary to infill those missing data prior to the use of the data for hydrological analysis (Maidment, 1993).

Incorrect and missing values have computed where possible by estimated values based on other observations at the same station or at neighboring stations. Several researchers recommend different methods of infilling or extending rainfall data. The most commonly used methods include arithmetic mean, normal ratio, inverse distance weighting, and linear regression methods. The classical method of arithmetic-mean is widely applicable to topographically homogeneous catchments in which the variation in normal rainfall of the station with missing rainfall is within10% compared to the neighboring stations.

- R-Programing

Like other statistical software packages, R is capable of handling missing values. However, to those accustomed to working with missing values in other packages, the way in which R handles missing values may require a shift in thinking. In R-Programming first the basics thing is how missing values are represented in R and I have outlined some of the differences between missing values in R and missing values elsewhere. Finally, I have used some of the tools for working with missing values in R, both in data management and analysis by using mice package.

3.4.2 Tests for consistency

In order to discover a large long-term systematic shift in rainfall data, the double mass analysis technique is utilized in validation. A significant variation in the slope of the double mass curve indicates data inconsistency, which can be caused by a change in instrument placement, exposure, or measurement technique. It does not suggest that either era is inaccurate; rather, it implies that they are incompatible (DHV Consultants BV & DELFT, 1999). The data can be made consistent by changing the double mass curve such that it does not break. The presence of a discontinuity in the double mass plot does not automatically indicate which part of the curve has to be changed (before or after the break). Adjusting the previous part of the record such that the complete record is consistent with the current and continuing record is a typical approach. However, there may be times when the adjustment is made to the later segment, such as when the source of the inconsistency is recognized to be incorrect or when the record has been discontinued.

The annual application of the double mass curve technique is usually done on accumulated monthly (rather than daily) data. However, in some cases, the technique could be used to date the start of an instrument defect, such as a leaking gauge, using daily data. The correction should be applied to all data intervals after a discrepancy has been found (DHV Consultants BV & DELFT, 1999).

In this study, the double mass curve is used to examine the consistency of all of the stations' annual rainfall data, and any discrepancies detected within the rainfall datasets are corrected by modifying the slope of the curve. In addition, the double mass curve is utilized to choose the optimal approach for filling in missing rainfall data.

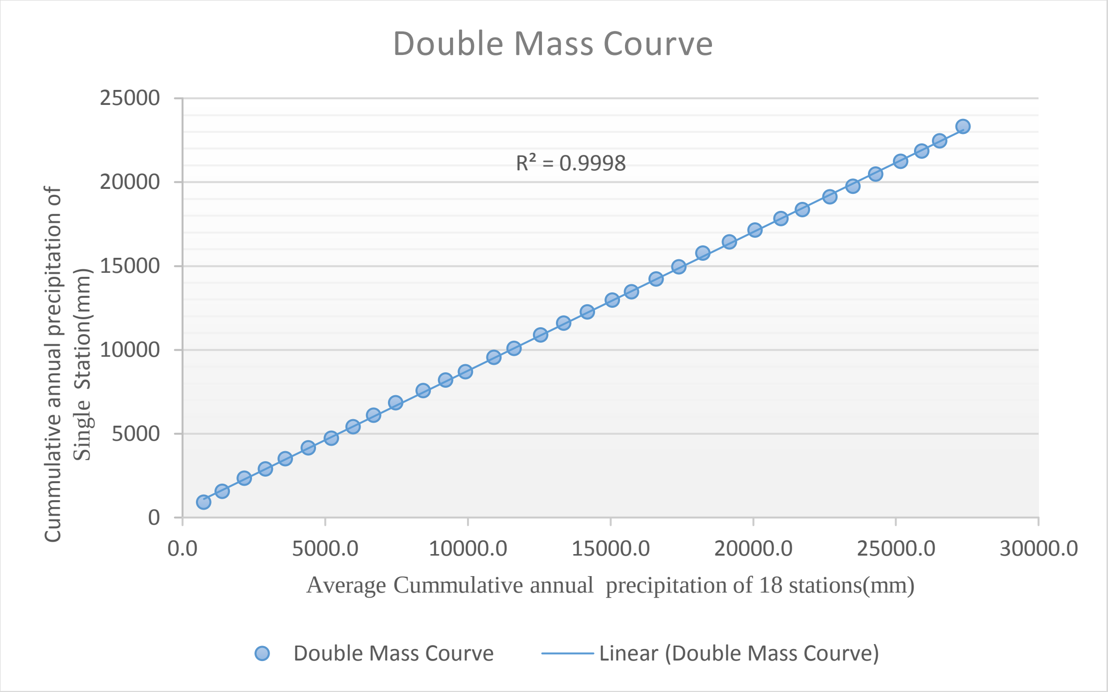


Fig 3- 7: Double mass curve of the station

3.4.3 Checking the adequacy of rain gauge network

The fundamental goal of an optimal rain gauge network with a significant major river basin is to acquire data that characterize the statistical distribution of meteorological elements with sufficient precision for practical circumstances (Das, 2002). The ideal number of rain-gauge stations to be created in a given area is provided by the following equation if there are already existing rain-gauge stations in the area:

$$N = \left(\frac{Cv}{\varepsilon}\right)^2 \text{ ....Equation 3 - 4}$$

Where N is the optimum number of rain gauge stations to be installed in the area,

Cv is the coefficient of variation of the rainfall of the existing rain gauge stations (say n) and

ε is an allowable degree of error in the estimate of mean rainfall

If there are m stations in a catchment and P1, P2………Pm is a recorded rainfall at a known time at 1, 2….m stations, and then the coefficient of variation C_V is calculated as:

$$Cv = \frac{100 * \sigma m-1}{\overline{p}} \text{ ....Equation 3 - 5}$$

$$\sigma_{m-1} = \frac{\sqrt{\sum_{i=1}^{m}(pi^2 - m * \overline{p}^2)}}{(m-1)}$$ ……………………………………………………..Equation 3 - 6

Where, Pi is monthly average precipitation at the i^{th} station and $\overline{P}$ is the average rainfall of 'm' number of stations, given by:

$$\overline{P} = \frac{\sum_{i=1}^{m} Pi}{m}$$ ……………………………………………………………….Equation 3 - 7

It is usual practice to take ε =10%. $\sigma m - 1$ Is used for calculation of C_V when the number of stations, m, in the network is less than 30 otherwise σ_m can also be used (Lohani, 2012).

In this study, the long term averaged monthly rainfall data of 18 stations found within the river basin listed in Appendix-I was used for checking the adequacy of the existing rain gauge stations. In addition to the methods of the coefficient of variation and the allowable percentage of error, the criteria recommended by World Meteorological Organization (WMO) have been used for assessing the adequacy of the existing rain gauge stations.

The WMO-recommended criteria, on the other hand, indicate the basic standards that apply to all storms and are not specifically tailored to a certain storm type (WMO, 2008). Rainfall from various storm types may have varying degrees of spatial variance. Hourly rainfall from convective storms and typhoons, for example, is highly varied in space, but hourly frontal rainfall is less so. Furthermore, rainfall over a longer period of time (monthly, seasonal, and annual) is more spatially homogeneous than rainfall over a shorter period of time (hourly and daily). Long-duration rainfall is a major problem when assessing potential water supplies, whereas short-duration rainfall is frequently employed for desalination.

3.4.4 Rain gauge network optimization framework

The optimal locations for the development of additional rain gauge stations in the basin's current rainfall monitoring networks are determined in this study, with the goals of minimizing redundant data and rainfall estimation error. The steps in the process are outlined below.

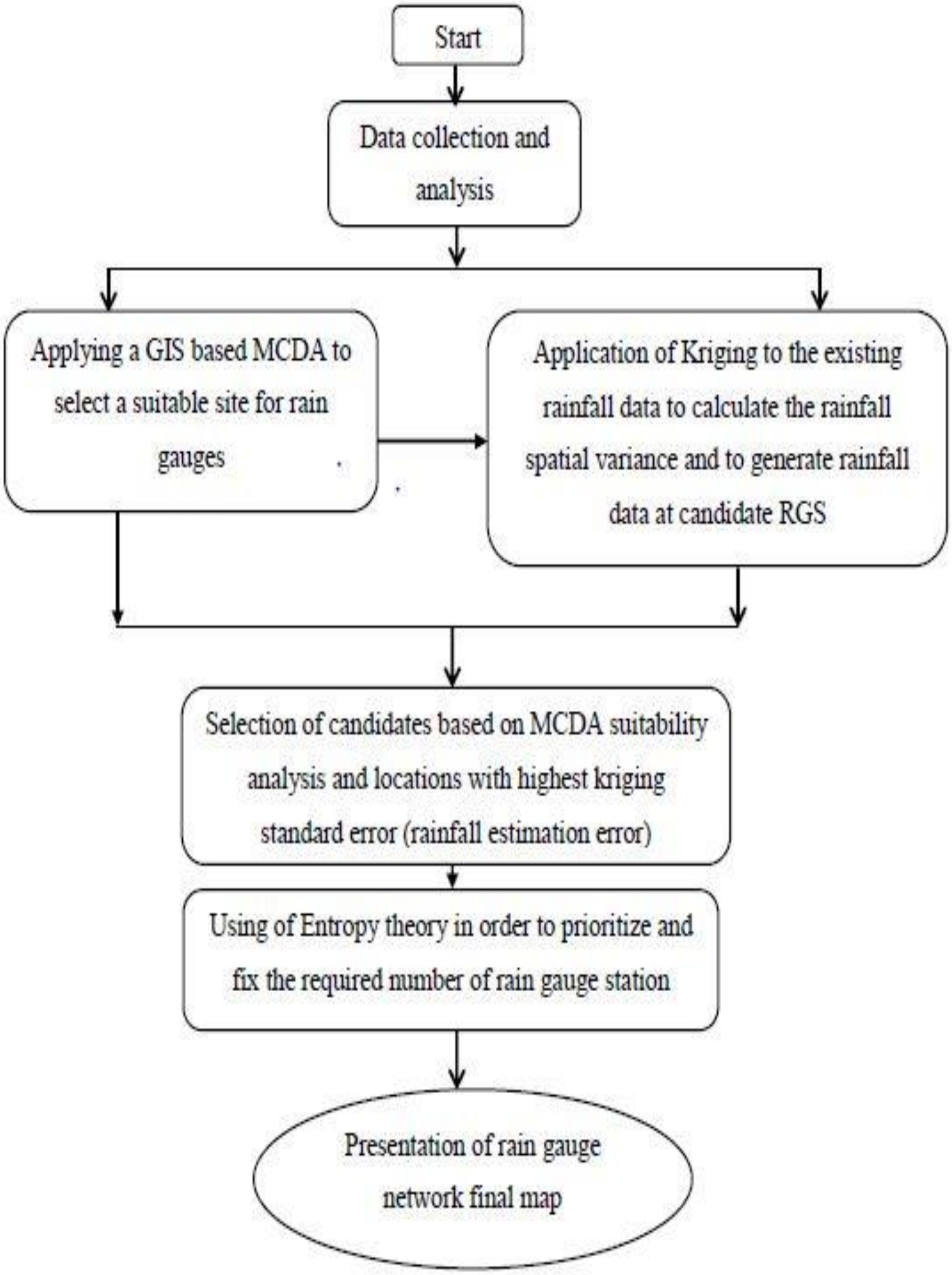


Fig 3- 8: Simplified Flow Chart of the study

3.5 Development of criteria for selecting the location of rain gauge Stations

The number and position of rain gauge stations within the study area are crucial because they impact how effectively the measurements represent the actual amount of precipitation falling in the area (WMO, 2010). The GIS-based Multi Criteria Decision Analysis (MCDA) has employed in this study to determine the best places for sitting rain gauge stations.

3.5.1 Preparation of spatial datasets and pre-processing

Land use/land cover, road networks, drainage networks, slope, and orographic impacts are the five criteria used to determine the best locations for additional rain gauge stations in the river basin.

The land use/land cover map was created by clipping from currently available Ethiopian land use/land cover raster data sets obtained from Ethiopian MoWIE, GIS, and Remote Sensing Department, and the slope expressed in percentage of rising was generated from the DEM 30*30m of raster data set of the Awash river basin using the spatial analyst tool in Arc GIS 10.5. The roads and drainage networks were also received in a spatial data format from the MoWIE GIS and Remote Sensing Department.

Prior to going to weight overlay analysis the spatial datasets for the criterions are reclassified to a common scale of suitability shown below (Table 3-3). The weights were given to the criteria using the well-known approach of Analytical Hierarchal Process (AHP).

Table 3- 2: Reclassification of each Thematic Layer into Level of Suitability (Worqlul et al., 2015)

Class Group	Scale of Suitability	Description
4	Highly suitable	Land having no limitation to sustain a given land utilization type.
3	Moderately Suitable	Land with minor limitations which when combined Together is less severe to sustain the intended activity.
2	Marginally suitable	Land having limitations which when combined together are moderate to sustained the intended activity and will so, reduce productivity
1	Not suitable	Not Suitable as the range of inputs required is beyond the set requirements.

3.5.2 Suitability analysis using GIS-based Multi-Criteria Decision Analysis (MCDA)

The criteria evaluation factors were based on the WMO (2010) recommendations as well as literature on rain gauge network optimization. The WMO's basic standards for rainfall measurements take into account a number of factors in order to reduce error sources. Rain gauge measurements are subject to a variety of errors, including sampling, exposure, evaporation, adhesion, splash, and wind ventilation effects. Inherent sample difficulties are the most common source of rainfall measurement error from gauges. Wind-induced precipitation loss is the second-largest source of rain gauge error, after sampling error (2010).

Therefore, to do the spatial GIS analysis preparation of spatial datasets and pre-processing of the above mentioned datasets will be done. GIS-based Multi-Criteria Decision Analysis (MCDA) is then applied to find the suitable areas using the above criteria after setting or assigning weights to criteria using the Analytical Hierarchical Process (AHP). The criteria that have been used in the decision-making process to select a suitable location for the development of new rain gauges in the basin were discussed in the following sections.

a) Land use/ land cover the ground surrounding the gauge site should be covered with short grass or gravel or shingle, according to WMO (2010), however a hard flat surface, such as concrete, causes excessive splashing. The location should be free of trees, buildings, walls, and other obstacles. Any such obstacle (including fencing) should be at least twice the height of the object above the gauge's rim, preferably four times the height, from the rain gauge. In addition, the site should have the same ground cover as the natural cover found in the area. The weather conditions in all forms of land use in a region must be recorded for optimal network design. Furthermore, locations for rain gauge stations must be in an open region with less dense vegetation cover, and such locations were scored better since they are less likely to be influenced by trees that do not deflect wind-blown rain into the gauge. Rain gauges were least ideal for siting in places with dense vegetation, which were given lower values (Feloni, 2018). The kinds of land use/land cover in the research region were reclassified as shown below.

Table 3- 3: Land use/ cover reclassification

Land use/cover type	Values	Suitability
Grassland, Bare land, shrub land, and settlement	4	Highly suitable
Cultivation and, plantation	3	Moderately suitable
Forest, and woodland	2	Marginally suitable
Water body	1	Not suitable

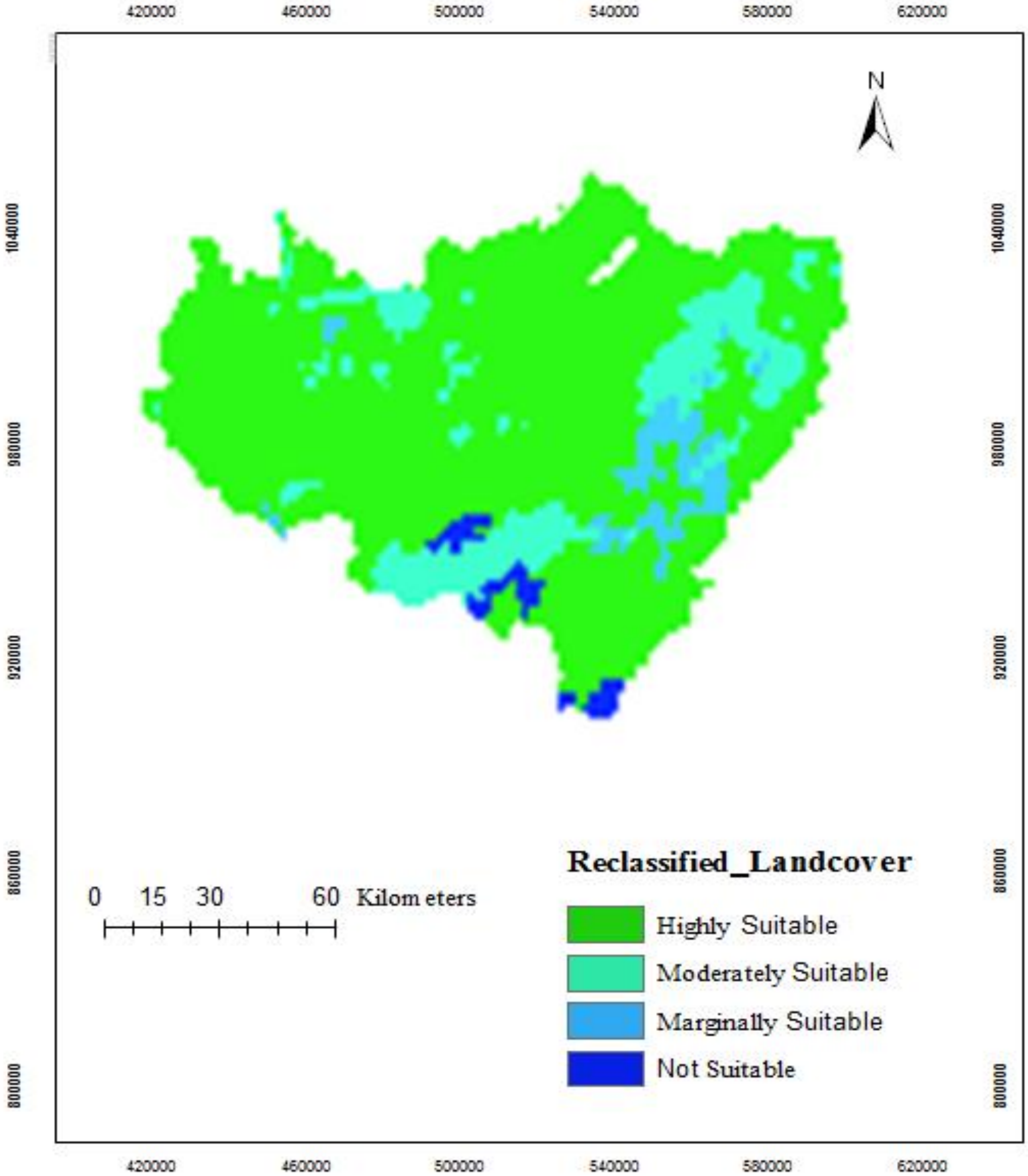


Fig 3- 9: Reclassified land cover

b) Drainage networks during the selection of a suitable location, drainage networks consisting of all possible rivers, streams, and lakes within the river basin were also taken into account. According to the World Meteorological Organization (WMO), the sensors should be located at least 100 meters away from lakes, ponds, and rivers and they should not be impacted by flooding.

