Addis Ababa University
Addis Ababa Institute of Technology
School of Postgraduate Studies
Civil and Environmental Engineering Department.

Thesis Submission to the School of Postgraduate Studies in Partial Fulfillment of the Requirements for the Degree of Master of Science in Hydraulic Engineering at Addis Ababa Institute of Technology, Addis Ababa University, Ethiopia;

on

Development of Intensity Duration Frequency(IDF) Curves for Bahir Dar City from Daily Rainfall Data by Using Simple Scaling Method,

Bahir Dar, Ethiopia.

Advisor:
Agizew NIGUSSIE ENGIDA (Dr.)

Prepared and submitted by:
Simeneh GEBEYEHU MELESSE

Addis Ababa

Ethiopia

December 2016

CERTIFICATION

It is certainly to declare that the thesis work named under "Development of Intensity Duration Frequency(IDF) Curves for Bahir Dar City from Daily Rainfall Data by Using Simple Scaling Method" is my own work. This thesis work was submitted to the partial fulfillment of Master of Science (M.Sc.) in hydraulic engineering from school of postgraduate studies in Addis Ababa University Institute of Technology, Addis Ababa University, Ethiopia. I would like to confirm that this thesis work has not been published by other professionals or graduates to be accepted for the award of any other degree from other university except proper citations or acknowledgements were given to the notes taken from the publications.

Prepared and submitted by:

Simeneh GEBEYEHU: ……………… ……………..
ID.NO:GSE/1553/05 Signature Date

Approved by:

Agizew NIGUSSIE (Dr.) ………………… ……………..
Advisor, Signature Date
Dean, School of Civil
and Environmental Engineering Department,
Addis Ababa Institute of Technology,
Addis Ababa University, Ethiopia.

Examiners:

1.Yilma SELESHI (Dr.) ...…………………………… ………………………
Signature: Date:

2. Yenesew MENGISTIE (Dr.) ...…………………………… ………………………
Signature: Date:

ABSTRACT

This research mainly dealt the development of intensity duration frequency curves for Bahir Dar city from annual maximum daily rainfall data by using simple scaling method. Regarding the rainfall data, 41 years of daily rainfall data and 4 years sub hourly and hourly rainfall data were collected from National Meteorological Agency (NMA) in Addis Ababa, Ethiopia.

Hydrological data checking was conducted which were mainly testing for independence, stationary, homogeneity and outliers of the rainfall data. And the data were found independent, stationary, homogenous and with no outliers. There were missing data for 12 months in 1991, 2005 and 2014 and in-filling of the data were done using simple average method.

By using easy fit software, it was possible to know that the annual maximum daily rainfall data to Bahir Dar city followed the general extreme value(GEV) distribution function and the three parameters of the distribution function were also determined. The annual maximum daily rainfall were found to be time scale invariant which helped to develop the intensity duration frequency relationships and curves using simple scaling method. The scale exponents of the scale factor was found constant in two ranges of durations $15\min \le d < 1hr$ and $1hr \le d \le 24hr.$ Because of the distinction of the linearity of the slopes for these two durations, it was possible to detect the slopes difference at one hour duration.

The automatically 15 minutes intervals recorded for 4 years (2012-2015) rainfall data were the back bone of the successful development of the curves for Bahir Dar city. The determination of the design daily rainfall for 2 years, 5 years, 10 years, 25 years, 50 years, and 100 years return periods was done. The comparison of the regional IDF curves and simple scaling method generated IDF curves showed that the intensity of the simple scaling method generated curves were higher value than the intensities generated using the regional IDF curves. Moreover, the 24 hour rainfall values calculated from simple scaling model fall above the design rainfall values calculated from observed rainfall data for different selected return periods using frequency factor method. In addition, the comparison of the regional IDF curves, observed IDF curves, and simple scaling IDF curves were also analyzed in order to understand and check the performance of simple scaling modeling with regards to the observed rainfall data. Based on this analysis, it was learnt that the intensities or rainfall values from simple scaling model fall above the values of the observed rainfall or intensity values. In contrary, the values from regional IDF curves fall below both observed rainfall and design rainfalls calculated values. This proved that simple scaling model performs better than the regional IDF curves. From the study, it was also possible to learn that the intensities at shorter rainfall durations are higher than the intensities at longer rainfall durations.

ACKNOWLEDGEMENTS

Firstly, I would like to express my deep seated and sincere acknowledgment and gratitude to my advisor, Agizew NIGUSSIE (Dr.), for his determined guidance, coaching, valuable and super essential comments from the stage of the thesis proposal inception to the final indispensible findings of the thesis. I would imagine without his clinical and continuous guidance and support where I can stage myself by this time. I would also like to express my high appreciation to my advisor‟s professionalism and depth of rich knowledge which enabled me to reach to this level of successful arena, which equipped me with the required wonderful skills and knowledge for my future endeavors.

Secondly, it is mind urging to fully acknowledge and thank to the support and guidance that I received from my classmates and instructors in the course of the years to conclude my study at Addis Ababa Institute of Technology successfully. Equally important, I would also like to thank to the staff of the National Meteorological Agency(NMA) for enabling me to get the daily and hourly rainfall data without which this study would have been futile.

Thirdly, I am extending my everlasting appreciation and gratitude to my family for their support to my successes achieved during the study period. Words never explained how deep physical and moral support that my wife along with my kids availed to me during my study. I am also privileged to extend my gratitude to UNHCR management who let me go for attending classes, exams and regular thesis findings presentation in Addis Ababa Institute of Technology from deep field of my regular works located very far away from the capital Addis.

Lastly, the glory of God was beyond mind when I realized and compared the challenges that I came through and the successes that I am crowning today.

DEDICATION

To my family!

Table of Contents

Lists of tables **pages**

Lists of figures

Abbreviations

AA	Addis Ababa
CDF	Cumulative distribution function
ECDF	Empirical cumulative distribution function
EV1	Extreme value type I
EV II	Extreme value type II
G-B	Grubbs and Becktest
GEV	General extreme value
IDF	Intensity duration frequency
LN	Lognormal
LP	Log Pearson
LMS	Least median square
Med	Median
M-W	Mann-Whitney
NMA	National Meteorological Agency
SS	Simple scaling
PWM	Probability weighted moments
W-W	Wald-Wolfowitz
UNHCR	United Nations High Commissioner for Refugees
US	United States
U.S.S.R	Union of Soviet Socialist Republics

CHAPTER 1: INTRODUCTION

1.1. Background

Bahir Dar city is located on the southern shore of lake Tana at latitude of $11^0 35$‟N and longitude of $37^0 23$‟E in Ethiopia and it is the capital of Amhara region. It is rapidly growing and expanding city. According to Ethiopia Meteorology Agency, the annual mean precipitation of Bahir Dar city is around 900mm, where 54% of the rain falls in July and August only and 3% falls during dry months, the rest falls in the remaining months. The urban storm runoff and domestic waste receiving water body, Lake Tana, with surface area of 3150km^2 is the largest fresh water body in the country and the source of Blue Nile

In Bahir Dar city, the degradation of Lake Tana water quality was caused by urban storm runoff and it is a serious problem and affecting the ecosystem of the lake and significant portion of the community living around the gulf principally Bahir Dar city residents. The dramatic expansion of the city has resulted in an increase of the area of that doesn‟t have the capacity to store and infiltrate water(impervious surface).The transformation of these natural land surfaces in to impervious surfaces(like streets, parking lots, foot path and other pavements, etc.) are radically decreasing the rate of infiltration and hereby increasing surface runoff volume from precipitation. This change in land use increases impervious cover leads to flooding, erosion, habitat degradation, and water quality impairment. Researchers have identified storm water runoff as a major contributor to water quality degradation in urbanizing water shed. Generally discharges of large storm event may shock the receiving water body many times greater than the small ones (Wondie, 2009).

Thus, the rainfall Intensity Duration Frequency(IDF) relationship is one of the most commonly used tools for various engineering projects against floods. IDF curves expressed the relationship between rainfall intensity, duration and return periods of the rainfall. In order to construct IDF curves, a historical series of the maximum rainfall intensities at a higher time resolution(with sub hourly interval) is required. Such rainfall data are available only from a limited number of rain gauging stations or may not be available at all (Jaleel and Farwan, 2013).

Hence, in order to properly combat these types of major flooding problems in the city, engineers required sub daily rainfall data so that the required design of the hydraulic structures attained. In the contrary, Bahir Dar city has no any reliable and relatively accurate sub daily data available for these types of engineering interventions in order to balance the effect of surface runoff developed from the rainfall. So, in order to generate adequate and most reliable sub daily rainfall intensity data to Bahir Dar city for further engineering actions, it was aimed to develop intensity duration frequency (IDF) curves using simple scaling method.

My research helped to have intensity duration frequency curves for sub daily rainfall data that were generated from the annual maximum daily rainfall data of Bahir Dar city using simple scaling method.

1.2. Problem Statement

The regional intensity duration frequency (IDF) curves developed in Ethiopia were not having good accuracy with regards to estimating the design flood for the catchment of a given size. This was the main reason and logic to study the sub daily intensity duration frequency curves for Bahir Dar city using simple scaling method so that it is possible to determine the design flood for constructing drainage and flood protection structures in the city as an example. Doing the IDF curves from sub hourly rainfall data is relatively very accurate and has a very good level of efficiency for fixing the intensity of the Bahir Dar city, by using the specific concentration time and the selected return period of the design flood. The sub daily IDF curves have more advantage than the regional IDF curves when it comes to design the drainage structures in a specific places.

At the moment, in order to design drainage structures at Bahir Dar city, one should use the regional IDF curves developed from the rainfall data using stations at Gondar, Debre Tabor, Bahir dar, Debre Markos, Fitche and Addis Ababa. It is very clear how disperse rainfall data to represent these far regions in one IDF relationship while all the mentioned locations having different hydrological data characteristics. This is a clear justification to show how the rainfall data in the regional curves is less accurate as compared to the rainfall data collected just at Bahir Dar city. This less accurate regional IDF curve has a clear impact when it comes to affecting the drainage structures in Bahir Dar as well as Addis Ababa as an example, by developing road side flooding especially during rainy season. This takes us to a clear conclusions that the flooding of the newly built road was mainly due to the contribution from under design rainfall calculated from the regional intensity duration frequency curves (ERA, 2013).

The degradation of Lake Tana water quality was caused by urban storm runoff was a serious problem and affecting the ecosystem of the lake and significant portion of the community living around the gulf principally Bahir Dar city residents. The dramatic expansion of the city resulted in an increase of the area of that doesn‟t have the capacity to store and infiltrate water(impervious surface).The transformation of these natural land surfaces in to impervious surfaces(like streets, parking lots, foot path and other pavements, etc.) were radically decreasing the rate of infiltration and hereby increasing surface runoff volume from precipitation. This change in land use increased impervious cover leads to flooding, erosion, habitat degradation, and water quality impairment. Researchers have identified storm water runoff as a major contributor to water quality degradation in urbanizing water shed. Generally discharges of large storm event may shock the receiving water body many times greater than the small ones (Wondie, 2009)

1.3. Objectives:

- To develop intensity duration frequency(IDF) curves for Bahir Dar City in Ethiopia from annual maximum daily rainfall data by using simple scaling method.