The method of buffering grid cells was used to reclassify distances for drainage networks; regions adjacent to drainage networks, such as highway drainage or runoff zones, were given lower values and were deemed unsuitable for rain gauge placement (Feloni, 2018).

Table 3- 4: Drainage networks reclassification

Drainage network buffer distance (m)	Values	Suitability
29,090- 38,737	4	Highly suitable
19,393 – 29,090	3	Moderately suitable
9,696 – 19,393	2	Marginally suitable
0 – 9,696	1	Not suitable

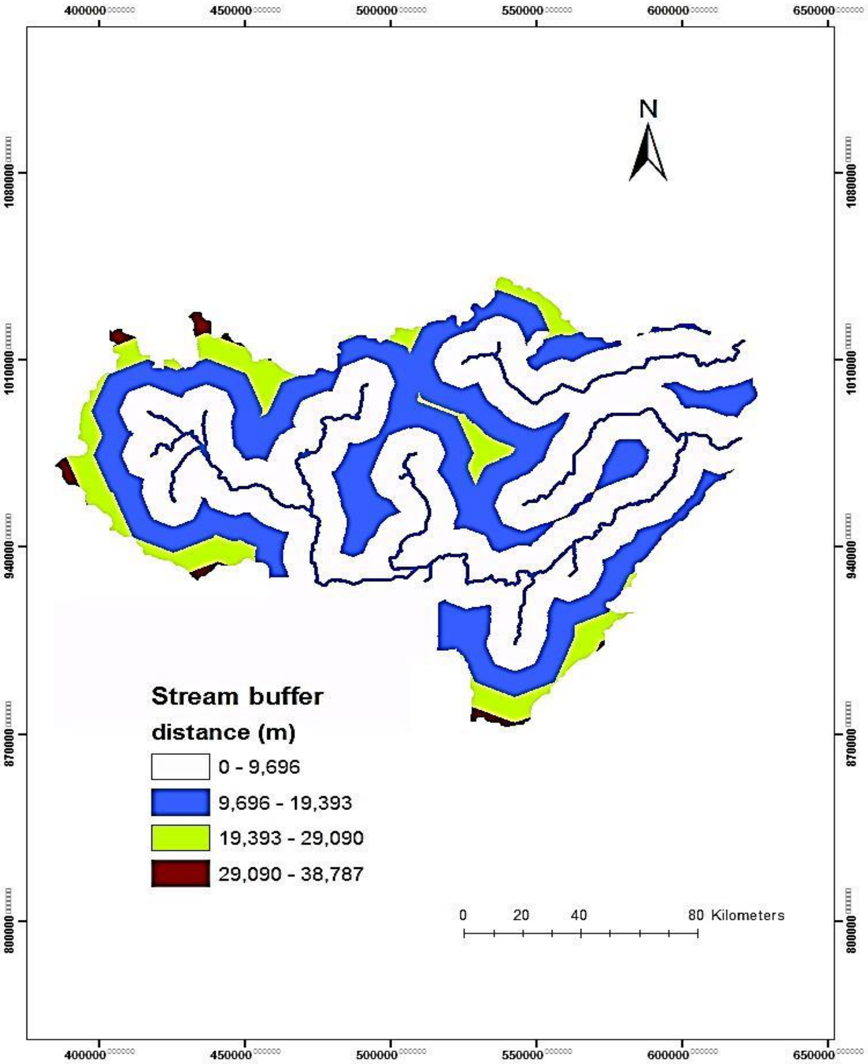


Fig 3- 10: Reclassified drainage network buffer distance of Awash River basin

c) The station's distance from roadways and urban areas ensures the most convenient access and station security. According to Feloni (2018), the optimal location for rain gauges is up to 100 meters away from roads and urban areas, although a distance of about 500 meters away from roads and 2,000 meters away from settlements is also appropriate for the highest elevation zones. Furthermore, in locations where road networks do not exist, a more flexible criterion should be established to satisfy the zone's representation. Different buffer distance criteria have been established in the literature, and in this study, different classes of distances were established based on the elevation zones of the study location.

Table 3- 5: Road network buffer distance reclassification

Buffer distance (m)	Values	Suitability
0 – 50	4	Highly suitable
50 – 100	3	Moderately suitable
100- 200	2	Marginally suitable
200 – 350	1	Not suitable

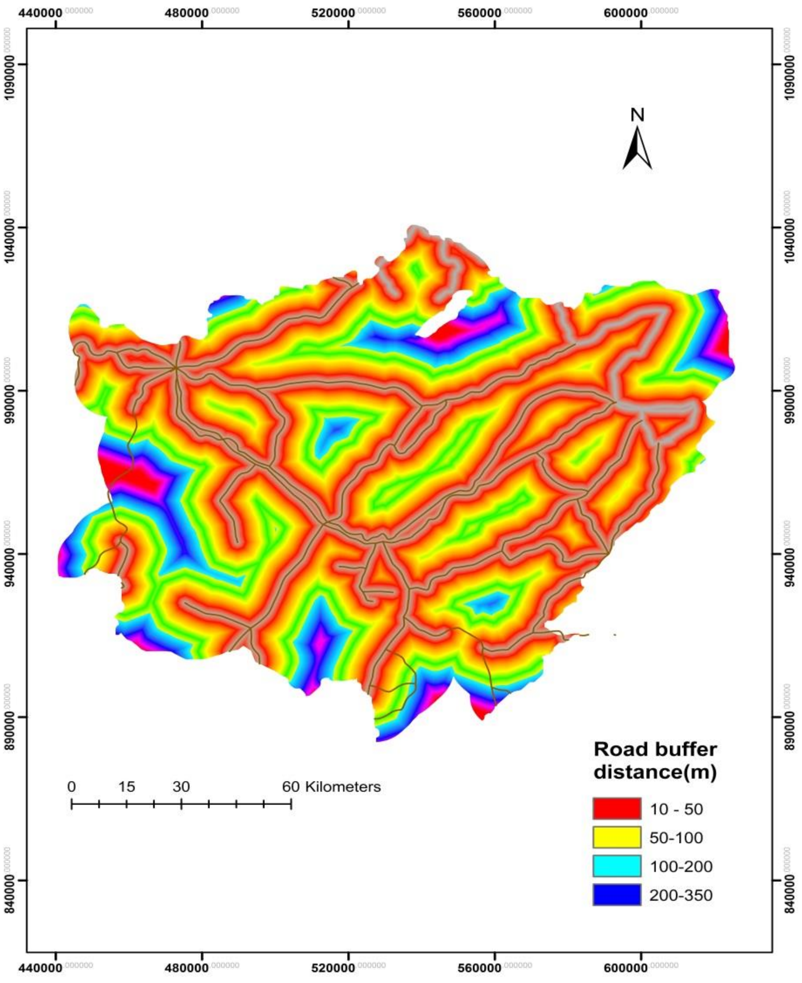


Fig 3- 11: Reclassified buffer distance of roads in the Awash River basin

d) Slope Rain gauge instruments should be situated on a level piece of land with no steeply sloping ground in the neighborhood and not in a hollow location, and should be set up around 1.50 meters above the ground. If these prerequisites aren't met, the observations may reveal oddities that are only of local interest (Hong et al., 2016).

Steep surfaces should be avoided when it comes to the terrain slope. The station should be placed on a slope of less than 5% in the best case scenario (Feloni et al., 2018).During the selection of a suitable site, avoid sites on a slope where the ground slopes away sharply in one direction, particularly if this direction is the same as the prevailing rain-bearing wind. For correct operation of the rain gauge, it must have to be installed in an accurately horizontal plane, in a flat area away from obstructions and free from waterlogging.

The slope density of the river basin under study was calculated first using the Arc GIS 10.5 spatial analysis tools and were classified according to FAO slope classification guidelines

Table 3- 6: Slope classification according to FAO (De Winnaaret al., 2007)

No.	Relief	Slope
1	Flat	< 3%
2	Undulating/ gently sloping	3 - 8%
3	Sloping	8 - 15%
4	Hilly	15 - 30%
5	Mountainous to steep mountainous	>30%

Steep gradients were assigned lower values in the reclassification of the slope layer, making them the least suited for rain gauge placement. Rain gauges should be installed in a flat area away from impediments and in a horizontal plane that is accurate for proper operation. The slopes were categorized for this investigation as shown below.

Table 3- 7: Slope reclassification

Slope (%)	Values	Suitability
0 - 1	4	Highly suitable
1 - 2	3	Moderately suitable
2 - 4	2	Marginally suitable
4-27	1	Not suitable

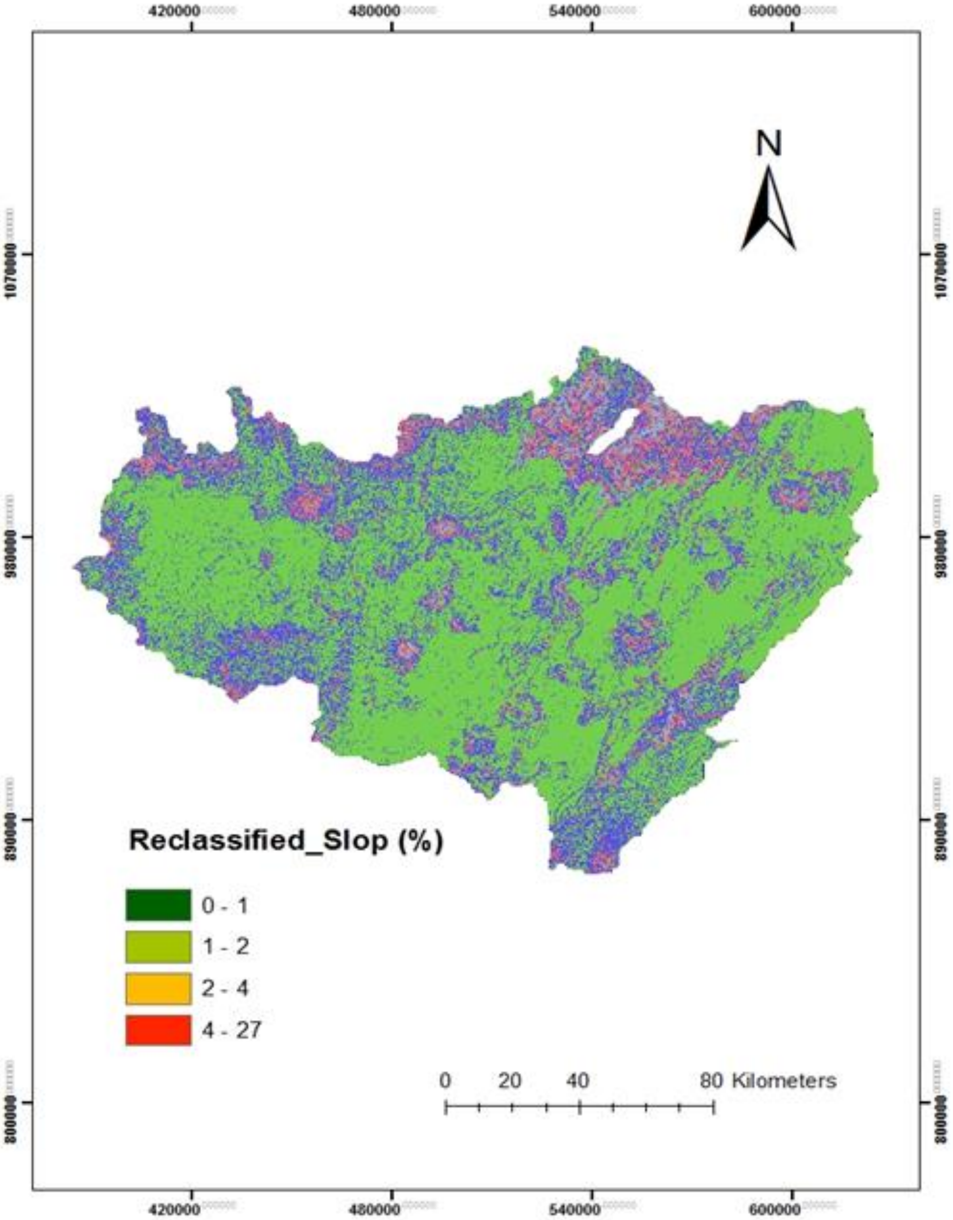


Fig 3- 12: The reclassified slope of Awash River Basin

e) Effects of orography Rain rarely fall vertically; instead, it is blown to a greater or smaller extent in the direction of the wind. The rain gauge should be placed in an open region so neighboring obstacles such as buildings, walls, and trees will not deflect wind-blown rain into the gauge, ensuring constant and precise observations (Shepherd, 2004).

In tropical countries, the orographic effects will be prominent above an elevation of 800 m (DHV Consultants and Delft Hydraulics with Hal crow, 2001). Hence, the area under study is located in the tropical region with an elevation varying from 589 to 4455 m and so that the influence can be considered to exist in the river basin for elevations greater than 800 m. To reclassify the orographic effects, the elevation ranges of the currently existing rain gauge stations were used as threshold values and most of the stations are installed at an elevation of lesser than 4000 m above sea level. The locations with an elevation of lesser than 800 m were given higher values and those with an elevation greater than 4000 m were given lower values.

Table 3- 8: Reclassification of orographic effects based on elevation

Elevation (m)	Values	Suitability
667 - 1395	4	Highly suitable
1395 - 1917	3	Moderately suitable
1917 - 2425	2	Marginally suitable
2425 - 4171	1	Not suitable

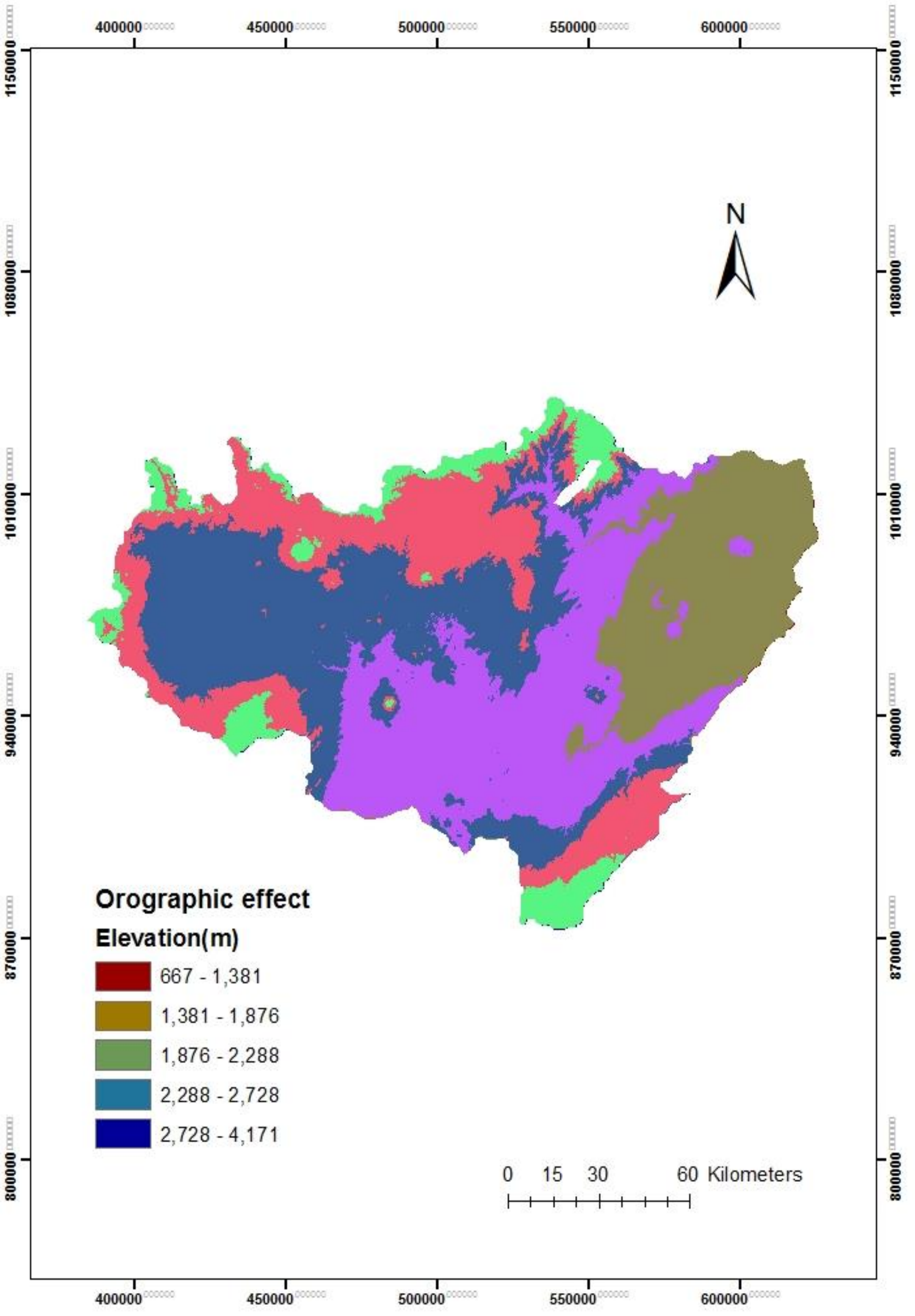


Fig 3- 13: Reclassified orographic effect of the upper Awash River Basin

3.6 Analytical Hierarchical Process (AHP) for assigning weights to criteria

Because it gives a straightforward technique to determine the relationship between criteria and options, the AHP method is exceedingly versatile. These approaches evaluate the criteria's relevance in the real world and establish how the criteria interact. By providing weights to the components for decision making, complex problems involving numerous criteria functions can be handled. When a rational and well-structured decision-making procedure is not followed, the chances of misunderstanding increase dramatically. This study adopted the widely used AHP multi-criteria decision-making method, which was first introduced by (Saaty, 1980).because of the number of criteria selected AHP is proportional method.