1.3.1. Specific Objectives

- Evaluate the daily and hourly data of Bahir dar city.
- Check the scale invariance nature of rainfall data of the study area.
- Determine the design daily rainfall for different return periods.
- Develop IDF curves for different return periods from daily rainfall data using simple scaling method.

1.3.2. Research Questions

- Are there data quality problems in the daily and hourly rainfall data?
- What is the extent of missing daily and hourly rainfall data?
- Is the annual maximum rainfall series stationary, independent and homogenous?
- Which of the probability distribution functions best fit with annual maximum rainfall series?
- Is there scale invariance in the rainfall data records for Bahir Dar city?
- What are the design rainfalls for 2, 5, 10, 25, 50 and 100 years of return periods?
- How comparable are the regional IDF curves and IDF curves developed by the simple scaling method?

1.4. Scope and limitation of the research

1.4.1. Scope of the research

The scope of the research focused on the developing the intensity, duration and frequency relationship using simple scale method from the Bahir Dar city annual maximum daily, hourly and sub hourly rainfall data. From this rainfall data, it was also possible to determine the intensity, duration and frequency curves for the city. After the development of the intensity, duration and frequency curves using the simple scaling method, the research also analyzed the relationship between the regional IDF curves in Ethiopia and also the IDF curves developed for Bahir dar city from simple scaling method. This relationship helped to recommend the users the IDF curves developed by the simple scaling method and relative factor also developed for future use as well.

1.4.2. Limitation of the research

The limitation of the research was absence of annual maximum sub hourly and hourly rainfall for the required 40 years data request from National Metrology Agency in Addis Ababa, but it was only possible to consider 4 years automatically recorded reliable sub hourly rainfall data for Bahir Dar city. The absence of the sub hourly rainfall data for about 40 years like annual maximum daily rainfall data might probably provide slightly different rainfall data.

CHAPTER 2: LITRATURE REVIEW

2.1. Previous similar or related works or studies on IDF curves

Through analysis of historical rainfall data from 4 rainfall stations in Singapore, the study showed that the rainfall in Singapore displayed scale invariance property and simple scaling model was thus applicable to Singapore rainfall. The results of this study showed that the Singapore rainfall displayed scale invariance property in 2 scaling regimes-from 45-min to 24-hour and 15-min to 45min.(Chang and Hiong, 2013). The simple scaling theory was applied to the intensity-duration-frequency(IDF) characteristics of a short duration rainfall. This method allowed for estimation of the design values of rainfall of selected recurrence interval and duration shorter than a day by using only the daily data. The scaling behavior of rainfall intensities was examined, and the possibility of using simple scaling in Slovakia was verified(Bara et al., 2010). The IDF curves for short duration (hourly) were derived from 24-hour data. The simple scaling property verified by local data; then IDF relationships were deducted from daily rainfall which showed good results as compared to IDF curves obtained from at-site short duration rainfall data at Yodo catchment of Japan(Nhat et al.,2006). As short duration rainfall records were not available at the Uberaba gauging station in Brazil, the simple scale invariance properties were verified for sites with sub daily records and then transferred to the location of interest (Naghettini, 2012). The magnitude and frequency of extreme events such as high intensity rainfall, flashing flooding, severe droughts, etc. are expected to be altered in future as a consequence of this change. The scaling properties of extreme rainfall were examined to establish scaling relationship behavior of statistical moments over different durations. The results showed that a rainfall property in time did follow a simple scaling process(Afrin et al, 2015). The establishment of IDF curves for precipitation remained a powerful tool in the risk analysis of natural hazards. Indeed, the IDF-curves allowed for the estimation of the return period of an observed rainfall event or conversely of rainfall amount corresponding to a given return period for different aggregation times.There was a high need for IDF-curves in the tropical region of Central Africa but unfortunately the adequate long-term data sets were frequently not available. More physically based models for the IDF curves were proposed(Mohymont et al., 2004). IDF relationships or IDF curves established for stations with long record length has good accuracy and good representation of the reality than those stations with short record length, but this thesis used short length of record that was ten years of data since there were no scenario data better than these data length even if it fulfills the minimum requirement)(Merra,2011). In urban drainage systems, knowledge of short duration rainfall events can be considered as one of the most critical elements when their hydrological behavior wants to be investigated. The proposed method is based on the scale invariance theory whose concepts imply that statistical properties of extreme rainfall processes for different temporal sales are self-related by scale changing operator involving only the scale ratio (Nhat et al.,2008).

2.2. Rainfall data

2.2.1. Annual maximum daily and hourly rainfall data

In many hydraulic engineering applications such as those concerned with floods, the probability of occurrence of a particular extreme rainfall, e.g. a 24-hour maximum rainfall, will be of importance. Such information is obtained by the frequency analysis of the point rainfall data. The rainfall at a place is a random hydrologic process and the rainfall data at a place when arranged in chronological order constitute time series. One of the most commonly used data series is the annual series composed of annual values such as annual rainfall. If the extreme values of a specified event occurring in each years listed, it also constitutes an annual series. Thus for example, one may list the maximum 24-h maximum rainfall values. The probability of occurrence an event in this series is studied by frequency analysis of this annual data series (Subramanya,1984)

2.3. Tests on hydrological data

2.3.1. Test for Independence and Stationarity

Under this section, it is mandatory first to check the stationarity, homogeneity, independence and outliers of the daily rainfall data series. First, let us define what does it mean by stationarity, homogeneity, independence and outlier of the rainfall data.

- **Independence:** It implies that no observation in the data series has any influence on any following observations.
- **Homogeneity:** It means that all the elements of the data series originate from a single population.
- **Stationarity:** It means that, excluding random fluctuations, the data series is invariant with respect to time. Types of non-stationarity include trends, jumps and cycles. In flood analysis, jumps are generally due to an abrupt change in a basin or river system, such as the construction of dam. Trends may be caused by gradual changes in climatic conditions or land use, such as urbanization. Cycles may be associated with long term climatic conditions.
- **Outliers:** An outlier is an observation that deviates significantly from the bulk of the data, which may be due to errors in data collection, or recording, or due to natural causes.

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

$$R = \sum_{i=1}^{N-1} x_i x_{i+1} + x_1 x_N \qquad 2.1.1.1.$$

1

When the elements of the sample are independent, R follows a normal distribution with mean and variance 2 and 3 below.

$$\overline{R} = \frac{(s_1{}^2 - s_2\ )}{N-1} \qquad 2.1.1.2$$

$$Var(R) = \frac{S^2{}_2 - s_4}{N-1} - \overline{R}^2 + \frac{(s^4{}_1 - 4s_1{}^2 s_2 + 4s_1\ s_3 + s_2{}^2 - 2s_4}{(N-1)(N-2)} \qquad 2.1.1.3$$

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

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

2.3.2. Test for Homogeneity and Stationarity

In this test two samples of size *p* and *q* with $p \le q$ are compared. The combined data set of size $N=p+q$ is ranked in increasing order. The Mann-Whitney (M-W) test considers the quantities V and W in equations below.

$$V = R - \frac{(p(p+1))}{2} \qquad 2.2.2.1$$

$$W = pq - V \qquad 2.2.2.2$$

R is the sum of the ranks of the elements of the first sample(size p) in combined series(size N),and V and W are calculated from R, p, and q. V represents the number of times an item in sample 1 follows an item in sample 2 in the ranking. Similarly, W can be computed for sample 2 following sample 1.The M-W statistic U is defined by the smaller of V and W. When N>20 and p, q>3 and under the null hypothesis that the two samples came from the same population, U is approximately normally distributed with mean $\overline{U} = \frac{pq}{2}$ and variance var (U),

$$\mathrm{var}(U) = \left[\frac{pq}{N(N-1)}\right]\left[\frac{N^3 - N}{12} - \sum T\right] \qquad 2.2.2.3$$

Where $T = (J^3 - J)/12$ and J is the number of observations tied at a given rank. T is summed over all groups of tied observations in both samples of size *p* and *q*. The statistic $u = (U - \overline{U}) / \left[\mathrm{var}(U)\right]^{1/2}$ is used to test the hypothesis of homogeneity at significance level α by comparing it with the standard normal variant for that significance level (Rao and Hamed, 2000).

2.3.3. Test for Outliers

An outlier is an observation that deviates significantly from the bulk of the data, which may be due to errors in data collection, or recording, or due to natural causes. The presence of outliers in the data causes difficulties when fitting a distribution to the data. Low and high outliers are both possible and have different effects on the analysis. The Grubbs and Becktest(G-B) may be used to detect outliers. In the this test the quantities x_H and x_L are calculated by using the equations mentioned here below.

$$x_H = \exp(\bar{x} + k_N s) \qquad 2.2.3.1$$

$$x_L = \exp(\bar{x} - k_N s) \qquad 2.2.3.2$$

Where $\bar{x}$ and s are the mean and standard deviation of the natural logarithms of the sample, respectively, and k_N is the G-B statistic by Grubbs and Becks(1972).At the 10% significance level, the following approximately proposed by Pilon et al.(1985) is used, where N is the sample size.

$$k_N = -3.62201 + 6.28446N^{1/4} - 2.49835N^{1/2} + 0.491436N^{3/4} - 0.037911N \qquad 2.2.3.3$$

Sample values greater than x_H are considered to be high outliers, while those less than x_L are considered to be low outliers.

The other methods of determining outliers includes the US water resources council(1981) method, which has a very high threshold(McCormick and Rao,1995).The least median square(LMS) method is also the alternative method of testing outliers. The general equation of the least median square is of the form below.

$$Y_i = \beta_o + \beta_1 X_i + \varepsilon_i \qquad 2.2.3.4$$

In the above equation, Y_i is the response variable; β_0 and β_1 are the parameters to be estimated; X_i is independent variable; and ε_i is the derivation from predicted value. The estimated value from the above equation is given below.

$$\hat{Y}_i = b_o + b_1 X_i \qquad 2.2.3.5$$

where $\hat{Y}_i$ is the estimated response variable; b_0 and b_1 are the estimates of the parameters β_0 and β_1.The residual value, the difference between predicted and observed values, is then defined as below.

$$e_i = Y_i - \hat{Y}_i \qquad 2.2.3.6$$

In the LMS method, b_0 and b_1 in the above equation are determined such that the median value of the squared residual is minimized:

$$\text{Minimize}\left[med(e_i^2) \right] \qquad 2.2.3.7$$

A robust method is one which will not overemphasize any particular portion of the data range, such as the higher values of the dependent variable. The LMS method is a robust method. A scale estimate is used by the LMS method to define how well the data are fitted by the straight line. the initial scale estimate, s^0 is given by below.

$$s^0 = 1.486\left(1+\frac{5}{n-2}\right)\sqrt{MED(e_i^2)} \qquad 2.2.3.8$$

Each observation is then assigned a weight, corresponding to whether it is within a reasonable range of the initial scale estimate.

$$w_i = \begin{cases} 1 if \ |e_i| / s^0 \leq 2.5 \\ 0 otherwise \end{cases} \qquad 2.2.3.9$$

The final scale estimate used in the LMS method is then:

$$\sigma^* = \sqrt{\frac{\sum_{i=1}^{n} w_i e_i^2}{\sum_{i=1}^{n} w_i - 2}} \qquad 2.2.3.10$$

Outliers may be detected from plot of e_i vs. $\hat{Y}_i$,also known as a residual plot. The criterion to test whether an observation is an outlier is whether it has a residual value of greater than a multiple factor of the final scale estimate (Rao and Hamed,2000).