AHP is a widely established statistical method for calculating factor weights using a preference matrix in which all potentially key relevant criteria are contrasted against each other using replicable preference factors. AHP will consider the elements chosen as critical for the decision-making process, and all factors will be compared to one another in a pair-wise comparison matrix, which will measure the way to express relative preference among the factors. As a result, each element must be allocated numerical values conveying a judgment of its relative relevance to other factors (Fu et al., 2016).Preferences to the parameters under consideration are given using preference description scale proposed by Saaty (1980) given below (Table 3.10).

Table 3- 9: The Saaty scale of importance

Verbal judgment	AHP numeric value (scale)
1	Equal importance
2	Equal to moderate importance
3	Moderate importance
4	Moderate to strong importance
5	Strong importance
6	Strong to very strong importance
7	Very strong importance
8	Very to extremely strong importance
9	Extreme importance
Reciprocal	Values for Inverse Comparison

In addition, the AHP provides a scientific framework for detecting and correcting inconsistencies in interpreting the relative importance of components in a site appropriateness analysis. Estimating the consistency ratio, which is the comparison between the consistency index and the random consistency index, can be used to assess the consistency of the subjective judgment. The following equation can be used to calculate the consistency ratio (CR).

$$\mathrm{CR} = \frac{\mathrm{CI}}{\mathrm{RI}} \quad \text{……………………………………………………………Equation 3 - 8}$$

Where CI is the Consistency index and RI is the Random consistency index. The consistency index is a measure of consistency can be estimated using the following equation

$$\mathrm{C\ I} = \frac{\lambda\mathrm{max} - \mathrm{n}}{\mathrm{n} - 1} \quad \text{……………………………………………………… Equation 3 - 9}$$

Where λmax is the Principle Eigenvalue obtained from the priority matrix and n is Size of comparison matrix. Saaty has determined an average random consistency index (RI) on the basis of various sample sizes. The average random consistency ratios for different size of matrices are given below:

Table 3- 10: Random consistency index of different sample sizes (Saaty, 1987)

N	1	2	3	4	5	6	7	8	9	10
RI	0	0	0.58	0.90	1.12	1.24	1.32	1.41	1.45	1.49

If the value of the consistency ratio is smaller or equal to 10% the consistency is acceptable. If the consistency ratio is greater than 10%, we need to formulate again the subjective judgment.

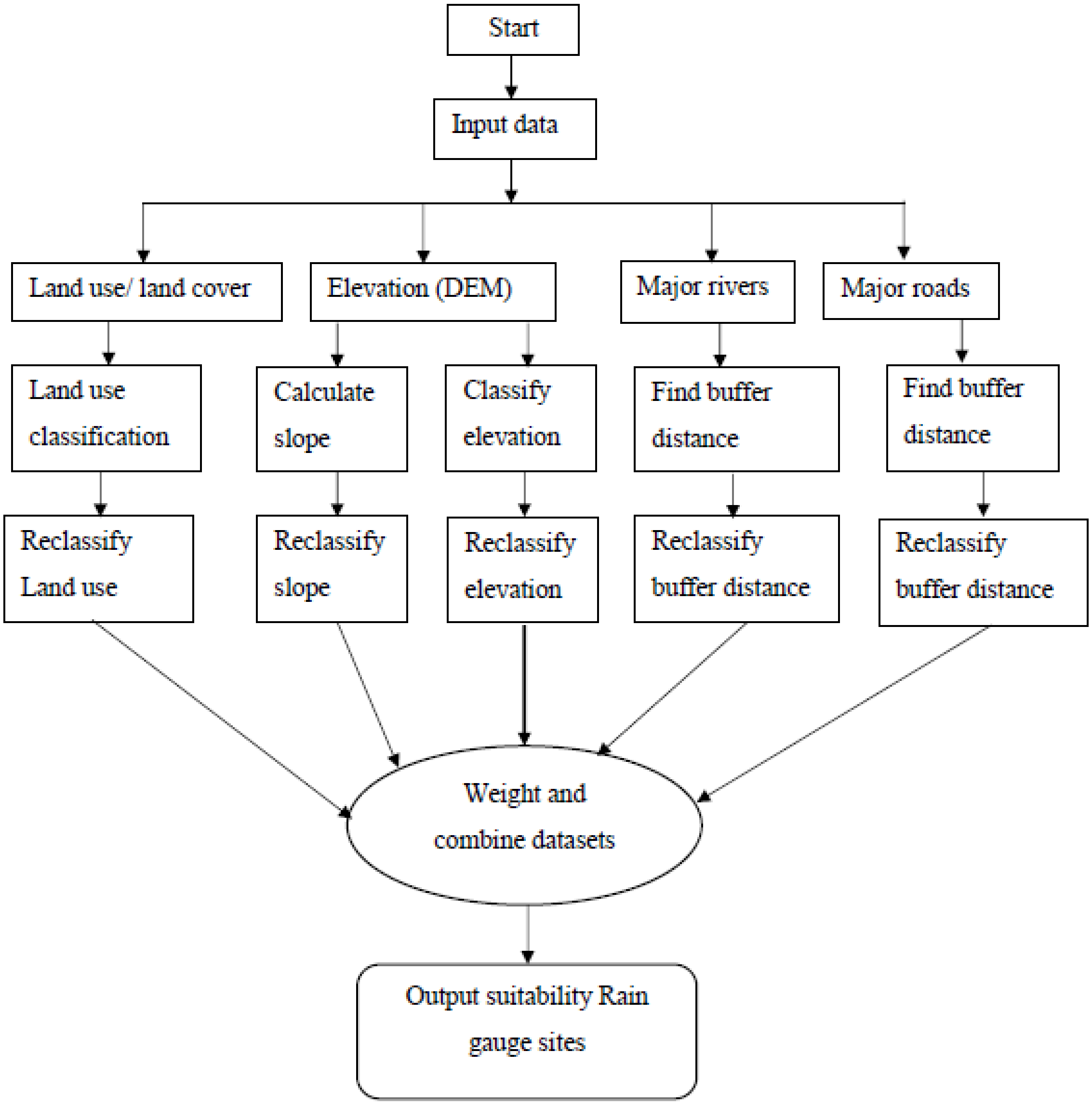


Fig 3- 14: The general flowchart of the GIS-based MCDA work

3.7 Application of the geo-statistical methods

3.7.1 The Variogram modeling and Kriging interpolation

In order to apply the Kriging-based geo-statistical approach for rain gauge network optimization Data preparation is the first step prior to precede the model analysis. In this paper the input data downloaded from satellite so no need of preparation for the exploratory data analysis (i.e. detection and removal of outliers, carrying out the normality test for the observed data and applying the data transformation for non-normal datasets)

Kriging interpolation is a sophisticated geostatistical approach for estimating a surface from a collection of distributed observation points. It refers to a family of generalized least square regression methods in Geo-statistics that is the best linear unbiased estimator of unknown variable values at ensample locations in space where no measurements are available based on known sampling values from the surrounding and is the best linear unbiased estimator of unknown variable values at ensample locations in space where no measurements are available based on known sampling values from the surrounding (Yeh et al., 2011).

Ordinary Kriging (OK) technique from the family of the classical geo-statistical methods was used in this study for interpolation of the rainfall data and estimation of the Kriging error considering the physical characteristics of the catchment under study. The Kriging estimator is expressed as:

$Z^* (x_0) = \sum wi\ (xi)$ ……………………………………………………………Equation 3 - 10

Where: $Z^*(x_0)$ - refers to the estimated value of Z at the desired location x_o;

wi - represented weights associated with the observation at the location xi with respect to x_o, and

n - Indicates the number of observations within the domain search neighborhood of xo for performing the estimation of $Z^*(x_0)$.

In Kriging, the Variogram model is used to express the degree of spatial dependence. A Variogram is a mathematical function that quantifies the spatial autocorrelation in regionalized data by measuring the distance and direction between two sites (RVs). An RV is a variable that can take on different values depending on where it resides in space. The process of constructing a link among the sampling locations in order to measure the variability associated with RV is known as Variogram modeling.

In the calculation process, the Kriging approach requires a proper Variogram model that characterizes the spatial structure of the observed data. Initially, the observed data set is used to create an experimental Variogram. After that, the experimental Variogram model is fitted with a functional Variogram model. The Variogram modeling technique allows for the fitting and selection of an appropriate Variogram model. Following the selection of an appropriate Variogram for the observed dataset, Kriging interpolation is used to anticipate interpolated surfaces across the river basin and estimate the relevant estimation errors.

The Kriging approach necessitates the fitting of a theoretical Variogram function to an experimental Variogram of the observed data. The experimental Variogram, (h), is derived from observed data as a function of separation distance, h, and is denoted by

$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [Z(xi + h) - Z(xi)]^2$ ……………………………………Equation 3 - 11

Where, $N(h)$ is the number of sampled data points separated by a distance h; xi and (xi +h) represent sampling locations separated by a distance h; Z(xi) and Z (xi +h) indicate values of the observed variable Z, measured at the corresponding locations xi and (xi +h) respectively.

The theoretical Variogram function* (h) allows for the analytical determination of Variogram values across any distance and gives the unique weight solution for Kriging interpolation. Exponential, Gaussian, spherical, circular, linear, k- Bessel, J- Bessel, rational quadratic, stable, and hole effect models are all feasible based on the shape of the Variogram function (Johnston et al., 2001; Webster and Oliver, 2007). The exponential, Gaussian, and spherical Variogram models, which are expressed by the following equations, are the most commonly employed in hydrology.

Exponential Variogram: $\gamma^*(h) = C_o + C_1 [1 - \exp\left(\frac{-3h}{a}\right)]$……………………Equation 3 - 12

Gaussian Variogram:$\gamma^*(h) = C_o + C_1[1 - exp\left(\frac{-3h^2}{a}\right)]$ ……………………...Equation 3 - 13

Spherical Variogram: $\gamma^* (h) = C_o + C_1[1.5\left(\frac{h}{a}\right) - 0.5\left(\frac{h^2}{a^2}\right)]$…………………..Equation 3 - 14

Where: C_0, a, and (C_0 + C_1) represent nugget, range, and sill, respectively, commonly called Variogram parameters. These parameters describe a Variogram model and hence affect the Kriging computation. Nugget represents measurement error and/or micro scale variation at spatial scales that are too fine to detect and is seen as a discontinuity at the origin of the Variogram model. The range is a distance beyond which there is little or no autocorrelation among variables. Sill is the constant semi-variance of the RV beyond the range (Adhikary, 2015).

Three isotropic theoretical Variogram functions (exponential, Gaussian, and spherical models) were fitted to the experimental Variogram, ignoring directional influences and assuming isotropic conditions, for Variogram modeling. Isotropy is a property in which the direction has no bearing on the spatial dependency or autocorrelation, which varies only when the distance between two points increases. On the basis of the experimental Variogram, the theoretical models' relevant Variogram parameters were inferred. To acquire the best-fitted parameters for Variogram models, manual (visual) and automatic fitting approaches (ESRI, 2016) were used. To get the best-fitted model, the Variogram parameters (nugget, sill, and range coefficients) were modified iteratively

Because the monthly patterns of rainfall data vary, data normalization was utilized separately for each month in this investigation. Then, for each month, a Kriging model was created and the related estimation error was determined. Monthly data has been overlay with a GIS tool to present an overall perspective of the Kriging errors at different points of the region in different months. Due to the considerable variability of monthly precipitation, the ratio of long-term monthly average rainfall to long-term yearly average rainfall was used as a monthly weight for overlaying monthly layers of Kriging error variance.. The points with the high values of Kriging standard error variance in the overlaid map was considered as the candidate points for rain gauge development.

An example of selection of the best Variogram model for April monthly rainfall data using samples of 16 stations were illustrated in Table below. The selected Variogram models for each month following the same procedures were also shown below. The prediction maps and prediction standard errors of each month using the selected Variogram models shown in table below

Table 3- 11: Comparison of the best Variogram models for April month

Prediction errors	Variogram models		
	Spherical	Exponential	Gaussian
Mean	-0.12	-0.01	0.31
Root-Mean-Square	11.78	11.58	13.7
Mean Standardized	0.009	0.002	0.16
Root-Mean-Square Standardized	1.18	1.08	2.86
Average Standard Error	8.83	9.64	7.92

Table 3- 12: The selected best Variogram models in each month

Month	Variogram model selected
March	Exponential
April	Spherical
May	Gaussian
June	Exponential
July	Exponential
August	Exponential
September	Gaussian
October	Spherical
November	Gaussian

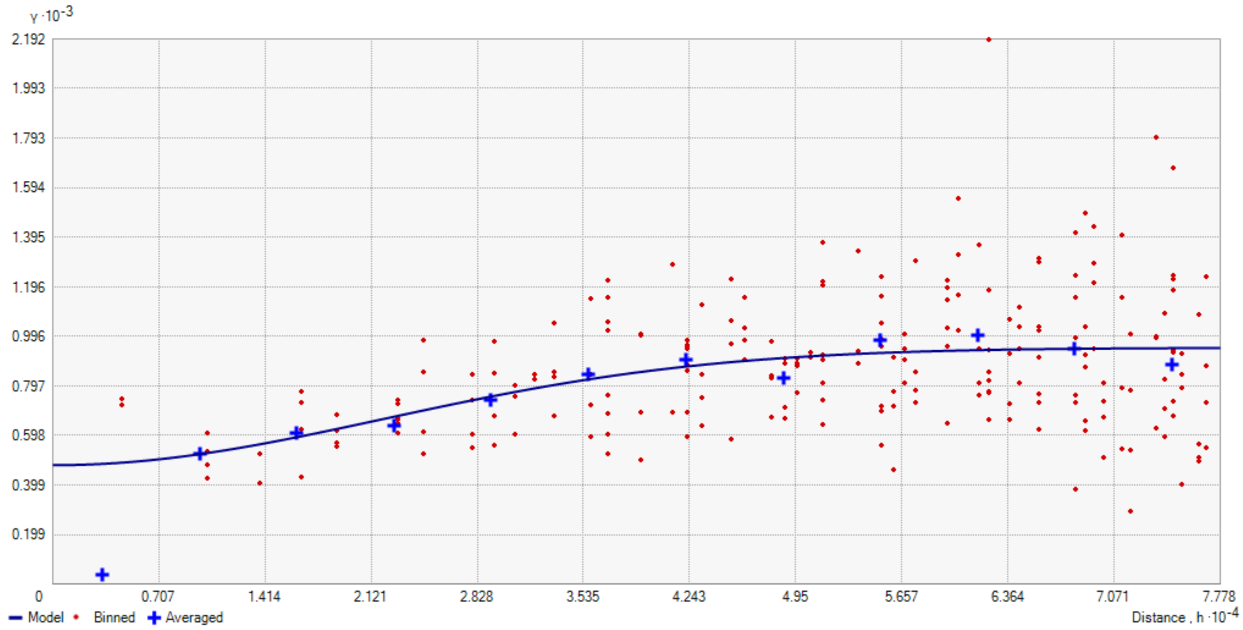


Fig 3- 15: Variogram model of rainfall data

3.8 Application of the entropy and remote sensing approaches

3.8.1 Entropy approach

In thermodynamics, the concept of entropy has been understood as a measure of disorder or randomness of a system. While Shannon (1948) extended the entropy concept to information theory by recognizing that uncertainty in a system will be decreased when information is added to the system. Therefore, the term entropy in information theory, introduced by Shannon in 1948 describes the amount of information contained in a random variable. Different measures of entropy including the marginal entropy, joint entropy, conditional entropy, and Transformation are used in the entropy theory (Masoumi & Kerachian, 2009). Some details of the different forms of entropy were discussed below.

a. Marginal entropy:

When information is delivered to a system, one can anticipate the system's uncertainty to be reduced; thus, marginal entropy refers to the amount of information given to a system by knowing a variable. If a random variable X is expected to have N outcomes with a probability distribution P = {P_1, P_2, _ _ _, P_N}, the (weighted) average information provided by the N joint events is given by:

H (x) = – P_1 log P_1 – P_2 log P_2 - ……….– P_N log P_N = – Σ ($PilogPiNi$=1)…… Equation 3 - 15

Where: H(x) is the marginal entropy of a random variable X with $\sum_{i=1}^{N} Pi = 1, Pi \geq 0$

The continuous probability approach can be used to replace any discrete probability when the rainfall data seems to be a continuous random variable. Equation 3-15 can thus be rethought as a continuous probability condition by altering it with the probability density function. If an event has the potential for multiple outcomes, the differential entropy can be represented as follows:

$H(x) = \int_{-\infty}^{\infty} f(x) log f(x) dx$ …………………………………………………..Equation 3 - 16

Where: f(x) is the probability density function. If f(x) is a normal distribution,

$H(x) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma}} \exp[\frac{1}{2}(\frac{x-\mu}{\sigma})^2\ ] ln \frac{1}{\sqrt{2\pi\sigma}} \exp[-\frac{1}{2}\Big(\frac{x-\mu}{\sigma})^2\Big] dx = \frac{1}{2}\ln(2\pi e\sigma^2)$ ……Equation 3 - 17

Where:μ is mean value, and σ is the standard deviation

b. Joint entropy:

Because rainfall data gathered from rain gauge stations may overlap, the joint entropy can be used to estimate the uncertainty of the entire set of data when the two sets of data are the two variables x and y. This is provided by

$H(x, y) = -\int i \int j\ P_{ij} \ln\big(P_{ij}\big)\, dx\, dy$ ………………………………..…………Equation 3 - 18

Where: H (x, y) and P_{ij} are the joint entropy and joint probability of x and y, respectively. Likewise, for three variables (or three stations), the joint entropy of x, y, and z is

$H(x, y, z) = -\int i \int j \int k\, p_{ijk} \ln\big(p_{ijk}\big)\, dxdydz$ …………………………………Equation 3 - 19

Where: H (x, y, z) and p_{ijk} are the joint entropy and the joint probability of x, y, and z, respectively.