2.4. Data evaluation: Missing and In-filling of data

In the area of water resources planning and management, complete sets are required on many variable such as rainfall, stream flow, evapotranspiration and temperature. Unfortunately, records of hydrological processes are usually short and often missing observations. The existence of data gaps might be attributed to a number of factors such as interruption of measurements because of equipment failure, existence of extreme natural phenomena or of human –induced factors such war and civil unrest, mishandling of observed records by field personnel, or accidental loss of data files in the computer system (Elshorbagy et.al, 2009).

According to (Elshorbagy et.al., 2009), missing data can be categorized in to three groups:

a) Data trivial importance missing(e.g. a few sparsely distributed, not consecutive, missing observations in a long historical record).In this case, a simple in-filling method such as in-filling by suing series average or simple interpolation could be satisfactory. Peak or extreme values should not be encountered in the data gaps to be in-filled by the simple in-filling methods.
b) Fundamental data are missing (e.g. lengthy segments or many intermittent observations) where data patterns or structure cannot be recognized from the remaining record. In this case, any attempt at in-filling such a record may be unreliable and therefore the whole record should be dropped according to the current available state of knowledge and techniques.
c) Significant data are missing(e.g. a segment of consecutive observations).In the latter case, the missing data are considered important enough(quantitatively or qualitatively)

to deserve developing a technique that estimates them as accurately as possible. At the same time, the data gaps are too short to have significant damaging effect on the patterns and structure of the whole record.

2.5. Frequency analysis

The primary objective of frequency analysis is to relate the magnitude of extreme events to their frequency of concurrence through the use of probability distributions. Data observed over an extended period of time in a river system are analyzed in frequency analysis. The data are assumed to be independent and identically distributed. The flood data are considered to be stochastic and may be even be assumed to be space and time independent. Further, it is assumed that the floods have not been affected by natural or manmade changes in the hydrological regime in the system.

2.5.1. Return periods, Probability and plotting formulas

Flood peaks don"t occur with any fixed pattern in time or magnitude. Time intervals between floods vary. The definition of return period is the average of these inter-event times between flood events. Large floods naturally have large return periods and vice versa. The definition of the return period may not involve any reference to probability. However, a relationship between the probability of occurrence of a flood and its return period can be justified. A given flood q with return period T may be exceeded once in T years. Hence, the probability of exeedence is $P(Q_T > q) = 1/T$ .The cumulative probability of non-exceedence, $F(Q_T)$ is given below.

$$F(Q_T) = P(Q_T \le q) = 1 - P(Q_T > q) = 1 - \frac{1}{T} \qquad 2.4.1.1$$

The above equation, it is the basis for estimating the magnitude of flood, Q_T given its return period T .Substituting $F(Q_T) = 1 - 1/T$ in a unknown statistical distribution function, one can solve for the magnitude of Q_T given its return T .Often ,the data are plotted on probability paper to check whether they follow a particular distribution, to detect errors, and to check for outliers. Probability plots require an initial estimate of the probability of non-exceedence $F = (F(Q_T))$,which is called a "plotting position." Plotting positions are also used to estimate parameters by using the probability weighted moments (PWM) method. A plotting position formula by Hosking(1990).

$$F = \frac{i - 0.35}{N}, i = 1, 2, ..., N \qquad 2.4.1.2$$

where N is the sample size and i is the rank of the observations in ascending order. The formula in the above is believed to give acceptable results for some common three parameter distributions and is used in in the PWM method.

2.5.2. Hydrologic risk

A relation between the magnitude, design in years and the probability of not exceeding that magnitude in the design period was derived by Riggs. A method based on Poisson distribution, to compute the probability of occurrence or exceedence of a flood of a specified magnitude at least once in a given time period, was developed by Hall and Howell(1963).Yen also derived an expression for the risk of failure associated with a return period and the expected life of a project. The bias in computed flood risk was discussed by Hardison and Jennings who recommended that the accuracy of procedures used in flood frequency analysis be appraised in terms of standard errors.

2.5.3. Regionalization

The availability of data is an important aspect in frequency analysis. The estimation of probability of occurrence of extreme floods is an extrapolation based on limited data. Thus the larger the database, the more accurate the estimates will be. From statistical point of view, estimation from small samples may give unreasonable or physically unrealistic parameter estimates, especially for distributions with a large number of parameters(three or more).Large variations associated with small sample sizes cause the estimates to be unrealistic. In practice, however, data may be limited or in some cases may not be available for a site. In such case, regional analysis is most useful. Regional analysis is based on the concept of regional homogeneity which assumes that annual maximum flow populations at several sites in a region are similar in statistical characteristics and are not dependent on catchment size although this assumption may not be strictly valid, it is convenient and effective.

Regionalization serves two purposes. For the sites where data are not available ,the analysis is based on regional measured at a site, called at-site data, and regional data from a number of stations in a region provides sufficient information to enable a probability distribution to be used with greater reliability. This type of analysis represents a substitution of space for time where data from different locations in a region are used with greater reliability. This type of analysis represents a substitution of space for time where data from different locations in a region are used to compensate for short records at a single site. Many types of regionalization procedures are available. One of the simplest procedures which has been used for a long time is the index flood method. The key assumption in the index flood method is that the distribution of floods at different sites in a region is the same except for a scale or index flood parameter, which reflects rainfall and runoff characteristics of each region. The index flood may be the mean flood, although any location parameter of frequency distribution may be used. In this case, regional quartile estimates $\hat{Q}_T$ at a given site for a given return period T can be obtained as below, where q_T is the quintile estimate from the regional distribution for the given return period, and μ_i is the mean flow at the site.

$$\hat{Q}_T = \mu_i q_T \qquad 2.4.3.1$$

The regional distribution parameters are obtained by using the regional weighted average of dimensionless rescaled data $q_{ij} = Q_{ij} / \hat{\mu}_i$ .Another method of obtaining the regional distribution parameters is station year approach where all the data are pooled, after dividing them by the mean μ_i at each site, and are treated as a single sample. The joint use of at-site and regional data is advisable, provided that a reasonably homogenous flood region can be identified. The data at a site may be used when the record at a station is exceptionally long, or when regional data are not available, or when a region is heterogeneous (Rao and Hamed, 2000).

2.6. Extreme maximum distributions and formulae

2.6.1. Normal distribution

The normal distribution is used in frequency analysis for fitting empirical distributions to hydrological, and in simulation of data. As many statistical parameters are approximately normally distributed, the normal distribution is often used for statistical inferences.

The probability density function of a normally distributed variable x is given by:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} \qquad 2.5.1.1$$

Where μ and σ are the parameters of the distribution. The variables x can take any value in the range (-∞,∞).The standard normal variant u is a normal variable with a mean equal to zero and standard deviation equal to one (Rao and Hamed, 2000).

2.6.2. Two-Parameter Lognormal (LN(2)) Distribution

The probability density function of a logarithmic normally distributed variable x with two parameters(LN(2)) is given by:

$$f(x) = \frac{1}{x\alpha_y\sqrt{2\pi}} \exp\left\{ \frac{\left[-\log x - u_y\right]^2}{2\sigma_y{}^2} \right\} \qquad 2.5.2.1$$

Where μ_y and σ_y are the mean and standard deviation of natural logarithm of x. If the variable log(x) is standardized as in Eq.2.5.1.3,

$$u = \frac{\log(x) - \mu_y}{\sigma_y} \qquad 2.5.2.2$$

The standard normal variant u is obtained with Eq.2 (Rao and Hamed, 2000).

.

2.6.3. Three Parameter Lognormal (LN(3)) Distribution

The three-parameter lognormal distribution is similar to the LN(2) distribution except that *x* is shifted by an amount *(a)* which represents a lower bound. The normally distributed variable becomes log(x-a) with the pdf,

$$f(x) = \frac{1}{(x-a)\sigma_y\sqrt{2\pi}} \exp\left\{ -\frac{1}{2\sigma_y^{\,2}} \left[\log(x-a) - \mu_y \right]^2 \right\} \qquad 2.5.3.1.$$

where μ_y and $\sigma_y^{\,2}$ are the location and scale parameters, which correspond to the mean and variance of the logarithm of the shifted variable $(x-a)$ .The standardized variable u is obtained as in Eq.2.5.1.5.

$$u = \frac{\log(x-a) - \mu_y}{\sigma_y} \qquad 2.5.3.2$$

Sangal and Biswas (1970) suggested a procedure to estimate the parameters of the LN(3) distribution in which only the mean, median and standard deviation of the data are used. The mathematical properties of the LN(3) distribution were discussed by Burges et al(1975).They also compared two methods of estimation of the third parameter „a‟ of the LN(3) distribution (Rao and Hamed, 2000).

Singh and Singh(1988) used the maximum entropy method to estimate the parameters of the LN(3) distribution.

2.6.4. Exponential distribution

The exponential is a special case of the Gamma family of distributions which include the Pearson(3),one and two parameter Gamma,log-Pearson and the generalized Gamma distributions (Rao and Hamed, 2000).

The pdf of the Pearson(3) distribution is given by:

$$f(x) = \frac{1}{\alpha^{\beta}.\Gamma(\beta)} (x-\varepsilon)^{\beta-1} e^{\frac{(x-\beta)}{\alpha}} \qquad 2.5.4.1$$

where $\Gamma(.)$ is the gamma function defined by Eq 2.5.1.7

$$\Gamma(y+1) = \int_0^{\infty} t^y e^{-t} dt, y+1 > 0 \qquad 2.5.4.2$$

with the following properties

$$\Gamma(y+1) = y\Gamma(y), y > 0 \qquad 2.5.4.3$$

$$\Gamma(y) = \Gamma(y+1)/y, y < 1 \qquad 2.5.4.4$$

$$\Gamma(n) = (n-1)!, \qquad 2.5.4.5$$

n is a positive integer

2.6.5. Two-Parameter Gamma(G(2)) Distribution

The two-parameter gamma(G(2)) distribution is a special case of the Pearson(3) distribution in which the parameter εin Eq.3.4.4.1 is equal to zero. The pdf of the G(2) distribution is thus

$$f(x) = \frac{1}{\alpha^{\beta}\Gamma(\beta)} x^{\beta-1} e^{-(x/a)} \qquad 2.5.5.1$$

The variable x in this case has a lower bound of zero instead of ε,0< x <∞.the distribution function of x cannot be obtained in closed form (Rao and Hamed, 2000).

2.6.6. Pearson(3) Distribution

The probability density function of the Pearson(3) distribution is given by Eq.2.5.6.2.

$$f(x) = \frac{1}{\alpha\Gamma(\beta)} \left(\frac{x-\gamma}{\alpha}\right)^{\beta-1} e^{-\left(\frac{x-\gamma}{\alpha}\right)} \qquad 2.5.6.1$$

The variable x in a Pearson(3) distribution can take values in the range $\gamma < x < \infty$ .Generally, α can be positive or negative, but for negative values of α the distribution becomes upper bounded and is therefore not suitable for analyzing maximum events (Rao and Hamed, 2000)..

2.6.7. Log-Pearson(3) Distribution

If the variable $\log x$ is assumed to have a Pearson(3) distribution when the distribution of the variable x is a log-Pearson(3)(LP(3)) distribution. The probability density function of a LP(3) distributed random variable is given by

$$f(x) = \frac{1}{\alpha x\Gamma(\beta)} \left[\frac{\log(x)-\gamma}{\alpha}\right]^{\beta-1} e^{-\left\{\frac{\log(x)-\gamma}{\alpha}\right\}} \qquad 2.5.7.1$$

The pdf of the log-Pearson(3) distribution may take many different shapes(Bobee and Ashkar,1991).

2.6.8. Generalized Extreme Value Distribution(GEV) Distribution

The probability density function of the GEV distribution is of form

$$f(x) = \frac{1}{\alpha}\left[1-k\left(\frac{x-u}{\alpha}\right)\right]^{1/k-1} e^{-\left[1-k\left(\frac{x-u}{\alpha}\right)\right]^{1/k}} \qquad 2.5.8.1$$

The range of the variable x depends on the sign of the parameter (Rao and Hamed, 2000).

2.6.9. The Extreme Value Type I EV1(2) Distribution

The probability density function of the EV1(2) distribution is given by Eq.2.5.9.1.

$$f(x)=\frac{1}{\alpha}\exp\left[-\left(\frac{x-\beta}{\alpha}\right)-e^{-\left(\frac{x-\beta}{\alpha}\right)}\right] \quad 2.5.9.1$$

The variable x take values in the range $-\infty< x <\infty$.

2.6.10. Parameter estimation for the distributions

2.6.10.1. Easy fit software application

Using the easy fit software, we can analyze the maximum daily annual rainfall data. Hence, by this software, it is quite possible to determine and select the best fit distribution functions for the rainfall data for Bahir dar city.

2.6.11. Goodness of tests and formulae

The goodness of fit test measures the compatibility of random sample with the theoretical probability distribution. The goodness of fit tests is applied for testing the following null hypothesis:

H_0 :the maximum daily rainfall data follow the specified distribution.

H_A :the maximum daily rainfall does not follow the specified distribution.

The following goodness of fit tests viz. Kolmogorov-Simirnov test and Anderson-darling test will be used along with the chi-square test at α level of significance for the selection of the best fit probability distribution.

a) Kolmogorov-Smirnov Test

The Kolmogorov-Smirnov statistic(D) is defined as the largest difference between the theoretical and empirical cumulative distribution function(ECDF):

$$D=\max\left(F\left(x_i\right)-\frac{i-1}{n},\frac{i}{n}-F\left(x_i\right)\right) \quad 2.5.11.1$$

$$\left(1\le i\le n\right)$$

where X_i is the random sample, i=1,2,……..n

$$CDF=F_n(x)=\frac{1}{n}.\left[Number \quad of \quad observations \quad \le \quad x\right] \quad 2.5.11.2$$

This test is used to decide if a sample comes from hypothesized continuous distribution.

b) Anderson-Darling Test

The Anderson-Darling Statistic(A^2) is defined as

$$A^2 = -n - \frac{1}{n}\sum_{i=1}^{n}(2i-1).\left[InF(X_i) + In(1-F(X_{n-i+1}))\right] \qquad 2.5.11.3$$

It is a test to compare the fit of an observed cumulative distribution to an expected cumulative distribution function. This test gives more weight to the tails then Kolmogorov-Smirnov test.

c) Chi-Square Test

The Chi-Square statistic is defined as

$$\chi^2 = \sum_{i=1}^{k}\frac{(O_i - E_i)^2}{E_i} \qquad 2.5.11.4$$

where O_i is the observed frequency, E_i is the expected frequency and „I‟ is the number of observations(1,2,……..k),which is calculated by $E_i = F(x_2) - F(x_1)$ and F=the CDF of the probability distribution being tested.

The observed number of observation(k) in interval „i‟ is computed from equation given below.

$$k = 1 + \log_2 n \qquad 2.5.11.5$$

n=sample size.

this test is continuous sample data only and is used to determine if a sample comes from a population with a specific distribution (Rao and Hamed, 2000).

2.7. Intensity Duration Curves(IDF) curves

2.7.1. History of IDF curves

The rainfall intensity-duration-frequency relationship is one of the most widely used methods in urban drainage design and flood plan management. The establishment of such relationships goes back to as early as 1928 by Meyer. After Meyer had developed a few, in 1931 Sherman derived applicable general intensity duration formula to other localities, and in 1932 Bernard made available for localities within limits of the study, rainfall intensity formulas for frequencies of 5,10,15,25,50,100 years, applicable to rainfall durations of 120 to 6000min (Abubakari, 2013).

In 1969, Bell developed IDF relationship using a formula which enabled him to compute the depth-duration ratio for certain areas of U.S.S.R. In February 1972, J.B Danquah developed IDF curves for town and cities in Ghana while 1982 Oyebande established IDF curves for Nigeria. In 1983, Chen developed a simple method to derive a generalized rainfall intensity duration frequency formula for any location in the United States (Sulemana Abubakari,2013).

In 1990‟s,some mathematically consistent approaches for IDF development had been proposed. In 1996, Burlando and Rosso proposed the mathematical framework to model extreme storm probabilities from the scaling properties of observed data of station precipitation, and the simple scaling and multiple scaling conjectures was thus introduced to describe the temporal structure of extreme storm rainfall (Abubakari, 2013).

During the 1994-1998,Koutsoyiannis proposed a new approach to the formulation and construction of the intensity-duration-frequency curves using data from both recording and non-recording stations. More especially, the approach discussed a general rigorous formula for the Intensity-Duration-Frequency relationship whose specific forms had been explicitly derived from the underlying probability distribution function of maximum intensities. He also proposed two methods for a reliable parameter estimation of the IDF relationship. Finally, it discussed a framework for the regionalization of IDF relationships by also incorporating data from non-recording stations (Abubakari, 2013).

More recently, in 2001,Garcia-Bartual and Schneider used statistical distributions and found thee Gumbel Extreme Value(GEV) distribution fitted to data well. in 2004,Yu with others, developed regional rainfall intensity duration frequency relations for non-recording sites based on scaling theory, which uses the hypothesis of piecewise simple scaling combined with the Gumbel distribution. In 2006, Di Baldassare analyzed the capability of seven different depth duration –frequency curves characterized by two to three parameters to provide an estimate of the design of rainfall for storm durations shorter than 1 hour.In 2007, Karahan and others estimated parameters of a mathematical framework for IDF relationship presented by Koutsoyiannis using genetic algorithm approach (Abubakari, 2013).

In 2009, Simonovic made the updating of rainfall intensity duration frequency curves for the City of London under changing climate. While in 2010,Vyver did the construction of IDF curves for the precipitation at Lubumbashi, Congo, under the hypothesis of inadequate data. In 2011,Lumbroso worked on the challenges of developing rainfall Intensity-duration-frequency curves and national flood hazard maps for the Caribbean. In 2012, Naghettini studied the application of scale invariance properties of rainfall for estimating the intensity-duration-frequency relations at Uberaba, in South-central Brasil. In 2013,Logah studied on developing Short Duration Rainfall Intensity Frequency Curves for Accra Ghana. In 2015, Islam and Rahman studied the development of IDF curve for Dhaka City Based on Scaling Theory under Future Precipitation Variability Due to Climate Change.

2.7.2. Application of IDF curves

The intensity-duration-Frequency curve(IDF) relationship of heavy rainfalls is certainly among the hydrologic tools utilized by engineers to design storm sewers, culverts, retention/detention basins, and other structures of storm water management systems. An at-site IDF relationship is statistical summary of rainfall events, estimated on the basis of records of intensities abstracted from rainfall depths of sub daily durations, observed at particular recording rainfall gauging station. At some particular site of interest, there might be one or more recording rainfall gauging stations operating for a time period sufficiently long to yield a reliable estimate of the at-site IDF relationship. In other location, however, these recording stations may either not exist or have too few records to allow a reliable estimation of IDF relationships (Naghettini, 2012).

The design of any infrastructure requires an understanding of the desired function of the structure and the physical environment in which it must perform this function. Thus, in the case of storm water management ,the dimensions of various components of the infrastructure system are based on the return period of heavy rainfall events. This information is often expressed as IDF curves obtained from a statistical study of extreme events. Depending on the application purpose they may be constructed using different time steps from instantaneous maximum daily intensity to annual, monthly and weekly intensities (Abubakari, 2013).

The major reasons for increased demand for rainfall IDF information can be summarized as follows: a)As the spatial heterogeneity of extreme rainfall patterns becomes better understood and documented, a stronger case is made for the value of "locally relevant" IDF information. b) As urban areas expand, making watersheds generally less permeable to rainfall and runoff, many older water systems fall increasingly in to deficit, failing to deliver the services for which they were designed. Understanding the full magnitude of this deficit requires information on the maximum inputs (extreme rainfall events) with which drainage works must contend. c) Climate change will likely result in an increase in the intensity and frequency of extreme precipitation events in most regions in the future. As a result, IDF values will optimally need to be updated more frequently than in the past and climate change scenarios might eventually be drawn upon in order to inform IDF calculations (Abubakari, 2013).

As per (Abubakari,2013), the properties and major uses of IDF curves were shown below respectively.

a) In logarithmic system of coordinates, the IDF relationships are almost parallel decreasing lines. These curves cannot cross each other.
b) For any duration of rainfall, one can establish the intensity of rainfall solong as the frequency of occurrence is given.
c) For any given return period, high rainfall intensities are recorded in short duration. In other words, the most intense rain are short duration.

The uses are also includes the following major points.

- Design of hydraulic structures(such as culverts and bridges), roads and urban drainage systems.
- Land use planning and soil conservation studies.
- Management of municipal infrastructure including sewers, storm water management ponds and street curb.

- Design of safe and economical structures for the control, storage and routing of storm water and surface drainage.
- Risk assessment of dams and bridges,
- Design of roof and storm water drainage systems.
- Flood plain management,
- Soil conservation studies,
- Water resources management,
- The curves can also be used as input to rainfall-runoff models that simulate floods for bridge and spillway design, and
- The IDF relationship are used in the rational method to determine the average rainfall intensity for a selected time of concentration.

2.7.3. Methods for IDF curves development

The rainfall Intensity-Duration-Frequency(IDF) relationship is one of the most important tools in water resource engineering to assess the risk and vulnerability of water resource structures as well as for planning ,design and operation (Rashid et. al, 2012).

2.7.3.1. Empirical Reduction Formula

Daily rainfall data were from the study for the available data in years. From this available data, the maximum values were extracted from each year and were converted in to shorter duration(1, 2, 3, 6 and 12-hrs) values using the reduction formula suggested by the Indian Meteorological Department (Lamia Abdul Jaleel, and Maha Atta Farwan, 2013).This will be done using Indian Meteorological Department(IMD) empirical reduction formula and it is shown below.

$$P_t = P_{24}\left(t/24\right)^{1/3} \qquad 2.6.3.1.1$$

P_t -the required precipitation depth in mm for the duration of t-hour

P_{24} -daily precipitation(mm). t -the time duration(in hours) for the required precipitation depth.