Therefore, when the assumed random variable x_1, x_2... x_n of an area is a multi-variable of a normal distribution with mean μ and covariance matrix k, the calculation of the joint entropy can be expressed as:

$H(x_1, x_2 \dots x_n) = \frac{1}{2}\ln[(2\pi e)^n|k|]$ …………………………………………..Equation 3 - 20

$= \frac{1}{2}|k| + \frac{n}{2}(\ln[2\pi + 1]$

Where: |K| is the determinant (Mahmoudi et al., 2016)

c. Trans-information entropy:

T (x, y) is the mutual information or overlapped information of two rain gauge stations x and y, where f (x, y) is the joint entropy of x and y, and f(x) and f(y) are the probability density functions of the variables x and respectively.

T (A, B) = H (A) – H (A\B) = H (B) - H (B\A) = T (B, A)……………………….Equation 3 - 21

Where: T (A, B) is Trans-information between the variables A and B. The larger the Trans-information is, the higher those variables depend on each other. In other words, the Trans-information indicates how much information content is transferrable from other variables. Trans-information shall be written as (Lathi, B.P. An Introduction to Random Signals and Communication Theory; International Textbook Company: Scranton, PA, USA, 1968).

Trans-information is typically used for measuring mutual information between two variables or two groups of variables as the generalized form for multivariate Trans information is given as:

$T\,[(X_1, X_2\ldots, X_K); (X_{K+1}, X_{K+2}\ldots, X_N)]$…………………………………………...Equation 3 - 22

$= H\,(X_1, X_2\ldots, X_K) – H\,[(X_1, X_2\ldots, X_K)/ (X_{K+1}, X_{K+2}\ldots, X_N)]$…………………...Equation 3 - 23

Trans-information, T (x, y) in the continuous form can be expressed as follows:

$T(x, y) = T(y, x)$

$= -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y)\ln(\frac{f(x,y)}{f(x)f(y)})$ ………………………………………………Equation 3 - 24

Where: f(x, y) is the joint entropy of stations x and y, f(x) and f(y) are the probability density function of the rainfall in stations x and y, respectively. (Mogheir, 2003) states that for variables with normal distribution Equation 3.-24 can be simplified as follows:

$T(x, y) = -\frac{1}{2}\ln(1 - r_{xy}^2)$ ……………………………………………………..Equation 3 - 25

Where: r_{xy} is the correlation coefficient between rainfall at stations x and y.

For the optimal design of rain gauge networks, the entropy approach is used in addition to Kriging in this study. The entropy method can be used with data that is roughly regularly distributed. As a result, the monthly rainfall data generated by each candidate rain gauge station from the anticipated monthly rainfall surface using the normalized monthly rainfall data developed for Kriging analysis are used here. The trans-information in the basin is estimated in the next step. New information is acquired by adding new candidate rain gauges, reducing the uncertainty in regional rainfall estimation.

3.8.2 Prioritization of candidate rain gauges

The entropy value of each rain gauge station can be used to determine its importance in supplying the required information in the network. The higher the entropy value, the more unpredictable the situation; as a result, each station is prioritized in descending order of entropy levels. After validating the first station with the highest marginal entropy value, the next stations are chosen and added one at a time, based on the system's inferiority and overlapped information. The standardization of finding the second most significant station in the sequence is established as follows to minimize the system's uncertainty:

Min $\{H(x_2) - H(x_2|x_1)\}$ …………………………………………………..Equation 3 - 26

Then, the criteria for the nth station to be added to the current network are as follows.

Min $\{H(x_1, x_2 \ldots\ldots x_{n-1}) - H[(x_1, x_2 \ldots\ldots x_{n-1}|xn]\}$ …………………………Equation 3 - 27

After selecting the optimal candidate points for the development of new rain gauges, it is necessary to prioritize them in order to make a decision about the number of stations that can be developed considering the limited budget for rain gauge network development.

The entropy-determined priority ranking of stations is utilized to eliminate stations. The joint entropies will steadily increase when additional stations are added based on their priority, until the amount of sent information increases insignificantly, and adding stations provides no significant value to the network.

Scientific studies describe that the optimal redundant information to pass by adding a station is recommended to be 60%. And the stations giving maximum redundant information beyond that are proposed for possible discontinuities from adding to the network (Vivekananda& Jagtap, 2013).

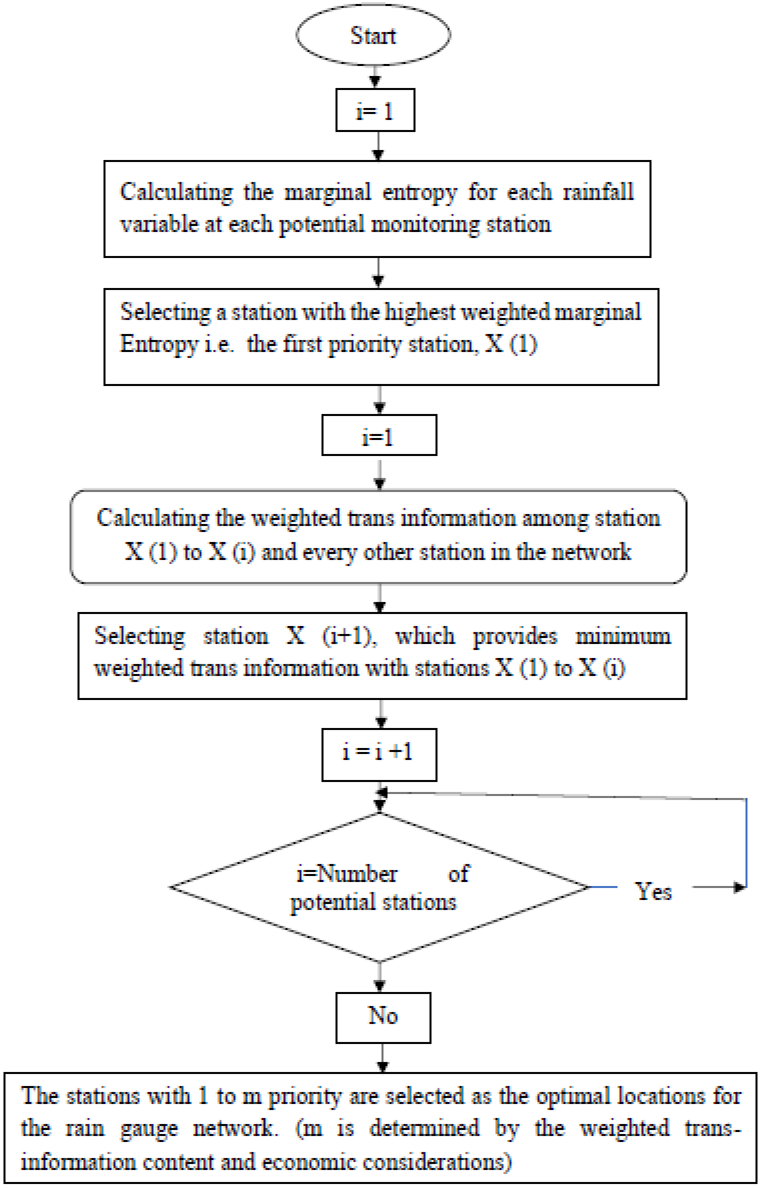


Fig 3- 16: Main steps for prioritizing the candidate rain gauge stations in the river basin (Mahmoudi et al., 2016)

3.9 Remote sensing approach

Earth observation satellites collect data on a variety of topics related to Earth resources, such as Earth surface features and atmospheric conditions. In digital form, satellite data is collected, transferred, stored, and evaluated. The spatial, spectral, radiometric, and temporal resolutions of satellite data are all different. The resolution of satellite data is increasing over time, which is a general indicator of progress. Satellites can be used to generate new data on their own or in conjunction with other geographic data sources. One of the most typical applications is to display satellite data in a map format for human viewing.

Remote sensing has many advantages over ground-based surveys, including the ability to survey enormous areas of land at any time. With the advent of satellite technology and multi/hyper spectral sensors, this capability has been further enhanced, allowing for the capture of images of very large areas of land in a single pass, as well as the collection of data about an environment that would otherwise be invisible to the naked eye. Collect comparable data over large area and repeated measurements over same area- dynamics over space and time can be analyzed remote sensing is increasingly becoming an important source of information in wide range of Hydrological applications. Mapping and monitoring of water bodies, Soil moisture estimation, Drainage basin mapping and watershed modeling

3.9.1Satellite Remote Sensing Data

Data from satellites is acquired using sensors installed on the satellites and kept primarily in digital form. Satellite images or pictures are created when satellite data is shown in a map format. Satellite imagery is another name for satellite data. Human eyes can only detect a tiny fraction of the electromagnetic spectrum (spectral band), while satellite sensors can take up data from a much larger range, allowing humans to perceive ultraviolet, infrared, thermal, and microwave signals. Color-coded maps are used to depict satellite data in order to give significant patterns and correlations. Aerial photography is occasionally used to enhance satellite imagery. Both are known as remote sensing or remotely sensed data; however the former can cover a greater viewing area more simply, whereas the latter can typically show more ground details. More crucially, satellite data may be obtained without regard for political boundaries (Tang, L.N., Shao, G.F., Piao, Z.J., Dai, L.M., Jenkins, M.A., Wang, S.X., Wu, G., Wu, J.G., and Zhao, J.Z.) (2010).Finally I have used the satellite derived rain is to know the spatial variability of the study area.

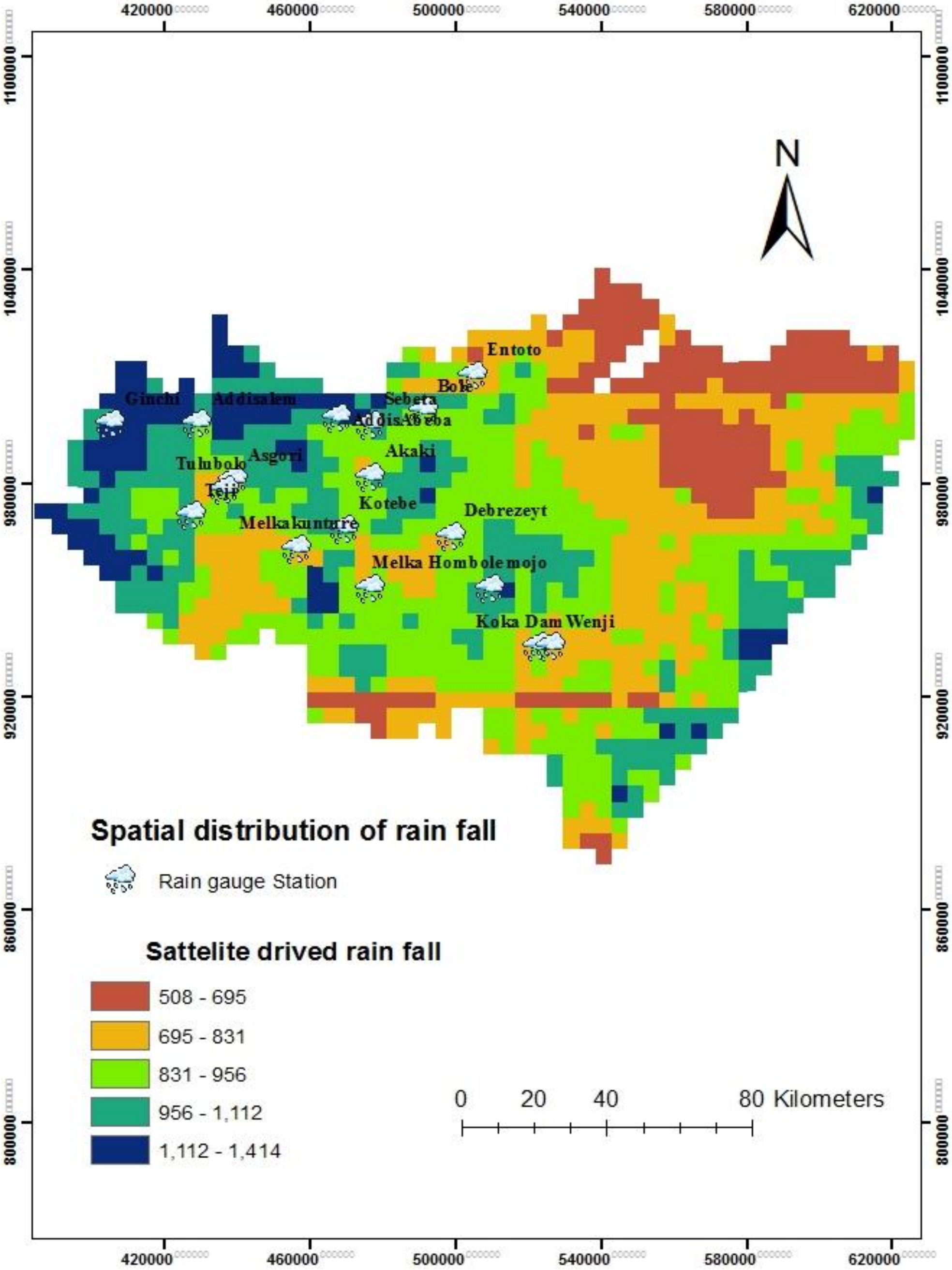


Fig 3- 17: Persian cloud classification system (P-CCS) based satellite precipitation product

4. Results and Discussion

4.1 Adequacy of rain gauge stations

In the Awash basin there are about 76 gauging stations, out of these 21 stations are found in the upper awash sub- basin particularly in the upstream of Koka Dam (Mengistu 2008). These all stations are not selected for modeling because some have defect, they are under the influence of natural and manmade factors. Some of the stations are set at upstream of Koka Reservoir that are directly subjected to the backup effect of the reservoir; others are set downstream of Legedadi Dam which has no natural record due to the influence of the release of the water through spillway and the rest are on the very plate and plain area that may be subjected to the over flow others are highly influenced by land use change such as urbanization which lead to wrong out comes.

The adequacy of rain gauge stations in the river basin was checked using the statistical methods of the coefficient of variation and allowable percentage of error. The optimal number of stations required in the river basin was determined as follows. The statistics of the average monthly rainfall data of 16 stations in the river basin were shown in Table 4-1 below:

Table 4- 1: Computation of the optimum number of stations

Parameters	Computed values
Mean monthly rainfall (mm)	89.21
Standard deviation	79.93
Coefficient of variation	89
Recommended accuracy (uncertainty levels) expressed at the 93% confidence interval for precipitation amount and form (3-7%)	Taken 7 %
Optimal number of stations with 7% allowable error	148

Having 16 rain gauge stations in the river basin an additional 136 rain gauge stations were demanded to be added to the network. However, this statistical method was used for preliminarily assessing the adequacy of the currently existing rain gauge stations in network. The results showed that the currently existing rain gauge stations in the network were not adequate enough, so that an additional rain gauge stations would be required in the network to obtain the required accuracy of rainfall data.

4.2 Rain gauge suitability mapping using Multi criteria Decision Analysis

The multi criteria decision analysis was used for selection of a suitable location for rain gauges. In this study, the criteria including the land use/ cover, slope, road networks, drainage networks, and orographic effects were considered. The weights to each of the criteria were assigned using AHP.

Table 4- 2: Weight of each criterion

Criteria	Land use	Slope	Drainage network	Road network	Orographic effect
Weight	0.21	0.56	0.06	0.05	0.1

The consistency of the criterions was calculated as 0.09 < 0.1. This indicates that the consistency is acceptable. A suitability map with the stations currently available at the river basin, totally 16 stations The suitability map's validation revealed that 57% of existing rain gauge stations were in very appropriate regions, 33% were in moderately suitable areas, and just 10% were in marginally suitable areas (Figure 4.1). According to the distribution, suitability mapping works well for identifying ideal places for rain gauge stations, and it may be utilized to expand the existing network in the river basin in the future.

The rain gauge stations that were to be installed in the river basin were initially placed in highly and moderately appropriate regions. The locations of potential rain gauge stations will be chosen based on their suitability were shown below in Figure 4.1.