2.7.3.2. IDF empirical equations

IDF empirical equations are the equations that estimates the maximum rainfall intensity for different duration and return period. IDF is a mathematical relationship between the rainfall intensity *i,* the duration t_d ,and the return period T.

$$i = \frac{a}{b+t_d} \text{ and } i = x*(t_d)^{-y} \qquad 2.6.3.2.1$$

where, i is the rainfall intensity in mm/hr , t_d is the rainfall duration in min. a, b, x and y are the fitting parameter. These empirical equations indicate that for a given return period the rainfall intensity decrease with increase in rainfall duration. Least square method can help to find the parameters for the rainfall IDF empirical formula (Rashid et. al., 2012).

The intensity duration function formulas are the empirical equations representing a relationship among maximum rainfall intensity (as dependent variable) and other parameters of interest such as rainfall duration and frequency (as independent variables).Four basic forms of equations or methods used to describe the rainfall intensity duration relationship are summarized as below:

a. Talbot Equation: $\mathrm{i} = \frac{a}{d+b}$ 2.6.3.2.2

b. Bernard Equation : $i = \frac{a}{d^e}$ 2.6.3.2.3

c. Kimijima Equation: $\mathrm{i} = \frac{a}{\mathrm{d}^e + b}$ 2.6.3.2.4

d. Sherman Equation: $i = \frac{a}{(d+b)^e}$ 2.6.3.2.5

Where i is the rainfall intensity(mm/hr);d is the duration(minutes);a, b and e are the constant parameters related to the metrological conditions. These empirical equations show rainfall intensity decreases with rainfall duration for a given return period. The least square can be applied to determine the parameters of the four empirical IDF equations that are used to represent intensity-duration relationships (Nhat et. al.,2006).

The Ontario drainage management manual recommends fitting the IDF data to the following parameter function:

$$i = \frac{A}{(t_d + B)^C} \quad 2.6.3.2.6$$

where i is the rainfall intensity(mm/hr), t_d the rainfall duration(min),and A, B, and C are coefficients. After selecting reasonable parameter B, method of least squares is used to estimate values of A and B. The calculation is repeated for a number of different values of B in order to achieve the closest possible fit of the data (Simonovic and Peck, 2009).

2.7.4. Scaling method

2.7.4.1. Simple scaling characteristics of rainfall

The daily precipitation data are by far the most accessible and abundant source of rainfall information, it is appealing to develop methods to derive the IDF characteristics of short-duration events from daily rainfall statistics. The early attempts to derive short-duration rainfall intensities from daily data made use of empirical proportionality factors, which are supposed to be valid at a specific location or over a geographic region. More recently, research has focused on the mathematical representation of rainfall fields both time and space, including the development of scaling invariance models to derive short-duration rainfall intensity-frequency relations from the daily data (Naghettini, 2012).

Scaling invariance occurs when the connections among the statistical descriptors of a given phenomenon at different scales are constant and defined by a scale factor. The statistical descriptors can be scaled either by a single (simple scaling) or by a more complex function of the scale(multi scaling).As rainfall is concerned, scaling its statistical description in time/space is related to the study of its fractal properties or, in other terms, the way in which rainfall organizes itself in self-affine cell clusters in time/space (Naghettini, 2012).

A scale invariant concept is explored for disaggregation (or downscaling) of rainfall intensity from low to high resolution and is applied to the derivation of scaling of IDF curves (Afrin et. al., 2015).

The scaling or scale-invariant models enable us to transform data from one temporal or spatial model to another one, and thus, help to overcome the difficulty of inadequate data. A natural process fulfills the simple scaling property if the underlying probability distribution of some physical measurements at one scale is identical to the distribution at another scale, multiplied by a factor that is a power function of the ratio of the two scales. The basic theoretical development of scaling has been investigated by many authors.
Let $X(t)$ and $X(\lambda t)$ denote measurements at two distinct time or spatial scales t and λt ,respectively. Definition of scaling property of the probability of the $X(t)$ is

$$X(t) = \lambda^{-H} X(\lambda t) \qquad 2.6.4.1.1$$

where λ denotes a scale factor and H is a scale exponent which varies with location. Gupta and Waymire introduced the notions of strict and wide sense simple scaling. The strict sense simple scaling above equation implies that $X(t)^q$ and $(\lambda^{-H} X(\lambda t))^q$ have the same probability distribution. The wide sense simple scaling is expressed as they have the same moments, i.e.

$$E\left[X(t)^q\right] = \lambda^{-Hq} E\left[X(\lambda t)^q\right] \qquad 2..6.4.1.2$$

The scaling exponent (Hq)can be estimated from the slope of linear regression relationship between the log transformed values of moment, $\log E\left[X(\lambda t)^q\right]$ and the scale parameters log λ for various order of moment q .This is definition of a “wide sense” simple scaling. A “wide sense” simple scaling with t=1 is given by:

$$E\left[X(1)^q\right] = \lambda^{-Hq} E\left[X(\lambda)^q\right] \qquad 2.6.4.1.3$$

If the scaling exponent H is not constant and changes probabilistically, equation 3 described as:

$$E\left[X(1)^q\right] = \lambda^{-K(q)} E\left[X(\lambda)^q\right] \qquad 2.6.4.1.4$$

where $K(q)$ is a function of the moment order. The procedure to test the suitability of scale invariant model to shown the figure below. The moments $E\left[X(\lambda t)^q\right]$ are plotted on the logarithmic chart versus the scale λ for different moments" order q .The slope $K(q)$ is plotted on the linear chart versus the moment order q .If the resulting graph is a straight line, the multi-scaling approach has to be considered. Let the random variable $I(d)$,the maximum annual value of local rainfall intensity over a duration d .It is defined as:

$$I(d) = \max_{0\le t\le 1year}\left[\frac{1}{d}\int_{t-d/2}^{t+d/2} X(\xi)d\xi\right] \quad 2.6.4.1.5$$

where $X(\xi)$ is a time continuous stochastic process representing rainfall intensity and d is duration. It is supposed $I(d)$ represents the annual maximum rainfall intensity of duration d ,defined by the maximum value of moving average of width d of continuous rainfall process.

Here, some concepts are introduced about scaling of the probability distribution of random functions. A generic random function $I(d)$ is denoted by simple scaling properties if it obeys the following.

$$I(d) = \left(\frac{D}{d}\right)^{-H_d} I(D) \quad 2.6.4.1.6$$

D is an aggregated time duration, i.e. 2,3,…,24hours.

Defining the scaling ratio is $\lambda_d = \frac{D}{d}$.

$$I(d) = \lambda^{-H_d} I(\lambda_d d) \quad 2.6.4.1.7$$

Equation 7 is rewritten in terms of the moments of order q about the origin, denoted by $E\left(I(d)^q\right)$.

The resulting expression is:

$$E\left[I(d)^q\right] = \lambda_d^{-H_d q} E\left[I(\lambda_d d)^q\right] \quad 2.6.4.1.8$$

If one assumes the wide sense simple scaling exists, the distribution of IDF for short-duration rainfall intensity can be derived from daily rainfall (Afrin et. al., 2015).

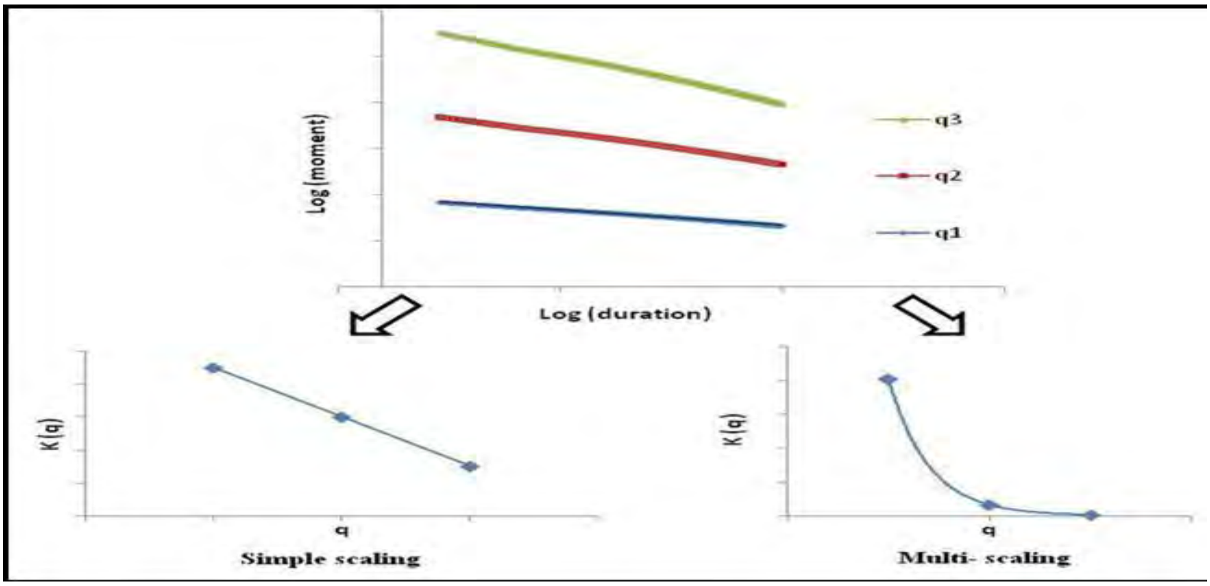


Figure 1. Simple and multiple scalings

2.7.4.2. Derivation of IDF for Short Duration

All forms of the generalized relationships assume that rainfall depth or intensity is inversely related to the duration of a storm raised to a power, or scale factor (Afrin et. al., 2015).

Scale invariance occurs when the connections among the statistical descriptors of a given phenomenon at different scales are constant and defined by scale factor. The statistical descriptors can be scaled either by a single factor(simple scaling or by a more complex function of the scale(multi scaling) (Naghettini, 2012).

According to Koutsoyannis, IDF relationships are particular cases of the following general formula.

$$i = \frac{w}{(d^{v} + \theta)^{\eta}} \qquad 2.6.4.2.1$$

Where i denotes the rainfall intensity of duration d, and w, v, θ ,and η are non-negative coefficients. Koutsoyannis also proposed a numerical exercise, in which they show that the errors resulting from imposing v=1 in equation (1) are much smaller than the typical parameter and quintile estimation errors from limited size samples of rainfall data. Hence, considering $v \neq 1$ as model over-parameterization, Koutsoyannis prescribed that, for a given period, the general expression for IDF relationships should be written as;

$$i = \frac{w}{(d + \theta)^{\eta}} \qquad 2.6.4.2.2$$

Rigorously, the coefficients w, θ ,and η depend on the return period. However, because the IDF curves for different return periods cannot intersect each other, this dependence cannot be arbitrary. Actually, this restriction imposes bounds for the range of variation of the parameters w, θ ,and η .For instance, if (w1, $\theta 1$,and $\eta 1$) and (w2, $\theta 2$,and $\eta 2$) denote the parameter sets for the return period T1 and T2 respectively, with T2<T1,Koutsoyannis et al.(1998) suggests the following restrictions on the range of variation of parameters.

$$\theta 1 = \theta 2 = \theta \geq 0 \qquad 2.6.4.2.3$$

$$0 < \eta_1 = \eta_2 = \eta < 1 \qquad 2.6.4.2.4$$

$$w_1 > w_1 > 0 \qquad 2.6.4.2.5$$

In this set of restrictions, it is worth to note the only parameter that can consistently increase with increasing return periods is w ,which results in substantial simplification of equation(2).In fact, these arguments justify the formulation of the following general model for IDF.

$$i_{d,T} = \frac{a(T)}{b(d)} \qquad 2.6.4.2.6$$

Which exhibits the advantage of expressing separable dependence relations between i and T, and between i and d. In equation(4), b(d)= $(d + \theta)^{\eta}$ with θ>0and 0<η <1,whereas a(T) is completely defined by the probability distribution function of the maximum rainfall intensities (Nhat et. al.,2006).