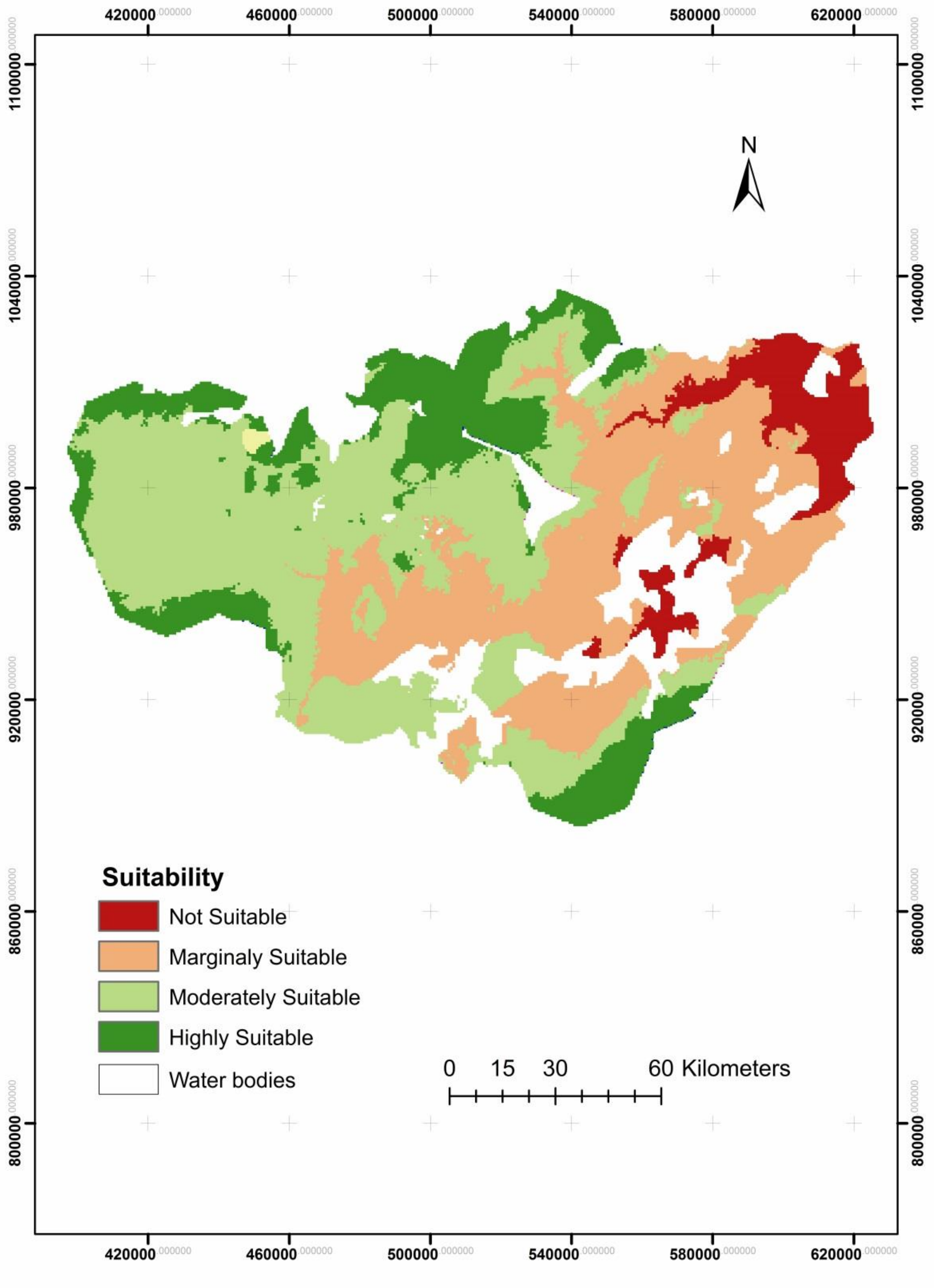


Fig 4- 1: The developed suitability map of Upper Awash River basin

The rain gauge stations that were to be installed in the river basin were initially placed in highly and moderately appropriate regions. The locations of prospective rain gauge stations inside each grid cell were chosen based on their suitability. Urban areas are also taken into account during the selection process, in addition to the suitability map. The candidate stations that were preliminarily chosen are listed below (Figure 4.2).

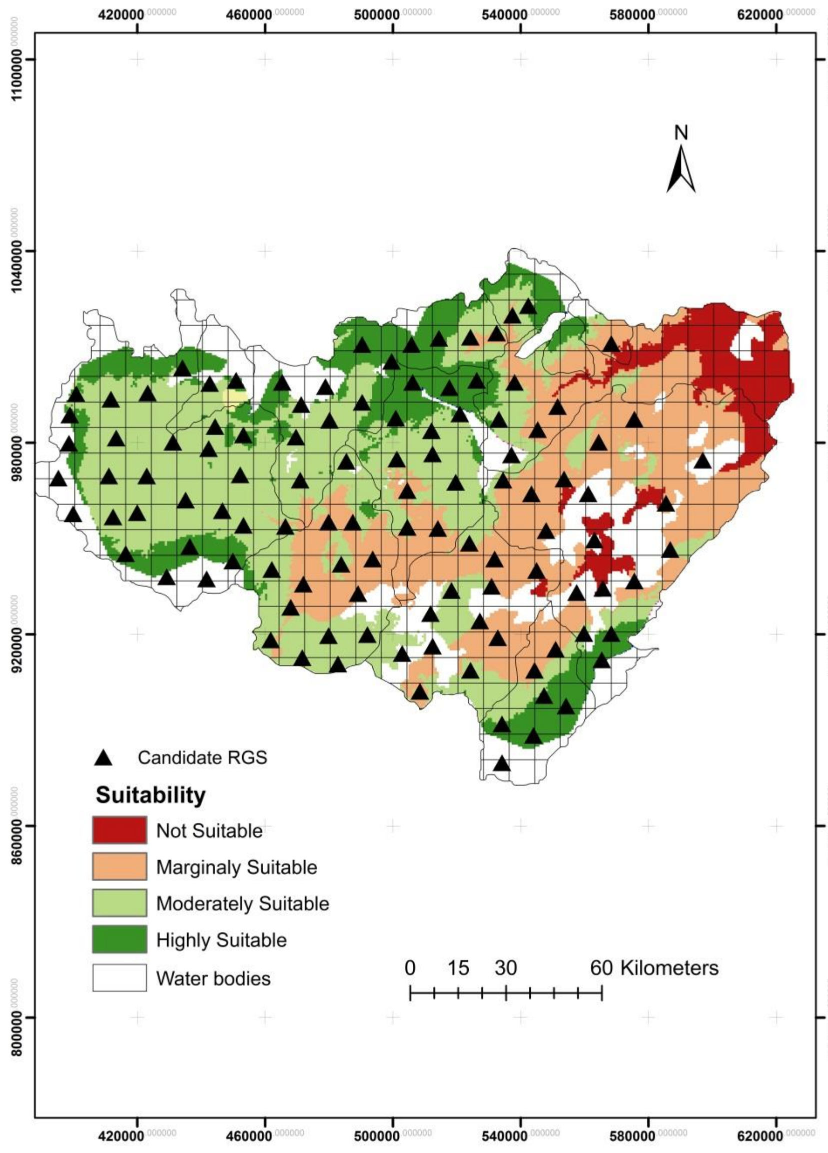


Fig 4- 2: The developed suitability map and the candidate rain gauge stations

4.3 Prediction of monthly rainfall data using Kriging method

After choosing the best Variogram models for each of the nine months, the average monthly rainfall in the basin surface was predicted. Rainfall data was generated at the basin's surface. Rainfall data has been created at each site in the river basin using the prediction map, and the prediction standard error map shows that locations with significant error require an extra rain gauge. The potential locations were chosen based on large standard error regions. Below are the prediction and Kriging standard error maps created with the specified Variogram models (Figure 4.2).

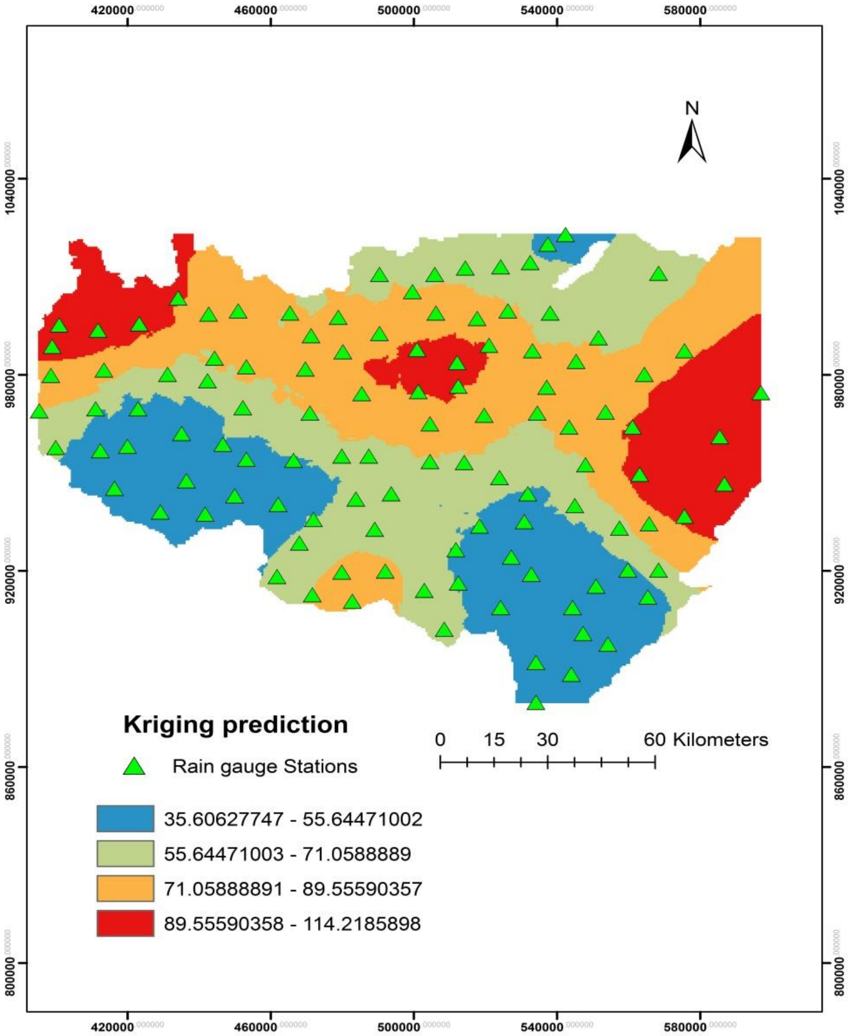


Fig 4- 3: Rainfall prediction map of the Awash River basin

The Spherical model was chosen based on the computed prediction errors, which included Mean error, Root Mean Square Error (RMSE), Mean Standardized Error (MSE), Root-Mean-Square Standardized Error (RMSSE), and Average Standard Error (ASE). For Kriging interpolation, the Variogram model with a Mean error around zero, a lowest RMSS error, an ASE equivalent to the RMSS error, and an MSE near zero was chosen.

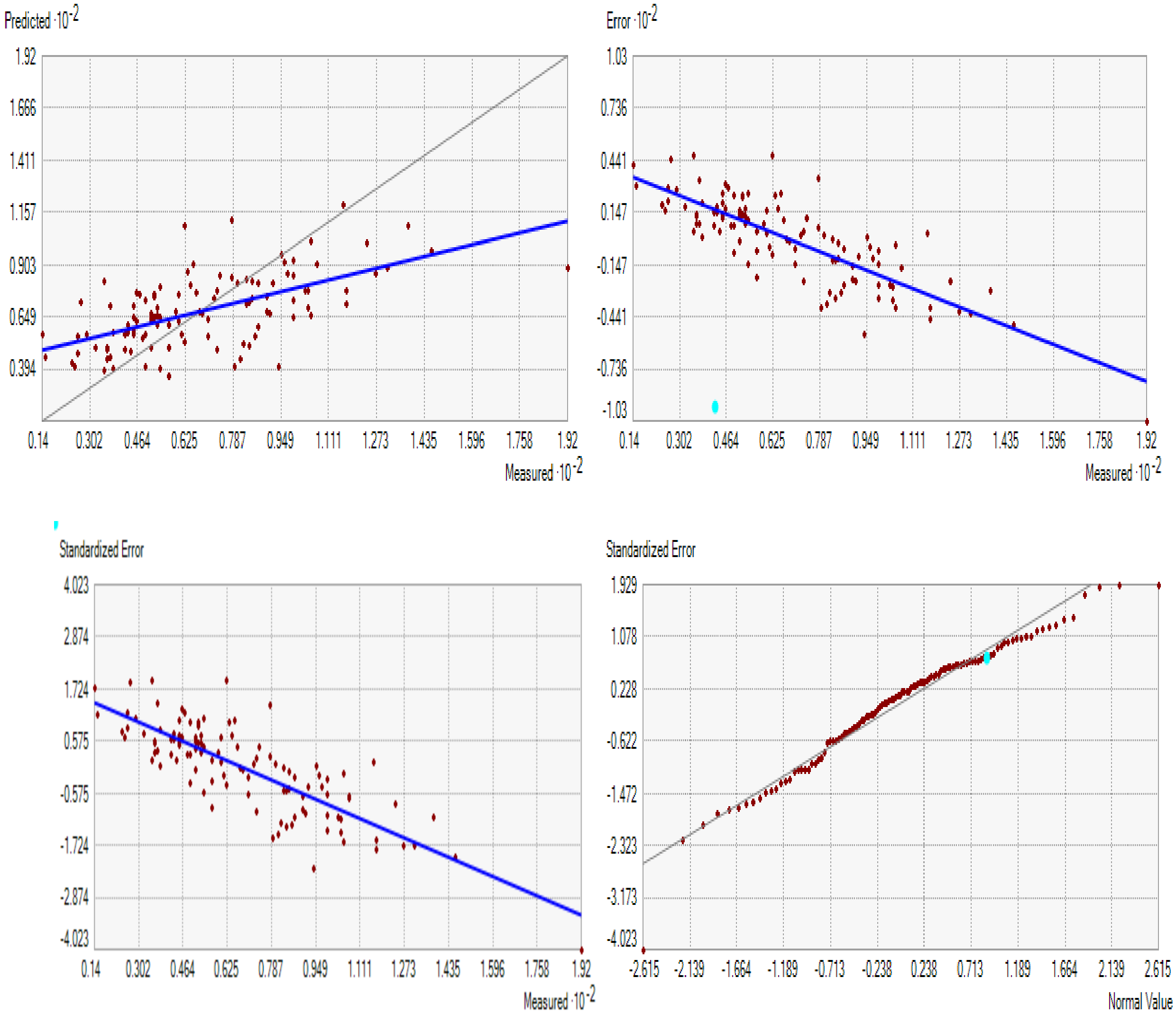


Fig 4- 4: Cross validation of Kriging

4.4 Determination of optimal number of stations using information entropy

The entropy technique was utilized to determine whether the selected candidate rain gauge stations should be included or removed. The entropy approach was used to prioritize the candidate rain gauge stations and to finally fix the optimal number of stations needed to be added to the network in this study. Monthly rainfall data of all the candidate rain gauge stations were generated for each month, and the entropy approach was used to prioritize the candidate rain gauge stations and to finally fix the optimal number of stations needed to be added to the network.

The computation of marginal and Tran's information entropy of each rain gauge stations was conducted. For the sake of space utilization, I have done only the three types of entropy Trans-information, marginal and joint entropy given satellite drive values and the reduction in uncertainty was described in Table 4.3 below

The results in Table 4.3 below show that the candidate station denoted as Wenji were the first priority station to be added in the basin. The second priority station were also Tulu bolo, a station with the smallest Trans information entropy and the last priority station were Kotebe as shown in transformation index matrix Table 4.3 below.

Finally, a literature-based judgment was taken to determine the total number of candidate stations to be added to the network. According to scientific studies, the recommended optimal redundant information reduction in a rain gauge network was a maximum of 60% because increasing this percentage beyond that would increase the number of stations required, thereby increasing the cost of installation and operation of rain gauges. In this study, an extra 148 rain gauge was recommended to be built, based on maximum redundant information of 60%. Finally, for the application of water resources projects in the Upper Awash river basin, an additional 49 rain gauges and 16 presently existing rain gauges totaling 65 stations were required.

Table 4- 3: Details of redundant information passed by each existing stations

The Priority Of each Station	Station Name	Transformation Index (T)	Optimal redundant Information Passed (T/2.91) *100 %
X(1)	Wenji	………	………
X(2)	Tulu bolo	0.83	28.70
X(3)	Awash Melka	0.56	19.22
X(4)	Ginchi	0.72	24.88
X(5)	Koka dam	1.06	36.41
X(6)	Teji	1.20	41.31
X(7)	Melka kunture	1.51	51.82
X(8)	Addis Ababa	1.61	55.42
X(9)	Sebeta	1.81	**62.35**
X(10)	Asgori	1.89	65.14
X(11)	Bole	2.01	69.12
X(12)	Addisalem	1.96	67.37
X(13)	Akaki	1.95	67.05
X(14)	Entoto	2.66	91.41
X(15)	Debre zeyt	2.89	99.36
X(16)	Kotebe	2.91	100.00

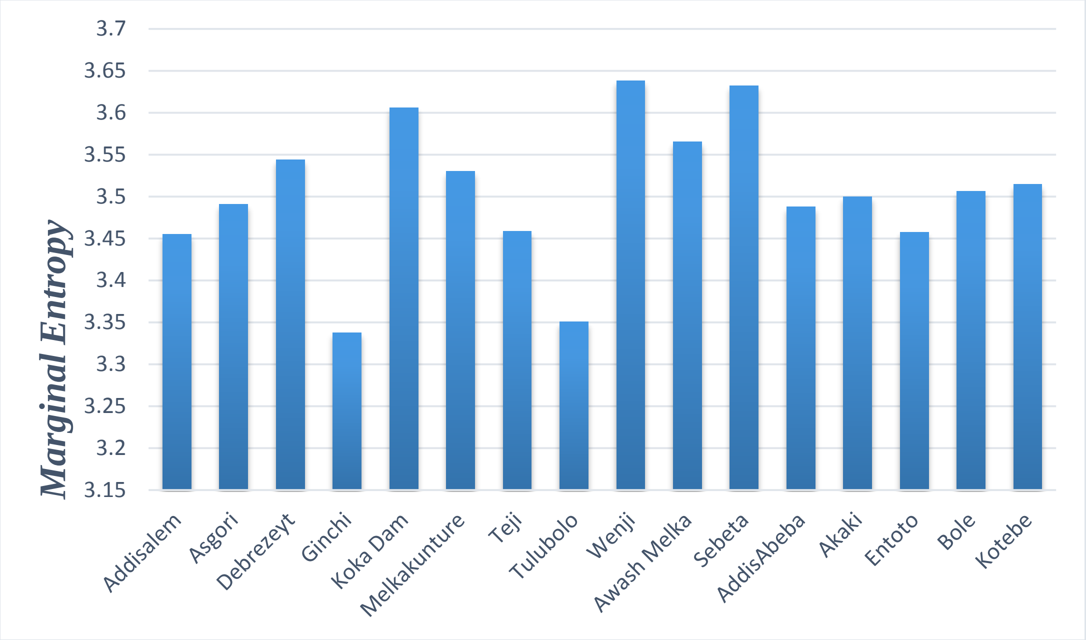


Fig 4- 5: The marginal entropy for each candidate stations at the upper Awash

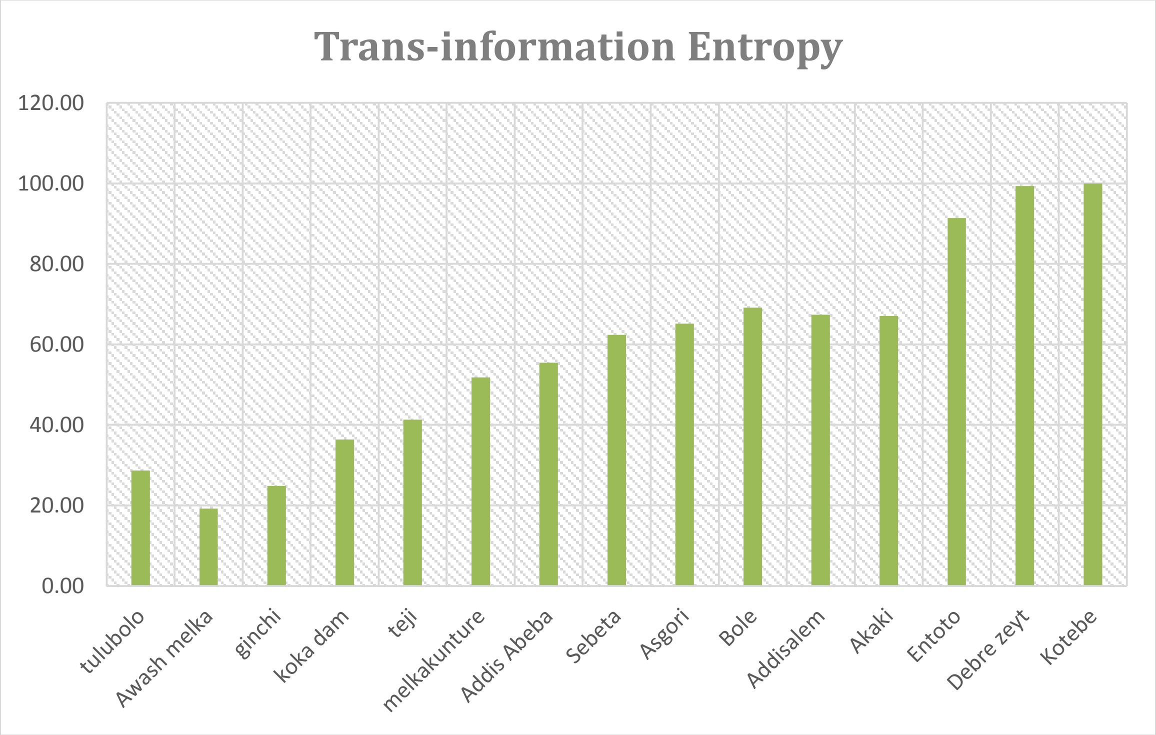


Fig 4- 6: The Trans-information entropy for each candidate stations at the upper Awash

Table 4- 4: New candidate rain gauge station and their rainfall amount

Station name	X(m)	Y(m)	Rain fall(mm)	Stn.no	X(m)	Y(m)	Rain fall(mm)
new1	512080	983751	95	new69	596968	974538	124
new2	512409	976512	96	new70	471939	935713	73
new3	501222	974867	127	new71	479835	919591	73
new4	514054	953151	49	new72	461739	918275	46
new5	519647	967629	80	new73	492009	919920	104
new6	504512	964996	87	new74	462068	940319	36
new7	511751	926500	43	new75	568343	1011060	51
new8	518331	933739	62	new76	551563	991318	51
new9	523925	948545	86	new77	538073	998886	53
new10	531821	943610	42	new78	533137	987370	54
new11	527215	924197	49	new79	545311	984080	91
new12	550905	915314	25	new80	537086	976183	85
new13	554195	897546	79	new81	471281	991976	66
new14	547286	900836	36	new82	442656	998557	92
new15	534125	879779	45	new83	450881	999544	90
new16	534125	891953	81	new84	411727	993621	146
new17	524254	908733	37	new85	446604	958745	94
new18	532808	918933	38	new86	483784	941965	45
new19	512409	916301	38	new87	493654	943610	82
new20	502867	913997	68	new88	471610	912681	72
new21	508461	902153	60	new89	482797	910707	117
new22	489048	932752	45	new90	467991	928475	61
new23	487403	955126	27	new91	466345	953809	45
new24	479835	955126	57	new92	441669	937358	35
new25	485429	974209	97	new93	499577	1005466	47
new26	470952	968287	99	new94	505828	1010731	59
new27	453184	954138	26	new95	532479	1014350	43
new28	435088	962035	32	new96	542350	1022905	48
new29	442327	978157	37	new97	537415	1019943	57
new30	452197	969932	52	new98	547944	952493	46
new31	453184	982435	52	new99	557485	933081	53
new32	431140	980131	87	new100	565382	912023	29
new33	423243	995596	62	new101	559789	920249	51
new34	434101	1003492	107	new102	544324	908733	24
new35	400869	995267	192	new103	543995	888333	49
new36	398895	988686	78	new104	530834	935055	44
new37	398566	979802	65	new105	544982	939990	43
new38	395276	968945	52	new106	504512	953480	67
new39	399882	957758	84	new107	500893	987699	99
new40	413372	981448	64	new108	517673	997241	104

new41	419953	958087	26	new109	526228	999544	83
new42	422914	969603	14	new110	412385	956771	42
new43	411069	969603	70	new111	444301	985067	105
new44	429166	938016	15	new112	534454	968287	60
new45	416334	945255	54	new113	417792	992037	89
new46	436404	947558	57	new114	415250	973928	99
new47	449894	942952	36	new115	437173	967573	136
new48	469636	981777	99	new116	428595	948510	58
new49	465358	998886	35	new117	401589	965349	183
new50	478848	997570	90	new118	481971	997439	87
new51	490364	992634	63	new119	514696	1011740	83
new52	490364	1010731	70	new120	508660	981870	132
new53	480164	987041	85	new121	522004	924999	188
new54	514383	1012705	54	new122	479430	923410	93
new55	524254	1013034	53	new123	488961	958995	28
new56	520964	989015	99	new124	543609	988225	39
new57	506157	998886	103	new125	558859	958042	45
new58	543337	964009	84	new126	562036	929447	44
new59	553537	968616	78	new127	549328	920551	95
new60	564395	980131	74	new128	551552	899899	46
new61	561105	964009	107	new129	542655	883696	68
new62	563079	949532	131	new130	524546	907524	52
new63	565711	934397	117	new131	516920	954547	95
new64	568343	920249	51	new132	468309	960583	76
new65	575582	936700	105	new133	461637	1002200	51
new66	586769	946571	138	new134	446705	1011420	45
new67	585452	961048	116	new135	449246	996168	64
new68	575582	987370	83	new136	409214	1005060	67

As I illustrate above in the Determination of optimal number of stations using information entropy by using existing station here the same I used new candidate rainfall data and I can select the priority of each station below in the table 4.5, finally due to limitation of cost we can provide only 49 rain gauge stations.

Table 4- 5: Finalized Upper Awash Station Transformation Index

Station Priority	Transformation index	Optimal Redundant information passed (T/8.78)*100
new2	4.31	49.14
new3	4.32	49.15
new4	4.32	49.23
new5	4.40	50.14
new6	4.40	50.14
new7	4.42	50.36
new8	4.44	50.60
new9	4.45	50.73
new10	4.46	50.77
new11	4.47	50.88
new12	4.47	50.95
new13	4.50	51.24
new14	4.51	51.40
new15	4.51	51.42
new16	4.52	51.47
new17	4.54	51.68
new18	4.54	51.70
new19	4.55	51.86
new20	4.56	51.94
new21	4.60	52.42
new22	4.62	52.61
new23	4.65	52.99
new24	4.66	53.08
new25	4.68	53.30
new26	4.74	54.02
new27	4.74	54.02
new28	4.75	54.08
new29	4.77	54.27
new30	4.78	54.41
new31	4.79	54.52
new32	4.79	54.53
new33	4.80	54.70
new34	4.82	54.85
new35	4.82	54.93
new36	4.83	54.98
new37	4.87	55.42
new38	4.89	55.69
new39	4.90	55.77

new40	4.90	55.80
new41	4.90	55.81
new42	4.92	56.01
new43	4.93	56.14
new44	4.99	56.80
new45	4.99	56.87
new46	5.01	57.05
new47	5.14	58.51
new48	5.19	59.10
new49	5.26	59.87
new50	5.30	**60.40**

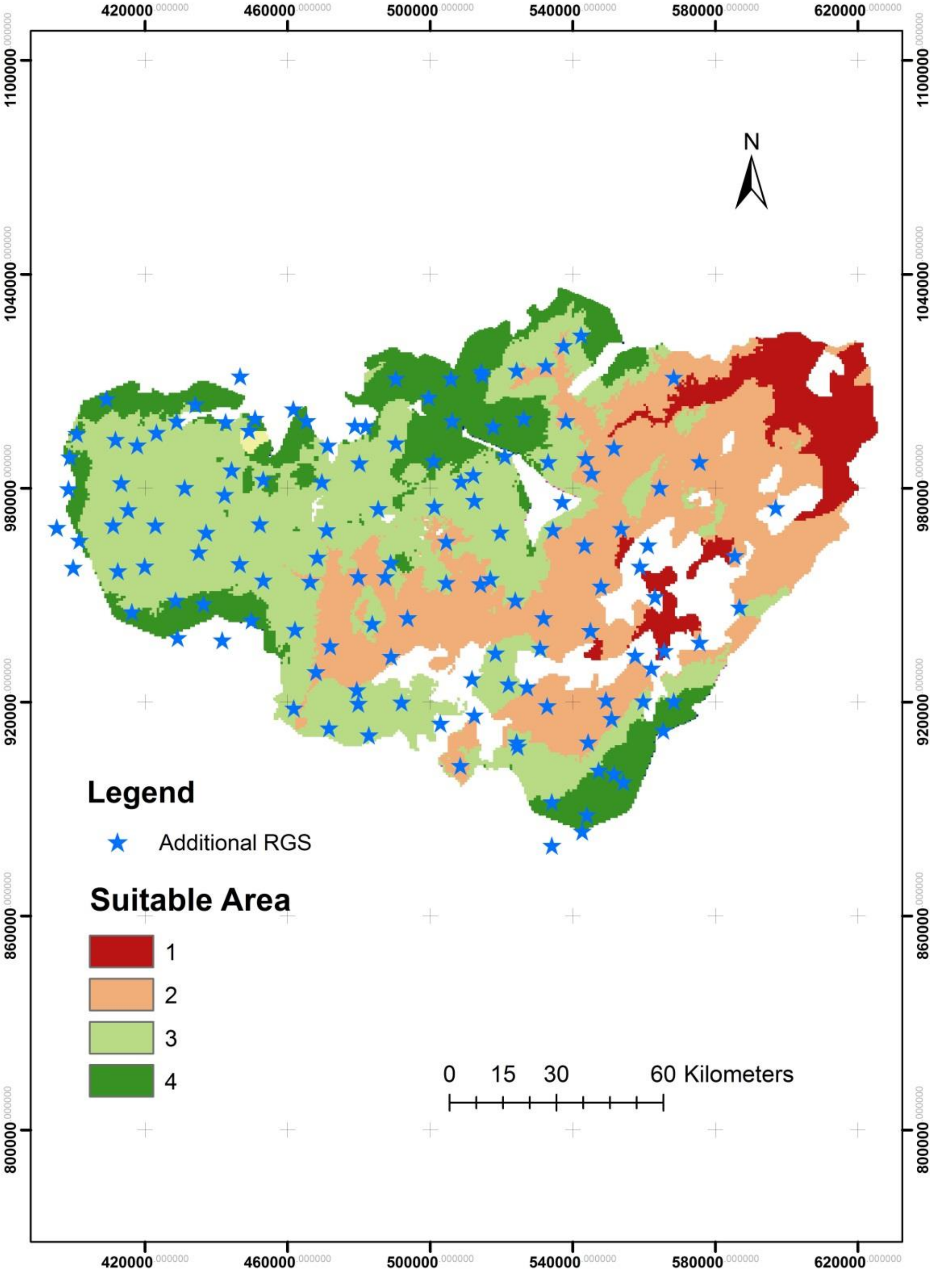


Fig 4- 7: Final proposed rain gauge stations in the Upper Awash river basin

5. Conclusion and Recommendations

5.1 Conclusion

It is self-evident that effective water monitoring networks are required for efficient water management. Although network design methodologies and applications have advanced significantly, a uniform design approach has yet to emerge. The optimization of rain gauge networks in the Upper Awash river basin using a geostatistical technique is presented in this study. The GIS-based MCDA was utilized to determine acceptable sites for rain gauge stations throughout the optimization phase. Following an extensive literature search, a number of appropriateness factors were identified and integrated suitably for the final site decision. The AHP was used to calculate factor weights as a methodological tool.

Without appropriate water monitoring networks, it is obvious that successful water management is impossible. Even though network design methodologies and applications have advanced significantly, a uniform design approach has yet to materialize. The use of a geostatistical method to optimize rain gauge networks in the Upper Awash river basin is presented in this study by using ordinary Kriging and by selecting appropriate Variogram model. A number of appropriateness factors were developed and suitably integrated for the final site decision after an extensive literature research. In comparison to the present rain gauge network, the result reveals that 49 RGS are required, consisting of 16 rain gauges. A total of 65 rain gauge stations were required to be operational in order to adequately run and manage the river basin's water resource projects.

5.2 Recommendation

The main recommendations arising from the results of the study were:

- WMO recommends most of the rain gauge stations to be install in the mountainous area but due to orographic effect couldn’t collect adequate data so in order to fill this gap I recommend further study.
- Owing to insufficient rain gauge stations, an additional 136 rain gauge stations in the Upper Awash River basin were needed, however due to financial constraints, only 49 RGS managed by Ethiopia's NMA could be erected first.
- Other water monitoring networks, such as stream flow networks, groundwater monitoring networks, and others, should be optimized to deliver the needed data with higher precision and at a lower cost than rain gauge networks.

6. References

A. Bechler, L. Bel, and M. Vrac (2015). Extremal t process conditional simulations: application to severe precipitation fields 109–127 in Spatial Statistics.

A. Chebbi, Z. K. Bargaoui, and M. Da Conceiço Cunha (2013). The development of a technique for optimizing rain gauge networks based on intensity-duration-frequency data. Hydrology and Earth System Sciences, vol. 17, no. 10, pp. 4259–4268, doi:10.5194/hess-17-4259-2013

A. K. Lohani, A. K. Lohani, A. K. Lohani, A. K. Lohani, A. K. Lohani, A (2012). National Institute of Hydrology, Roorkee.

A. Putthividhya and K. Tanaka (2012). Based on geo-statistical analysis of colocated elevation and humidity data, optimal rain gauge network design and spatial precipitation mapping have been developed. 3(2), 124 in International Journal of Environmental Science and Development.

Awadallah, A. G. Awadallah, A. G. Awadallah, A (2012). Using Kriging and entropy, find the best sites for rainfall stations. 36–41 in Int. J. Civil Environ. Eng., 12(1).

B. Manz, W. Buytaert, Z. Zulkafli, W. Lavado, B. Willems, L. A. Robles, and J. RodrguezSánchez (2016). The Tropical Andes' precipitation climatologies were integrated using high-resolution satellite gauges. Atmospheres, 121(3), 1190–1207. Journal of Geophysical Research: Atmospheres, 121(3), 1190–1207.

C. B. Jones, D. B. Kidner, L. Q. Luo, G. Ll. Bundy, and J. M. W. Jones, C. B. Jones, D. B. Kidner, L. Q. Luo, G. Ll. Bundy, and J. M. W. Jones, C. B. Jones (1996). A Multiscale Spatial Information System Database Design International Journal of Geographical Information Systems, 10, 901–920.

C. Camera, A. Bruggeman, P. Hadjinicolaou, S. Michaelides, and M. A. Lange (2017). A spatial rainfall generator is being tested for creating high-resolution precipitation predictions over orographically complicated terrain. 757–773 in Stochastic Environmental Research and Risk Assessment.