2.7.5. Use of IDF curve in Ethiopia or the regional IDF curves

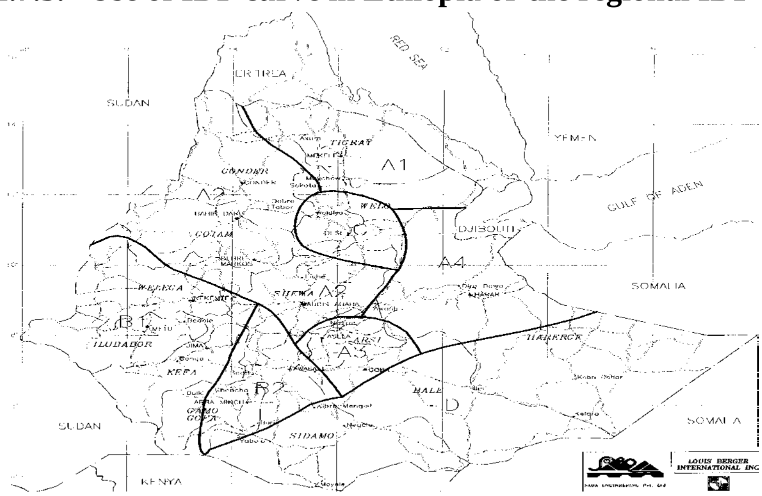


Figure 2: Categorization of regional IDF curves coverage

ERA(2002) drainage manual provides six region or regime areas as shown in the fig below which was established based on rainfall stations shown below as well. Also using the statistical analysis, rainfall intensity duration curves have been developed for commonly used design frequencies. These basic information will be used in estimating peak flood for small and large watersheds.

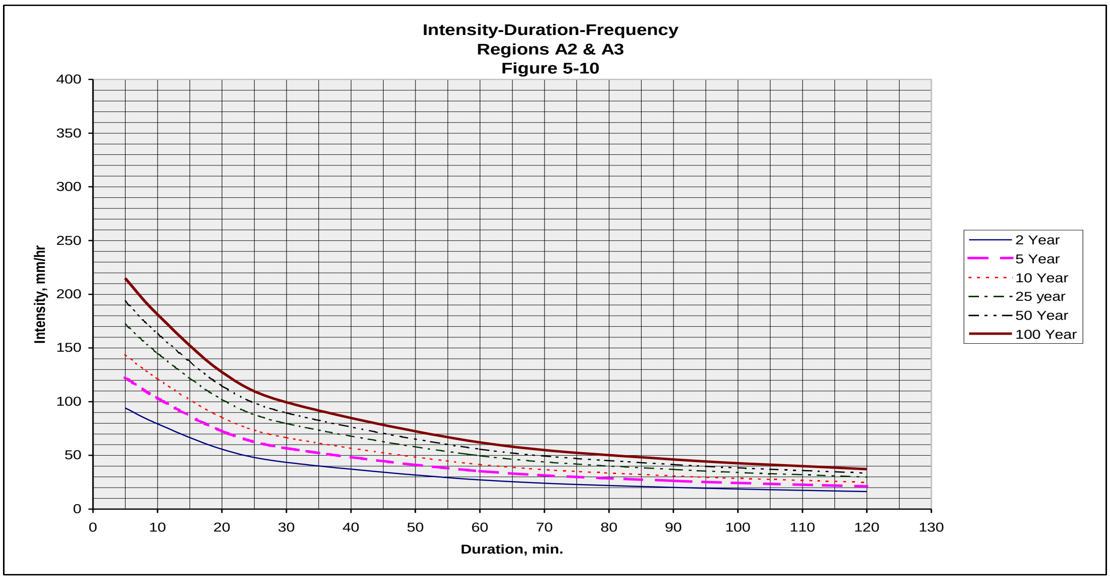


Figure 3:Intensity duration curves for regions A2 and A3

Table 1: Meteorological stations with relatively long record period

Meteorological stations with relatively long record period					
Meteorological Region	Station	Years of record	Meteorological Region	Station	Years of record
A1	Axum	18	B	Bedele	19
	Mekele	35		Gore	45
	Maychew	24		Nekemte	27
A2	Gondar	40		Jima	45
	Debre tabor	22		Arba minch	11
	Bahir dar	35		Sodo	28
	Debre markos	44		Awasa	26
	Fitche	25	C	Kombolcha	46
	Addis Ababa	33		Woldiya	23
A3	Nazareth	40		Sirinka	17
	Kulumsa	31	D1	Gode	29*
	Robe/Bale	19		Kebridihar	38
A4	Metehara	28	D2	Kibre mengist	24
	Diredawa	46		Negele	45
	Mieso	35		Moyale	18
*max 24 hour rainfall not given				Yabelo	34

CHAPTER 3: MATERIALS AND METHODS

3.1. Description of the study area

Bahir Dar city is located on the southern shore of lake Tana at latitude of $11^0 35$"N and longitude of $37^0 23$"E in Ethiopia and it is the capital city of Amhara region. It is rapidly growing and expanding city. According to Ethiopia Meteorology Agency, the annual mean precipitation of Bahir Dar city is around 900mm,where 54% of the rain falls in July and August only and 3% falls during dry months, the rest falls in the remaining months. The urban storm runoff and domestic waste receiving water body, Lake Tana, with surface area of 3150km^2 is the largest fresh water body in the country and the source of Blue Nile.

The climate of the study area varies from humid to semi -arid. Most precipitation occurs in the wet season(June to September) and the remaining precipitation occurs in the dry season(October to February) and in the mild season(March to May).The average annual daily maximum and minimum temperature of Bahir dar station(1961-2007) is 26.7^0C and 11.7^0C respectively .The mean annual relative humidity based on the Bahir dar station(1997-2007) data was 58%.The seasonal variation of temperature is between 3^0C to 6^0C from the warmest month and the coolest month. In summer, peak temperature is reduced because of rainfall and clouds while the highest temperature normally is expected in (April and May).The range of elevation within the net basin of lake Tana is from 1784 to 4109m and it has the major impact both on the climate and human activity. On average, the temperature falls by 5.8^0C for every 1000 meters increase in elevation(Gebremariame, 2009).

3.2. Rainfall data and its sources

- **Daily rainfall data**

The daily rainfall data for Bahir dar city was collected from the National Meteorological Agency in Addis Ababa, Ethiopia. The daily data was covering 41 years of data collected in the city. The primary source of the daily and sub daily rainfall data in order to undertake the this study was the National Meteorological Station found in the Ethiopian capital, Addis Ababa.From the daily rainfall data for Bahir Dar city, it was possible to develop or extract the annual maximum daily rainfall for the 41 years of data. This was the basic platform in order to develop the IDF relationship along with the sub daily rainfall data.

- **Sub hourly and hourly rainfall data**

The sub hourly and hourly rainfall data was also collected from the National Metrological Agency in Addis Ababa, Ethiopia. In addition, the automatic sub daily rainfall data recording was the backbone of this study without which have no any success to the study. The sub hourly and hourly rainfall data automatically recorded in 15minutes intervals was only possible to collect from 2012-2015, for about 4 years long data. The automatic recording was mainly based on daily recording of 15minutes intervals. In the data generation from this complex data base was very cumbersome. However, with careful analysis and extraction, it was possible to generate the annual maximum sub hourly rainfall data. Without these data, it would have been a failure in order to reach to this cornerstone findings of the IDF curves as well as relationship developments. Therefore, the annual maximum hourly rainfall data was generated for the 15minutes , 30mintes , 45minutes , 60minutes , 120mintes, 180mintes, and 24 hour durations.

- **Rainfall Data Evaluation**

The data evaluation was mainly focusing on the identifying the missing data and filling the missing rainfall data especially for the annual maximum daily data. The average method of calculating the average of every months over 41 years(1975-2015) and filling the missing data over the specified period of time. The study was interested in taking or selecting the years with no data missing value in the aspect of using the annual maximum sub daily or hourly rainfall data for the simple scaling method of developing the IDF relationship along with developing the IDF curves.

A. Rainfall Data Missing

In the course of the 41 years of data recording from the national metrological station, it was noted that there was 3 months(March-May) data missing in 1991,8 months(May-December) in 2005,and 1 month(May) in 2014.In general, it was noted that over the course of 41 years period of data collection, there were 3 times data missing for the month of May and other months were partly experiencing only a singular time of data missing. This was clearly in chapter 4 in order to clarify the discussion mentioned here.

B. Missing Rainfall Data Filling

The missing data filling technique was mainly based on taking the average value of a month over the period of 41 years and fill to make sure that all months have got their own maximum monthly rainfall data from the annual maximum daily data. And it was really practical that to do this way as it was noted that some years got their annual maximum rainfall data specifically from this average method of maximum monthly rainfall data calculation.

3.3. Tests for Hydrological Data

3.3.1. Test for Homogeneity and Stationarity

In this test two samples of size p and q with $p \le q$ are compared. The combined data set of size $N=p+q$ is ranked in increasing order. The Mann-Whitney (M-W) test considers the quantities V and W in equations below.

$$V = R - \frac{(p(p+1))}{2} \qquad 3.3.1.1$$

$$W = pq - V \qquad 3.3.1.2$$

R is the sum of the ranks of the elements of the first sample(size p) in combined series(size N),and V and W are calculated from R, p, and q. V represents the number of times an item in sample 1 follows an item in sample 2 in the ranking. Similarly, W can be computed for sample 2 following sample 1.The M-W statistic U is defined by the smaller of V and W. When N>20 and p, q>3 and under the null hypothesis that the two samples came from the same population, U is approximately normally distributed with mean $\overline{U} = \frac{pq}{2}$ and variance var (U),

$$\mathrm{var}(U) = \left[\frac{pq}{N(N-1)}\right]\left[\frac{N^3 - N}{12} - \sum T\right] \quad 3.3.1.3$$

Where $T = (J^3 - J)/12$ and J is the number of observations tied at a given rank. T is summed over all groups of tied observations in both samples of size p and q. The statistic $u = (U - \overline{U})/\left[\mathrm{var}(U)\right]^{1/2}$ is used to test the hypothesis of homogeneity at significance level α by comparing it with the standard normal variant for that significance level (Rao and Hamed, 2000).