C. D. Lloyd, C. D. Lloyd, C. D. Lloyd, C (2005). Examining the impact of including elevation data into monthly precipitation forecasting in the United Kingdom. 128–150 in Journal of Hydrology, 308(1–4).

C. E. Shannon and W. Weaver (1963). In 1949, he published The Mathematical Theory of Communication. University of Illinois Press, Urbana, IL.

C. Yoo, K. Jung, and J. Lee (2008). Comparison of mixed and continuous distribution function applications in the evaluation of a rain gauge network utilizing entropy theory. 226–235 in Journal of Hydrologic Engineering, 13(4). .

D. Aredo and Y. Seleshi (2003). The Case of Small-Scale Irrigation Projects in North Wollo, Ethiopia.” Participatory Local Economic Development: The Case of Small-Scale Irrigation Projects in North Wollo, Ethiopia. At the Second IPRW, a paper was presented.

DELFT & DHV Consultants BV (1999). What is the best way to amend and complete rainfall data? New Delhi is the capital of India.

Delft Hydraulics and Dhv Consultants With Halcrow, Tahal, Ces, O. & J. Halcrow, Tahal, Ces, O. & J. Halcrow, Tahal, Ces (2001). Design Manual for Hydro-Meteorology (Vol. 3).

E. Barca, G. Passarella, and V. Uricchio (2008). Optimal expansion of the Apulian Regional Consortium for Crop Protection's rain gauge monitoring network. 375–386 in Environmental Monitoring and Assessment, 145(1–3).

E. Pardo-Igzquiza, E. Pardo-Igzquiza, E. Pardo-Igz (1998). Using Geo-statistics and simulated annealing, the optimal number and position of rainfall gauges for areal rainfall estimation. 206–220 in Journal of Hydrology, 210(1–4).

Ethiopia'sMinistry of Water Resources (2009). MANAGEMENT OF WATER INFORMATION AND KNOWLEDGE. Ethiopia's capital, Addis Ababa.

Feloni, Feloni, Feloni, Feloni (2018). GIS Techniques and Multicriteria Decision Analysis for Optimal Hydro meteorological StationNetworkDesign (Vol.22,pp.1–8). https://doi.org/10.1061/(ASCE)HZ.2153-5515.0000397.

F. Masoumi and R. Kerachian (2009). Groundwater quality monitoring networks should be redesigned in the most efficient way possible. 247–257 in Environmental Monitoring and Assessment.

F. Merid (2004). Addis Ababa, Ethiopia, National Nile Basin Water Quality Monitoring Baseline Report

G. Das (2002). Soil conservation engineering and hydrology India's Prenctic-Hall is located in New Delhi.

G. De Winnaar, G. P. Jewitt, and M. Horan (2007). In the Thukela River basin in South Africa, a GIS-based method was used to locate suitable runoff collection locations. Earth's Physics and Chemistry, 1058–1067.

D. R. Maidment, D. R. Maidment, D. R. Maidment (1993). Hydrology handbook. Mcgraw-Hill,Inc.

H.-C. Yeh, Y.-C. Chen, C.-H. Chang, C.-H. Ho, and C.-H. Wei (2017). Using Radar and Entropy to Optimize Rainfall Networks Entropy, vol. 19, no. 10, p. 553.
https://doi.org/10.3390/e19100553

H. C. Yeh, Y. C. Chen, C. Wei, and R. H. Chen (2011). Rainfall network design using entropy and Kriging. 343–355 in Paddy and Water Environment, 9(3). https://doi.org/10.1007/s10333-010-0247-x

H. Feki, M. Slimani, and C. Cudennec (2016). Optimization of a rainfall monitoring network expansion using geostatistics: the instance of climatically diverse Tunisia. nh2016256, Hydrology Research.

H. Hurni, H. Hurni, H. Hurni, H (1993). Land degradation, hunger, and land resource possibilities in Ethiopia, according to the World Soil Erosion and Conservation Organization.

H. Mahmoudi-meimand, S. Nazif, and R. A. Abbaspour (2016). A GIS-based method for optimizing a rain gauge network utilizing geostatistics and entropy principles. 1–20 in Journal of Spatial Science, 8596. https://doi.org/10.1080/14498596.2015.1030789

J. Amorocho and B. Espildora (1973). In hydrologic systems and models, entropy is used to measure uncertainty. 1511–1522. Water Resources Research, vol. 9, no. 6, pp. 1511–1522.

J. J. Carrera-Hernández and S. J. Gaskin (2007). Daily precipitation and temperature in the Mexican Basin analyzed spatially and temporally. 231–249. Journal of Hydrology, vol. 336, no. 3–4, pp. 231–249.

J. Malczewsk, J. Malczewsk, J. Malczews (1999). Multi-criteria decision analysis using GIS. John Wiley & Sons is a publishing house based in the United Kingdom.

J. Nemec and A. J. Askew (1986). In network-design theories, there is a mean and a variance. IN: Integrated Hydrological Networks Design. Publication of the IAHS (158).

K. Cheng, Y. Lin, and J. Liou (2007). Geostatistics were used to evaluate and expand the rain gauge network. https://doi.org/10.1002/hyp

K. Cheng, Y. Lin, and J. Liou (2008). Geostatistics were used to evaluate and improve the raingauge network. 2554–2564 in Hydrological Processes: An International Journal, 22(14).

K. Johnston, J. M. VerHoef, K. Krivoruchko, L. N. Johnston, J. M. VerHoef, J. M. VerHoef, J. M. (2001). ArcGIS Geo-statistical Analyst is used. ESRI's ArcGIS Manual. Redlands, California, United States of America.

L. Alfonso, A. Lobbrecht, and R. Price (2010). An technique based on information theory for locating monitoring water level sensors in polders. Research on Water Resources, 46 (3).

M. Haggag, A. A. Elsayed, and A. G. Awadallah (2016). A case study in northern Oman was used to evaluate a rain gauge network in dry locations using a geostatistical method. Arabian Journal of Geosciences is a publication dedicated to geosciences in the Arabian Peninsula. https://doi.org/10.1007/s12517-016-2576-6

M. Khairul, B. Mohd, F. Yusof, Z. M. Daud, and Z. Yusop (2016). Geostatistics and particle swarm optimization were used to construct an optimal rain gauge network in Johor.

M. Kilibarda, T. Hengl, G. B. M. Heuvelink, B. Gräler, E. Pebesma, M. Perec Tadi, and B. Bajat (2014). At 1 km resolution, spatiotemporal interpolation of daily temperatures for worldwide land regions. 119(5), 2294–2313, Journal of Geophysical Research: Atmospheres.

M. L. Morrissey, J. A. Maliekal, J. S. Greene, and J. S. Wang (1995). Using rain gauge networks to investigate the uncertainty of simple geographical averages. 31(8), 2011–2017, Water Resources Research.

M. R. Shaghaghian and M. J. Abedini (2013a). Using a combination of geostatistical and multivariate methods, a rain gauge network was designed. 259–269. Scientia Iranica, vol. 20, no. 2, pp. 259–269.

M. R. Shaghaghian and M. J. Abedini (2013b). Using a combination of geostatistical and multivariate methods, a rain gauge network was designed. 259–269. Scientia Iranica, vol. 20, no. 2, pp. 259–269. https://doi.org/10.1016/j.scient.2012.11.014

M. Shafiei, B. Ghahraman, B. Saghafian, S. Pande, S. Gharari, S. Davary, K. Shafiei, B. Ghahraman, B. Saghafian, B. Saghafian, B. Saghafian, B. Saghafian, B. Saghafian, B. (2014). A probabilistic GIS-based method was used to assess rain-gauge networks. 551, Hydrology Research, 45(4–5). https://doi.org/10.2166/nh. 2013.042

M. Shafiei, E. Water, B. Ghahraman, B. Saghafian, S. Gharari, M. Shafiei, E. Water, B. Ghahraman, B. Saghafian, B. Saghafian, B. Saghafian, B. Saghafi (2013). A probabilistic GIS-based method was used to assess rain-gauge networks (April 2016). https://doi.org/10.2166/nh. 2013.042

M. Zaharifudin, M. Ali, and F. Othman (2018). By combining cross-validation and the geostatistical method, rain gauge network optimization in a tropical metropolitan region was achieved. 474–491. https://doi.org/10.1080/02626667.2018.1437271 Hydrological Sciences Journal, 63(3), 474–491.

N. HAREGEWEYN, M. Haile, G. Govers, and S. Deckers (2006). 230(August 2005), 211–230. Reservoirs In Tigray (Northern Ethiopia): Characteristics And Sediment Deposition Problems

N. Harmancioglu and V. Yevjevich (1987). The exchange of hydrologic data between river sites. 103–118 in Journal of Hydrology, 91(1–2).

N. T. Hong, P. Thi, T. Truc, N. D. Liem, and N. K. Loi (2016). Using GIS, Optimal Number and Location of Meteo-Hydrological Monitoring Networks for the Vu Gia – Thu Bon River Basin, 6(3), 324–328.

P. Camberlin (1997). Rainfall anomalies in the Nile's source area and their link to the Indian summer monsoon Journal of Climate, vol. 10, no. 6, pp. 1380–1392.

P. F. Krstanovic and V. P. Singh (1992a). I. Theoretical development of entropy-based evaluation of rainfall networks. 279–293. Water Resources Management, vol. 6, no. 4, pp. 279–293.

P. F. Krstanovic and V. P. Singh (1992b). II. Application of entropy in the evaluation of rainfall networks. 295–314, Water Resources Management, vol. 6, no. 4, pp. 295–314.

S. K. Adhikary, A. G. Yilmaz, and N. Muttil (2015). Rain gauge network optimization in the Middle Yarra River basin, Australia. 2582–2599 in Hydrological Processes, 29(11). https://doi.org/10.1002/hyp.10389

U. Haberlandt, U. Haberlandt, U. Haberlandt (2007). For a large-scale severe rainfall event, geostatistical interpolation of hourly precipitation from rain gauges and radar. 144–157 in Journal of Hydrology, 332(1–2).

W. H. Cheung, G. B. Senay, and A. Singh (2008). Annual and seasonal rainfall trends and geographical distribution in Ethiopia. 1723–1734 in International Journal of Climatology, 28(13).

W. Degefu, W. Degefu, W. Degefu (1987). Some elements of Ethiopia's meteorological drought. Cambridge University Press is a Cambridge-based publishing house.

WMO stands for World Meteorological Organization (1994). Guide on Hydrological Practices from the World Meteorological Organization. Geneva is a city in Switzerland.

WMO stands for World Meteorological Organization (2003). Current state and trends in hydrological data management. Geneva is a city in Switzerland.

WMO stands for World Meteorological Organization (2008). Hydrological Practices Guide (Vol. I). Switzerland, Geneva.

Y. Jung, H. Kim, J. Baik, and M. Choi (2014). Rain-Gauge Network Evaluations for Semi-Mountainous Regions Using Spatiotemporal Correlation Structure 25th Conference on Terrestrial, Atmospheric, and Oceanic Sciences (2).

Y. Mogheir, Y. Mogheir, Y. Mogheir (2003). The Gaza Strip case study was used to assess and rebuild groundwater quality monitoring networks using the entropy theory. Coimbra University is located in Portugal.

Y. Seleshi and U. Zanke (2004). Ethiopian rainfall and rainy days have recently changed. 973–983 in International Journal of Climatology.

7. Appendix

Appendix I: List of stations in the Awash River basin and their average monthly Rainfall

S.no	Latitude (m)	Longitude (m)	Candidate name	Elevation (m)	Monthly avg. rainfall (mm)
1	996668	428556	Addisalem	2372	89.03
2	1004010	446886	Asgori	2072	89.27
3	970876	426678	Debrezeyt	1900	76.24
4	981872	476177	Ginchi	2232	96.55
5	965283	498167	Koka dam	1618	76.25
6	961622	455998	Melka-Kunture	960	94.71
7	950551	476159	Teji	2091	84.67
8	950545	509170	Tulu bolo	2190	105.30
9	915560	544050	Wenji	2372	74.34
10	933970	525685	Awash Melka	2190	49.03
11	935809	518345	Sebeta	2220	91.00
12	933968	522015	Addis Ababa	2386	101.70
13	1013188	507324	Akaki	2057	91.43
14	978231	435855	Entoto	2903	109.38
15	998461	467028	Bole	2354	98.43
16	968968	498167	Kotebe	2755	100.01

Appendix II: Candidate stations in the Upper Awash river basin

S.no	Latitude (m)	Longitude(m)	Candidate name	March	April	May	June	July	August	September	October	November
1	996668	428556	Addisalem	34.62	44.88	59.21	111.94	221.00	207.06	101.35	12.50	8.68
2	1004010	446886	Asgori	28.62	58.24	54.00	114.50	221.65	225.21	86.59	11.68	2.97
3	970876	426678	Debrezeyt	29.1	46.3	44.0	78.0	193.2	193.7	84.9	14.0	3.0
4	981872	476177	Ginchi	38.76	65.21	72.62	130.74	209.32	212.35	116.32	18.35	5.26
5	965283	498167	Koka Dam	30.76	37.59	44.35	46.68	211.26	194.56	94.44	21.53	5.03
6	961622	455998	Melka-Kunture	61.71	76.38	56.10	102.97	226.98	212.50	88.80	23.26	3.71
7	950551	476159	Teji	35.35	56.12	58.26	109.35	201.59	201.50	84.71	11.18	3.94
8	950545	509170	Tulu bolo	30.85	47.71	68.32	179.65	252.26	250.47	105.26	11.44	1.71
9	915560	544050	Wenji	39.62	38.85	45.65	46.79	203.59	181.68	83.00	23.85	6.03
10	933970	525685	Awash Melka	44.74	38.74	33.00	21.26	104.94	116.38	53.68	19.06	9.44
11	935809	518345	Sebeta	28.32	49.62	51.65	97.06	218.82	242.06	113.79	13.65	4.03
12	933968	522015	AddisAbeba	43.03	67.24	66.21	109.65	228.62	245.76	127.56	23.29	3.97
13	1013188	507324	Akaki	34.15	63.12	56.59	93.82	222.32	231.15	106.76	12.32	2.68
14	978231	435855	Entoto	39.71	59.71	64.12	121.41	265.26	285.59	125.68	16.41	6.53
15	998461	467028	Bole	40.44	64.47	62.56	105.97	226.35	244.06	119.47	19.12	3.47
16	968968	498167	Kotebe	37.56	66.94	64.79	112.62	233.59	249.44	113.38	17.94	3.85

Appendix III: Trans information entropy index matrix

S.no	Latitude (m)	Longitude (m)	Candidate name	Marginal entropy	S2	S3	S4
1	996668	428556	Addisalem	3.45476	1.684	0.824	2.031
2	1004010	446886	Asgori	3.49064	1.669	0.848	1.904
3	970876	426678	Debrezeyt	3.54395	1.268	1.081	1.615
4	981872	476177	Ginchi	3.33772	1.796	0.723	….........
5	965283	498167	Koka Dam	3.60598	0.845	1.318	1.058
6	961622	455998	Melka-Kunture	3.53	1.205	1.049	1.429
7	950551	476159	Teji	3.45847	1.743	0.836	2.117
8	950545	509170	Tulu bolo	3.35037	…......... .	….........	….........
9	915560	544050	Wenji	3.63825	…......... .	….........	….........
10	933970	525685	Awash Melka	3.56516	0.559	….........	….........
11	935809	518345	Sebeta	3.63234	1.279	1.011	1.696
12	933968	522015	AddisAbeba	3.48772	1.290	1.012	1.889
13	1013188	507324	Akaki	3.49987	1.251	1.074	1.725
14	978231	435855	Entoto	3.45758	1.379	1.002	1.759
15	998461	467028	Bole	3.50619	1.305	1.030	1.856
16	968968	498167	Kotebe	3.51443	1.402	0.981	1.938

S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15
1.335	2.315	1.903	1.640	1.917	2.190	1.958	…………	…………	…………	…………
1.243	2.999	1.900	1.846	1.893	…………	…………	…………	…………	…………	…………
1.726	2.001	2.399	1.896	2.305	2.126	2.494	2.060	2.888	2.888	…………
…………	…………	…………	…………	…………	…………	…………	…………	…………	…………	…………
…………	…………	…………	…………	…………	…………	…………	…………	…………	…………	…………
1.338	1.874	1.506	…………	…………	…………	…………	…………	…………	…………	…………
1.201	…………	…………	…………	…………	…………	…………	…………	…………	…………	…………
…………	…………	…………	…………	…………	…………	…………	…………	…………	…………	…………
…………	…………	…………	…………	…………	…………	…………	…………	…………	…………	…………
…………	…………	…………	…………	…………	…………	…………	…………	…………	…………	…………
1.519	1.813	…………	…………	…………	…………	…………	…………	…………	…………	…………
1.462	1.898	2.736	1.611	…………	…………	…………	…………	…………	…………	…………
1.588	2.012	2.544	1.815	2.698	2.075	3.013	1.949	…………	…………	…………
1.502	2.073	3.050	1.666	2.537	2.195	2.811	2.118	2.657	…………	…………
1.483	1.988	2.890	1.680	3.700	2.009	…………	…………	…………	…………	…………
1.419	2.280	2.601	1.766	2.787	2.355	3.161	2.074	2.989	2.989	**2.907**

Appendix IV: Priority order and redundant information passed in the upper Awash River Basin

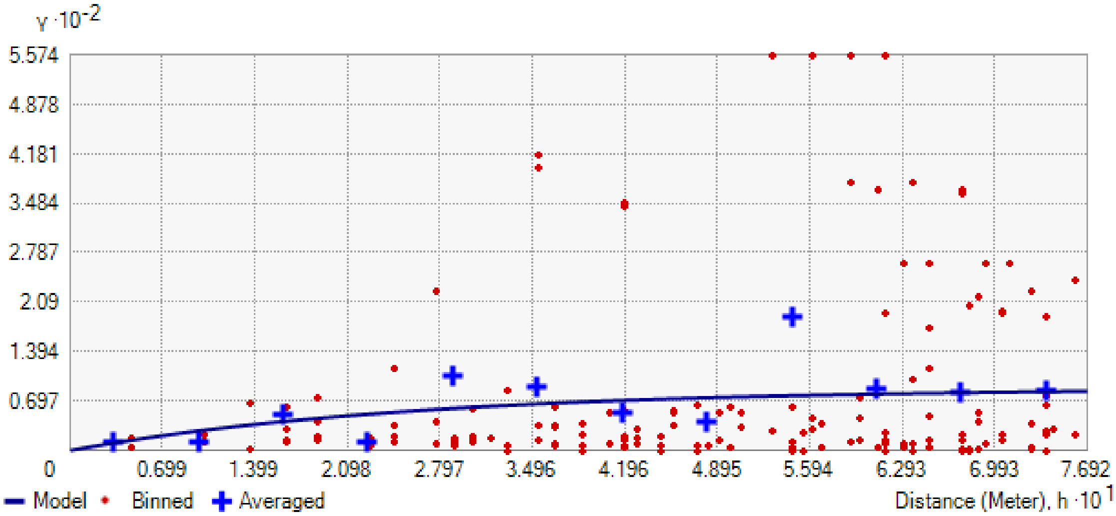
The Priority Of each Station	Station Name	Transformation Index (T)	Optimal redundant Information Passed (T/2.91) *100 %
X(1)	WENJI	………	………
X(2)	TULUBOLO	0.83	28.70
X(3)	AWASH MELKA	0.56	19.22
X(4)	GINCHI	0.72	24.88
X(5)	KOKA DAM	1.06	36.41
X(6)	TEJI	1.20	41.31
X(7)	MELKAKUNTURE	1.51	51.82
X(8)	ADDIS ABEBA	1.61	55.42
X(9)	SEBETA	1.81	**62.35**
X(10)	ASGORI	1.89	65.14
X(11)	BOLE	2.01	69.12
X(12)	ADDISALEM	1.96	67.37
X(13)	AKAKI	1.95	67.05
X(14)	ENTOTO	2.66	91.41
X(15)	DEBRE ZEYT	2.89	99.36
X(16)	KOTEBE	2.91	100.00

Appendix V: The recommended optimal redundant information to be reduced in a rain gauge network

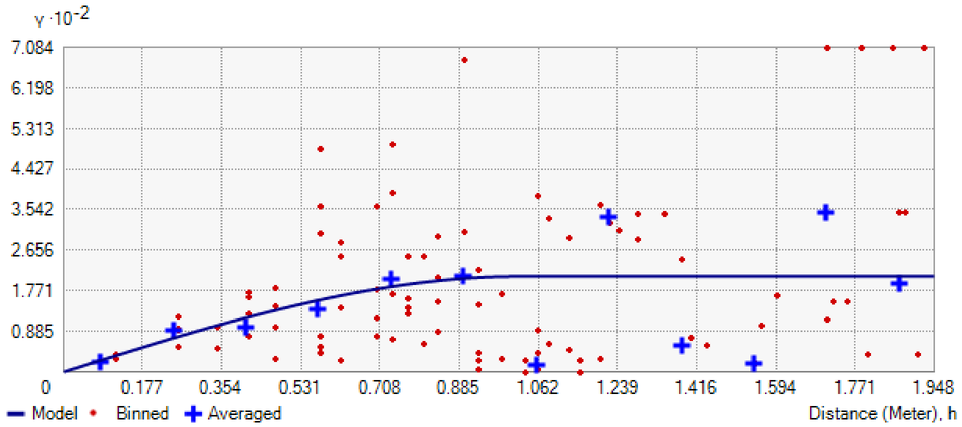
S.no	Candidate name	Mean	Std.	CV
1	Tulu bolo	105.30	98.72	0.94
2	Awash Melka	49.03	37.59	0.77
3	Ginchi	96.55	76.67	0.79
4	Koka dam	76.25	75.89	1.00
5	Teji	84.67	73.99	0.87
6	Melka kunture	94.71	77.28	0.82
7	Addis Ababa	101.70	86.01	0.85
	Average	86.89	75.16	0.86

Appendix VI: Selected Variogram model for nine wet months

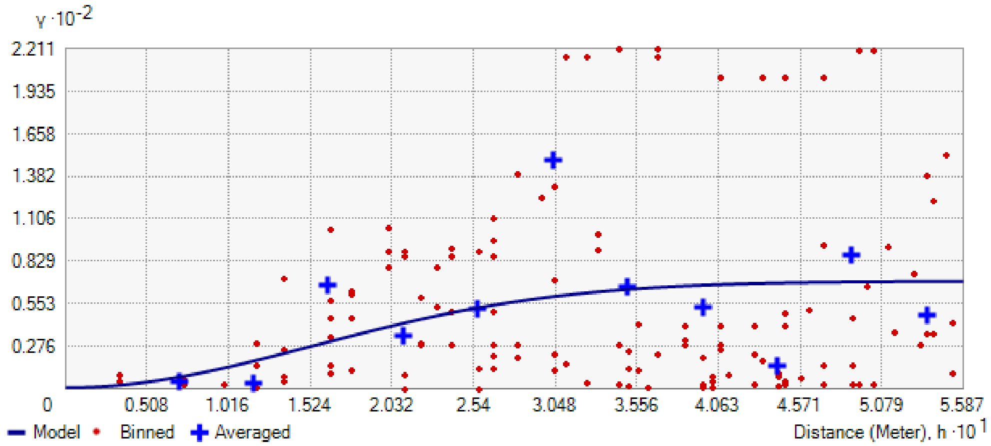
1. March (Exponential)



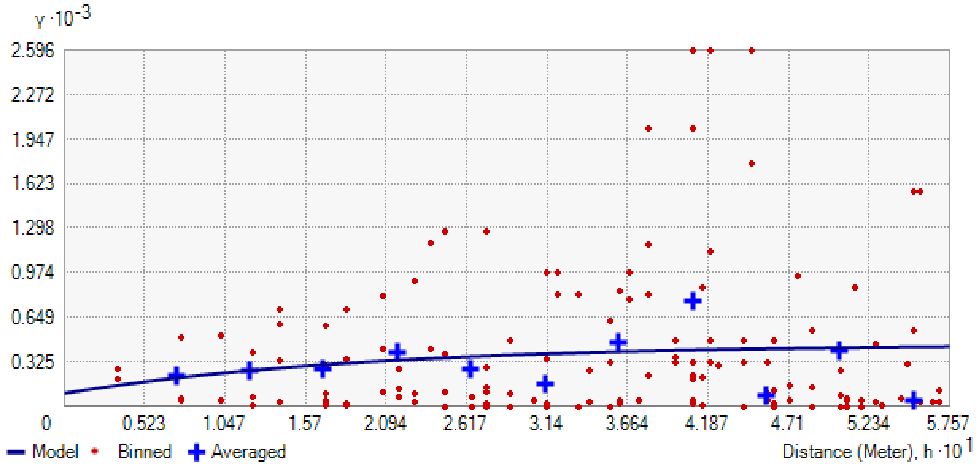
2. April (Gaussian)



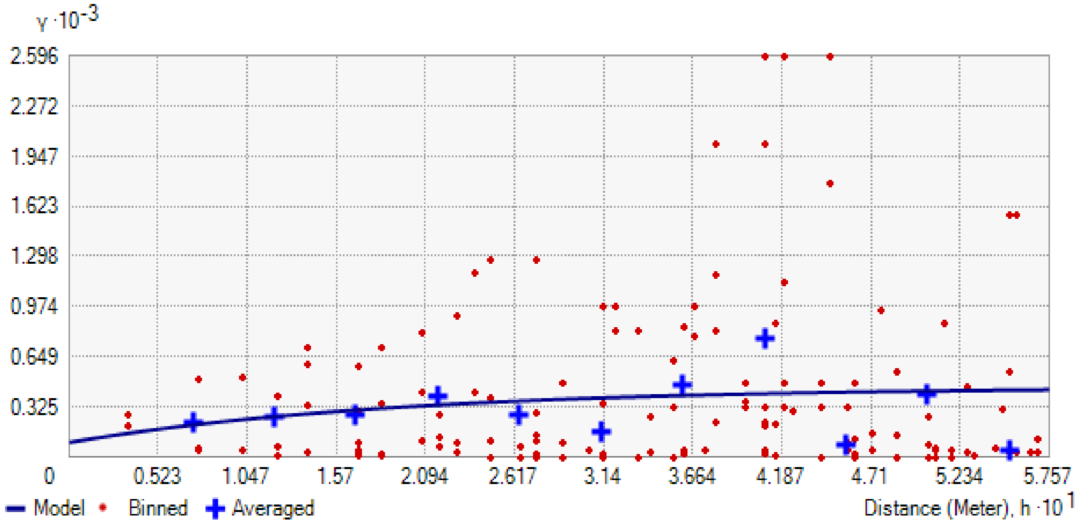
3. May (Gaussian)



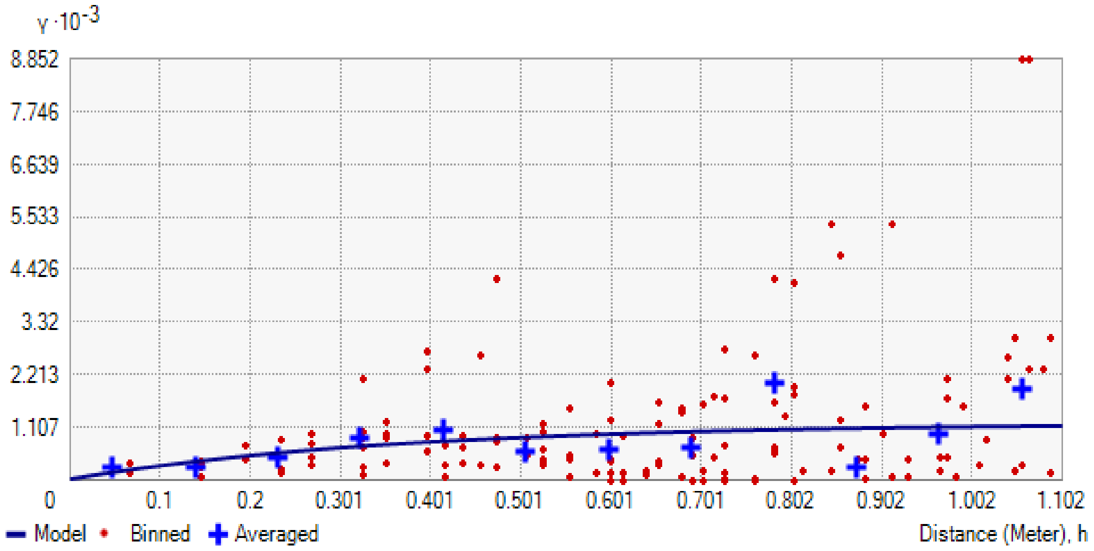
4. June (Exponential)



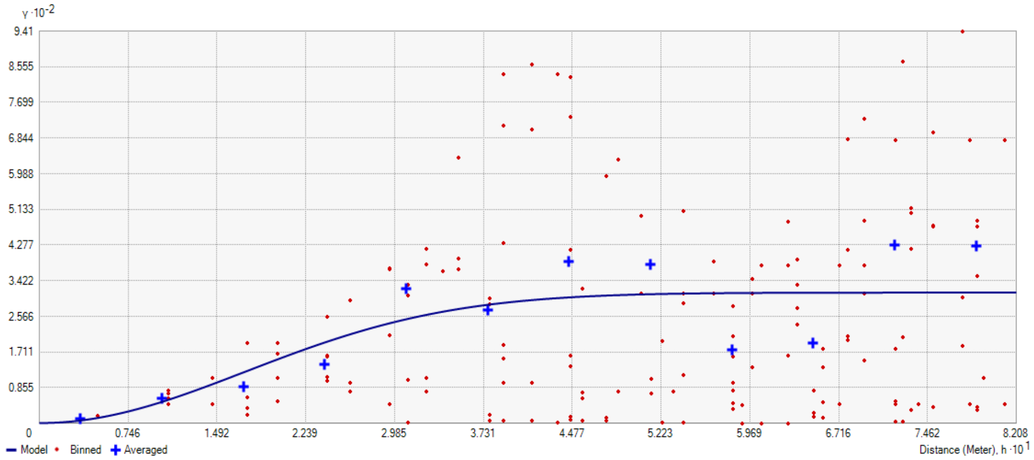
5.July (Exponential)



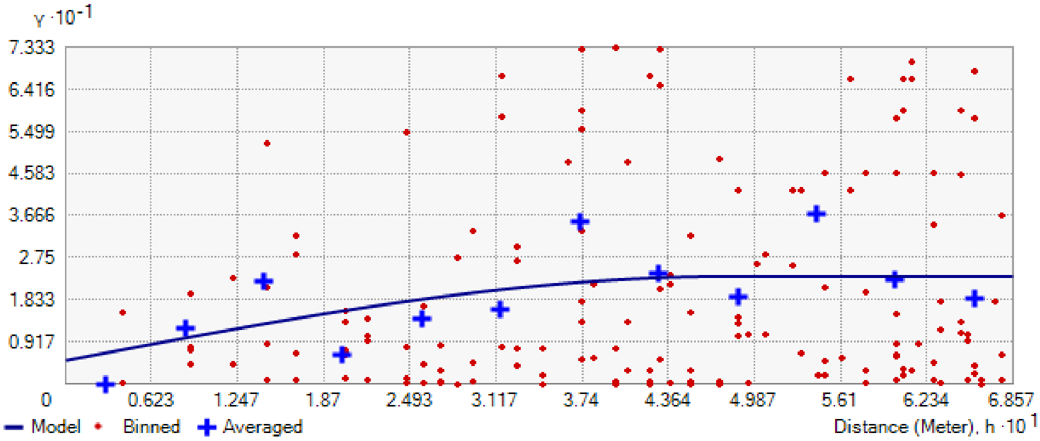
6.August (Exponential)



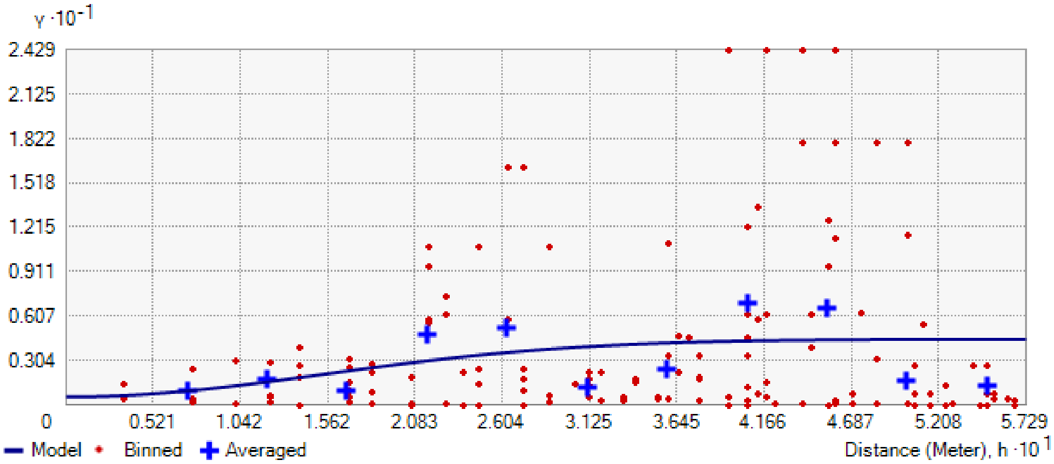
7. September (Gaussian)



8. October (Spherical)



9. November (Gaussian)



Appendix VI the Remaining Upper Awash Station Transformation Index

Station Priority	Transformation index	Optimal Redundant information passed (T/8.78)*100
new51	5.32	60.63
new52	5.35	60.92
new53	5.36	61.07
new54	5.40	61.51
new55	5.41	61.64
new56	5.53	62.97
new57	5.54	63.04
new58	5.55	63.18
new59	5.57	63.39
new60	5.58	63.54
new61	5.67	64.62
new62	5.67	64.62
new63	5.68	64.74
new64	5.69	64.78
new65	5.78	65.81
new66	5.79	65.96
new67	5.81	66.17
new68	5.83	66.37
new69	5.83	66.45
new70	5.90	67.19
new71	5.90	67.19
new72	5.99	68.25
new73	6.03	68.71
new74	6.06	68.99
new75	6.06	69.03
new76	6.09	69.39
new77	6.13	69.87
new78	6.14	69.98
new79	6.24	71.03
new80	6.25	71.16
new81	6.25	71.21
new82	6.31	71.86
new83	6.33	72.04
new84	6.37	72.60
new85	6.37	72.60
new86	6.42	73.06
new87	6.42	73.12

new88	6.43	73.27
new89	6.44	73.29
new90	6.45	73.41
new91	6.50	74.00
new92	6.64	75.58
new93	6.74	76.78
new94	6.75	76.84
new95	6.75	76.87
new96	6.76	77.01
new97	6.88	78.36
new98	6.95	79.18
new99	7.01	79.81
new100	7.04	80.14
new101	7.06	80.45
new102	7.06	80.45
new103	7.06	80.46
new104	7.10	80.84
new105	7.33	83.50
new106	7.40	84.33
new107	7.79	88.71
new108	7.84	89.31
new109	7.90	89.92
new110	8.38	95.46
new111	8.64	98.37
new112	8.74	99.58
new113	8.78	100.00