3.3.2. Test for Independence and Stationarity

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

$$R = \sum_{i=1}^{N-1} x_i x_{i+1} + x_1 x_N \quad 3.3.2.1$$

1

When the elements of the sample are independent, R follows a normal distribution with mean and variance 2 and 3 below.

$$\overline{R} = \frac{(s_1^{\ 2} - s_2\ )}{N-1} \quad 3.3.2.2$$

$$Var(R) = \frac{S^2{}_2 - s_4}{N-1} - \overline{R}^2 + \frac{(s^4{}_1 - 4s_1^{\ 2}s_2 + 4s_1\ s_3 + s_2^{\ 2} - 2s_4}{(N-1)(N-2)} \quad 3.3.2.3$$

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

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

3.3.3. Test for Outliers

An outlier is an observation that deviates significantly from the bulk of the data, which may be due to errors in data collection, or recording, or due to natural causes. The presence of outliers in the data causes difficulties when fitting a distribution to the data. Low and high outliers are both possible and have different effects on the analysis. The Grubbs and Becktest(G-B) may be used to detect outliers. In the this test the quantities x_H and x_L are calculated by using the equations mentioned here below.

$$x_H = \exp(\overline{x} + k_N s) \quad 3.3.3.1$$

$$x_L = \exp(\bar{x} - k_N s) \qquad 3.3.3.2$$

Where $\bar{x}$ and s are the mean and standard deviation of the natural logarithms of the sample, respectively, and k_N is the G-B statistic by Grubbs and Becks(1972).At the 10% significance level, the following approximately is used, where N is the sample size.

$$k_N = -3.62201 + 6.28446N^{1/4} - 2.49835N^{1/2} + 0.491436N^{3/4} - 0.037911N \qquad 3.3.3.3$$

Sample values greater than x_H are considered to be high outliers, while those less than x_L are considered to be low outliers.

The other methods of determining outliers includes the US water resources council(1981) method, which has a very high threshold. The least median square(LMS) method is also the alternative method of testing outliers (Rao and Hamed, 2000).

3.4. Identifying best fit distribution functions

3.4.1. Lists of the extreme maximum distribution functions

3.4.1.1. Normal distribution

The normal distribution is used in frequency analysis for fitting empirical distributions to hydrological, and in simulation of data. As many statistical parameters are approximately normally distributed, the normal distribution is often used for statistical inferences.

The probability density function of a normally distributed variable x is given by:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} \qquad 3.4.1.1.1$$

Where μ and σ are the parameters of the distribution. The variables x can take any value in the range (-∞,∞).The standard normal variant *u* is a normal variable with a mean equal to zero and standard deviation equal to one (Rao and Hamed, 2000).

3.4.1.2. Two-Parameter Lognormal (LN(2)) Distribution

The probability density function of a logarithmic normally distributed variable *x* with two parameters(LN(2)) is given by:

$$f(x) = \frac{1}{x\alpha_y\sqrt{2\pi}} \exp\left\{\frac{\left[-\log x - u_y\right]^2}{2\sigma_y{}^2}\right\} \qquad 3.4.1.2.1$$

Where μ_y and σ_y are the mean and standard deviation of natural logarithm of x. If the variable log(x) is standardized as in Eq.2,

$$u = \frac{\log(x) - \mu_y}{\sigma_y} \qquad 3.4.1.2.2$$

The standard normal variant *u* is obtained with Eq.2 (Rao and Hamed,2000).

3.4.1.3. Three Parameter Lognormal (LN(3)) Distribution

The three-parameter lognormal distribution is similar to the LN(2) distribution except that *x* is shifted by an amount *(a)* which represents a lower bound. The normally distributed variable becomes log(x-a) with the pdf,

$$f(x) = \frac{1}{(x-a)\sigma_y\sqrt{2\pi}} \exp\left\{-\frac{1}{2\sigma_y^{\ 2}}\left[\log(x-a)-\mu_y\right]^2\right\} \qquad 3.4.1.3.1$$

where μ_y and $\sigma_y^{\ 2}$ are the location and scale parameters, which correspond to the mean and variance of the logarithm of the shifted variable $(x-a)$ .The standardized variable u is obtained as in Eq.3.4.1.3.2(Rao and Hamed, 2000).

$$u = \frac{\log(x-a)-\mu_y}{\sigma_y} \qquad 3.4.1.3.2$$

Sangal and Biswas (1970) suggested a procedure to estimate the parameters of the LN(3) distribution in which only the mean, median and standard deviation of the data are used. The mathematical properties of the LN(3) distribution were discussed by Burges et al(1975).They also compared two methods of estimation of the third parameter „a" of the LN(3) distribution (Rao and Hamed, 2000).

Singh and Singh(1988) used the maximum entropy method to estimate the parameters of the LN(3) distribution.

3.4.1.4. Exponential distribution

The exponential is a special case of the Gamma family of distributions which include the Pearson(3),one and two parameter Gamma,log-Pearson and the generalized Gamma distributions (Rao and Hamed, 2000).

The pdf of the Pearson(3) distribution is given by:

$$f(x) = \frac{1}{\alpha^{\beta}.\Gamma(\beta)}(x-\varepsilon)^{\beta-1}e^{\frac{(x-\beta)}{\alpha}} \qquad 3.4.1.4.1$$

where $\Gamma(.)$ is the gamma function defined by Eq 3.4.1.4.2.

$$\Gamma(y+1) = \int_0^{\infty} t^y e^{-t}dt, y+1>0 \qquad 3.4.1.4.2$$

with the following properties

$$\Gamma(y+1) = y\Gamma(y), y>0 \qquad 3.4.1.4.3$$

$$\Gamma(y) = \Gamma(y+1)/y, y < 1 \qquad 3.4.1.4.4$$

$$\Gamma(n) = (n-1)!, \qquad 3.4.1.4.5$$

n is a positive integer

3.4.1.5. Two-Parameter Gamma(G(2)) Distribution

The two-parameter gamma(G(2)) distribution is a special case of the Pearson(3) distribution in which the parameter ε in Eq.3.4.4.1 is equal to zero. The pdf of the G(2) distribution is thus

$$f(x) = \frac{1}{\alpha^{\beta}\Gamma(\beta)} x^{\beta-1} e^{-(x/a)} \qquad 3.4.1.5.1$$

The variable x in this case has a lower bound of zero instead of ε,0< x <∞.the distribution function of x cannot be obtained in closed form (Rao and Hamed, 2000).

3.4.1.6. Pearson(3) Distribution

The probability density function of the Pearson(3) distribution is given by Eq. 3.4.1.5.1

$$f(x) = \frac{1}{\alpha\Gamma(\beta)}\left(\frac{x-\gamma}{\alpha}\right)^{\beta-1} e^{-\left(\frac{x-\gamma}{\alpha}\right)} \qquad 3.4.1.6.1$$

The variable x in a Pearson(3) distribution can take values in the range $\gamma < x < \infty$ .Generally, α can be positive or negative, but for negative values of α the distribution becomes upper bounded and is therefore not suitable for analyzing maximum events (Rao and Hamed, 2000).

3.4.1.7. Log-Pearson(3) Distribution

If the variable $\log x$ is assumed to have a Pearson(3) distribution when the distribution of the variable x is a log-Pearson(3),(LP(3)) distribution. The probability density function of a LP(3) distributed random variable is given by:

$$f(x) = \frac{1}{\alpha x\Gamma(\beta)}\left[\frac{\log(x)-\gamma}{\alpha}\right]^{\beta-1} e^{-\left\{\frac{\log(x)-\gamma}{\alpha}\right\}} \qquad 3.4.1.7.1$$

The pdf of the log-Pearson(3) distribution may take many different shapes (Rao and Hamed, 2000).

3.4.1.8. Generalized Extreme Value Distribution(GEV) Distribution

The probability density function of the GEV distribution is of form

$$f(x) = \frac{1}{\alpha}\left[1-k\left(\frac{x-u}{\alpha}\right)\right]^{1/k-1} e^{-\left[1-k\left(\frac{x-u}{\alpha}\right)\right]^{1/k}} \qquad 3.4.1.8.1$$

The range of the variable x depends on the sign of the parameter (Rao and Hamed, 2000).

3.4.1.9. The Extreme Value Type I EV1(2) Distribution

The probability density function of the EV1(2) distribution is given by Eq.3.4.1.9.1.

$$f(x)=\frac{1}{\alpha}\exp\left[-\left(\frac{x-\beta}{\alpha}\right)-e^{-\left(\frac{x-\beta}{\alpha}\right)}\right] \qquad 3.4.1.9.1$$

The variable x take values in the range $-\infty< x <\infty$ (Rao and Hamed,2000).

3.4.2. Use the Easyfit software

The easy fit software used to determine or identify the best fit extreme maximum distribution function for Bahir Dar city. In addition, this software also helped to determine the distribution parameters of the best selected function. Normally, the selection of the distribution function was mainly relying on the annual maximum daily rainfall data of the city.

3.5. Goodness of tests

The goodness of fit test measures the compatibility of random sample with the theoretical probability distribution. The goodness of fit tests is applied for testing the following null hypothesis:

H_0 :the maximum daily rainfall data follow the specified distribution.

H_A :the maximum daily rainfall does not follow the specified distribution.

The following goodness of fit tests viz. Kolmogorov-Simirnov test and Anderson-darling test will be used along with the chi-square test at α level of significance for the selection of the best fit probability distribution.

3.5.1. Kolmogorov-Smirnov Test

The Kolmogorov-Smirnov statistic(D) is defined as the largest difference between the theoretical and empirical cumulative distribution function(ECDF):

$$D=\max\left(F(x_i)-\frac{i-1}{n},\frac{i}{n}-F(x_i)\right) \qquad 3.5.1.1$$

$$(1\le i\le n)$$

where X_i is the random sample, i=1,2,……..n

$$CDF=F_n(x)=\frac{1}{n}.\left[Number\quad of\quad observations\quad \le\quad x\right] \qquad 3.5.1.2$$

This test is used to decide if a sample comes from hypothesized continuous distribution(Rao and Hamed,2000).

3.5.2. Anderson-Darling Test

The Anderson-Darling Statistic(A^2) is defined as

$$A^2 = -n - \frac{1}{n}\sum_{i=1}^{n}(2i-1).\left[InF(X_i) + In(1-F(X_{n-i+1}))\right] \qquad 3.5.2.1$$

It is a test to compare the fit of an observed cumulative distribution to an expected cumulative distribution function. This test gives more weight to the tails than Kolmogorov-Smirnov test (Rao and Hamed, 2000).

3.5.3. Chi-Square Test

The Chi-Square statistic is defined as

$$\chi^2 = \sum_{i=1}^{k}\frac{(O_i - E_i)^2}{E_i} \qquad 3.5.3.2$$

where O_i is the observed frequency, E_i is the expected frequency and „I" is the number of observations(1,2,……..k),which is calculated by $E_i = F(x_2) - F(x_1)$ and F=the CDF of the probability distribution being tested.

The observed number of observation(k) in interval „i" is computed from equation given below.

$$k = 1 + \log_2 n \qquad 3.5.3.3$$

n=sample size.

this test is continuous sample data only and is used to determine if a sample comes from a population with a specific distribution (Rao and Hamed, 2000).

3.6. Simple scaling and time scale invariant characteristics

The scaling or scale-invariant models enable us to transform data from one temporal or spatial model to another one, and thus, help to overcome the difficulty of inadequate data. A natural process fulfills the simple scaling property if the underlying probability distribution of some physical measurements at one scale is identical to the distribution at another scale, multiplied by a factor that is a power function of the ratio of the two scales. The basic theoretical development of scaling has been investigated by many authors.

Let $X(t)$ and $X(\lambda t)$ denote measurements at two distinct time or spatial scales t and λt ,respectively. Definition of scaling property of the probability of the $X(t)$ is

$$X(t) = \lambda^{-H} X(\lambda t) \qquad 3.6.1$$

where λ denotes a scale factor and H is a scale exponent which varies with location. Gupta and Waymire introduced the notions of strict and wide sense simple scaling. The strict sense simple scaling above equation implies that $X(t)^q$ and $(\lambda^{-H}X(\lambda t))^q$ have the same probability distribution. The wide sense simple scaling is expressed as they have the same moments, i.e.

$$E\left[X(t)^q\right]=\lambda^{-Hq}E\left[X(\lambda t)^q\right] \qquad 3.6.2$$

The scaling exponent (Hq)can be estimated from the slope of linear regression relationship between the log transformed values of moment, $\log E\left[X(\lambda t)^q\right]$and the scale parameters log λ for various order of moment q .This is definition of a "wide sense" simple scaling. A "wide sense" simple scaling with t=1 is given by:

$$E\left[X(1)^q\right]=\lambda^{-Hq}E\left[X(\lambda)^q\right] \qquad 3.6.3$$

If the scaling exponent H is not constant and changes probabilistically, equation 3 described as:

$$E\left[X(1)^q\right]=\lambda^{-K(q)}E\left[X(\lambda)^q\right] \qquad 3.6.4$$

where $K(q)$ is a function of the moment order. The procedure to test the suitability of scale invariant model to shown the figure below. The moments $E\left[X(\lambda t)^q\right]$are plotted on the logarithmic chart versus the scale λ for different moments" order q .The slope $K(q)$ is plotted on the linear chart versus the moment order q .If the resulting graph is a straight line, the multi-scaling approach has to be considered. Let the random variable $I(d)$,the maximum annual value of local rainfall intensity over a duration d .It is defined as:

$$I(d)=\max_{0\le t\le 1\,year}\left[\frac{1}{d}\int_{t-d/2}^{t+d/2}X(\xi)d\xi\right] \qquad 3.6.5$$

where $X(\xi)$ is a time continuous stochastic process representing rainfall intensity and d is duration. It is supposed $I(d)$ represents the annual maximum rainfall intensity of duration d ,defined by the maximum value of moving average of width d of continuous rainfall process.

Here, some concepts are introduced about scaling of the probability distribution of random functions. A generic random function $I(d)$ is denoted by simple scaling properties if it obeys the following.

$$I(d)=\left(\frac{D}{d}\right)^{-H_d}I(D) \qquad 3.6.6$$

D is an aggregated time duration, i.e. 2,3,…,24hours.

Defining the scaling ratio is $\lambda_d=\dfrac{D}{d}$.

$$I(d) = \lambda^{-H_d} I(\lambda_d d) \quad 3.6.7$$

Equation 7 is rewritten in terms of the moments of order q about the origin, denoted by $E\left(I(d)^q\right)$.

The resulting expression is:

$$E\left[I(d)^q\right] = \lambda_d^{-H_d q} E\left[I(\lambda_d d)^q\right] \quad 3.6.8$$

If one assumes the wide sense simple scaling exists, the distribution of IDF for short-duration rainfall intensity can be derived from daily rainfall (Afrin et..al., 2015).

3.6.1. Simple Scaling Parameters and Formula

All forms of the generalized relationships assume that rainfall depth or intensity is inversely related to the duration of a storm raised to a power, or scale factor(Islam and Rahman,2015).

Scale invariance occurs when the connections among the statistical descriptors of a given phenomenon at different scales are constant and defined by scale factor. The statistical descriptors can be scaled either by a single factor(simple scaling or by a more complex function of the scale(multi scaling) (Naghettini, 2012).

According to Koutsoyannis, IDF relationships are particular cases of the following general formula.

$$i = \frac{w}{(d^v + \theta)^\eta} \quad 3.6.1.1$$

Where i denotes the rainfall intensity of duration d, and w, v, θ ,and η are non-negative coefficients. Koutsoyannis also proposed a numerical exercise, in which they show that the errors resulting from imposing v=1 in equation (1) are much smaller than the typical parameter and quintile estimation errors from limited size samples of rainfall data. Hence, considering $v \neq 1$ as model over-parameterization, Koutsoyannis prescribed that, for a given period, the general expression for IDF relationships should be written as;

$$i = \frac{w}{(d + \theta)^\eta} \quad 3.6.1.2$$

Rigorously, the coefficients w, θ ,and η depend on the return period. However, because the IDF curves for different return periods cannot intersect each other, this dependence cannot be arbitrary. Actually, this restriction imposes bounds for the range of variation of the parameters w, θ ,and η .For instance, if (w1, $\theta 1$,and $\eta 1$) and (w2, $\theta 2$,and $\eta 2$) denote the parameter sets for the return period T1 and T2 respectively, with T2<T1,Koutsoyannis et al.(1998) suggests the following restrictions on the range of variation of parameters.

$$\theta 1 = \theta 2 = \theta \geq 0 \quad 3.6.1.3.$$

$$0 < \eta_1 = \eta_2 = \eta < 1 \quad 3.6.1.4$$

$$w_1 > w_1 > 0 \quad 3.6.1.5$$

In this set of restrictions, it is worth to note the only parameter that can consistently increase with increasing return periods is w ,which results in substantial simplification of equation(2).In fact, these arguments justify the formulation of the following general model for IDF.

$$i_{d,T} = \frac{a(T)}{b(d)} \quad 3.6.1.6.$$

Which exhibits the advantage of expressing separable dependence relations between i and T, and between i and d. In equation(4), b(d)= $(d \;\; +\theta)^{\eta}$ with θ>0and 0<η <1,whereas a(T) is completely defined by the probability distribution function of the maximum rainfall intensities (Nhat et.al.,2006).

3.7. Determining of the Design Rainfall using Frequency Factor

The Determination of the design rainfalls for 2,5,10,25,50 and 100 years of return periods were considered in order have permissible rainfall in any of the specified return periods. The design period calculation was followed through the use of frequency factor method. The formula for this application is shown here with below.

$$x_T = \bar{x} + K_T s \;, \quad 3.7.1$$

where x_T is the design rainfall to be calculated, $\bar{x}$ is the mean annual maximum daily rainfall, K_T is the frequency factor for extreme distributions, $K_T = -\frac{\sqrt{6}}{\pi}\left\{0.5772 + \ln\left[\ln\left(\frac{T}{T-1}\right)\right]\right\}$ and T -is the respective return period(Chow,1988).

3.8. Comparison of IDF curves developed using simple scaling method and Regional IDF curves in Ethiopia both for Bahir Dar city.

ERA(2002) drainage manual provides six region or regime areas. And it was also possible to know the regional have individual IDF curves in which the required study can be performed from these regional curves. At the same time, this study was engaged in developing IDF relationship and its respective IDF curves for Bahir Dar city using the annual maximum daily rainfall data over a per of 41 years for the city. The logic of comparison between the regional IDf curves developed in Ethiopia and the one of the IDF curves developed for Bahir Dar city was compared in order to know the relationship(the similarity of the rainfall intensity for Bahirdar city).Also using the statistical analysis, rainfall intensity duration curves have been developed for commonly used design frequencies. These basic information will be used in estimating peak flood for small and large watersheds.

The IDF curves from regional analysis was identified for the study area, Bahir Dar city in our case, and tried to be compared with the IDF curve developed from simple scaling method.

CHAPTER 4: RESULTS AND DISCUSSIONS

4.1. Rainfall data evaluation

4.1.1. Rainfall data missing and filling of the missing rainfall data

Table 2: Missing and infilling of annual maximum daily rainfall(mm) data calculation (1975-2015).

Years	Jan	Feb	Mar	Apr	May	Jun	July	Aug	Sept	Oct	Nov	Dec	Max daily Ann.RF(mm)
1975	9.8	3.2	0.0	6.0	6.6	47.1	64.3	44.4	45.6	45.0	14.0	14.8	64.3
1976	0.0	4.1	4.0	14.3	66.8	48.6	45.2	73.4	59.8	6.4	11.8	9.9	73.4
1977	0.0	0.0	4.2	0.0	13.3	55.4	133.2	35.7	32.5	30.4	7.4	4.2	133.2
1978	1.1	0	8.7	19.9	18.7	39.8	38	27.8	61.1	36.8	18	0.1	61.1
1979	9.2	0	0.1	0.8	28.3	32.4	56.2	44	29.6	36.3	0.1	0.2	56.2
1980	0	2.2	7.8	37.5	15.1	39.8	55.4	79.9	43.3	12.2	1.2	0	79.9
1981	0	0	0	52.8	16.3	19.9	108.7	58.9	32.3	26.8	3.8	0	108.7
1982	4.7	0.1	30	4.6	19.6	24.6	32.3	32.5	35.9	23.5	6	0	35.9
1983	0	0.3	0	0.5	18.2	40.3	46	64.3	36.5	43.3	4.4	0	64.3
1984	0	0	2.3	0.3	12.4	57.8	47.8	47.3	70.4	0	0.1	7.5	70.4
1985	3.5	0	2.9	16.1	34.2	83.3	82.2	42.3	46.2	14.7	2.2	0	83.3
1986	0	0.3	2.6	7.1	7.2	42.4	58.3	48.2	35.3	36.9	0	0	58.3
1987	0	0.1	0.7	6.4	40.9	47	21.5	47.4	30.3	35.3	9.3	0	47.4
1988	1.1	11	0	0	12.8	46.6	44.8	30.9	57.1	29.7	16.3	8.6	57.1
1989	0	0	3.6	5.2	61.1	45	50.7	57	41.9	15.2	15	2.1	61.1
1990	2.3	0.3	0.8	6.5	7.5	30.2	114.1	51.2	46.5	29.1	0	0	114.1
1991	0	0.1	4.54	12.41	22.59	42.3	90.5	42.3	34.5	58.2	0.8	1.8	90.5
1992	0	0	1.4	35.6	43	23.3	53.8	41	24.4	44.3	17.1	0	53.8
1993	7.2	0.2	4.3	12.5	25.2	41.3	52.7	58.3	48.6	25.5	18.5	0.3	58.3
1994	0	0	0.3	13.7	19.9	39	38.8	33.1	31.2	8.5	3.1	4.3	39.0
1995	0	3.2	4.4	14.4	27.8	58.6	56.3	38.8	17.2	13.6	8.4	3.8	58.6
1996	0	0.2	16.5	25.8	16.4	46.5	51.1	34.4	30.8	26.8	11.4	0	51.1
1997	0	0	7.8	18.3	69.2	28.1	16.1	28	25.5	24.4	10.2	10.1	69.2
1998	0	0	16.9	0.6	32.4	34.3	51.2	51.5	38.8	27.1	1.1	0	51.5
1999	7.9	0	0	3.7	14.1	33.8	55	64.6	48.8	37.1	3	0	64.6
2000	0	0	0.3	31.5	50.5	35.7	44.1	105.3	45.7	31.6	27.6	0	105.3
2001	0	0	0.5	10.5	30.4	53.7	46	124.7	23.5	22.4	9.9	9.1	124.7
2002	0	1.2	4.7	13.2	1	68.2	58.3	89.5	25	12.3	0	0.6	89.5
2003	0	0	0.2	0	0.5	56	76.9	49.3	70.3	32	5	4.6	76.9
2004	6.3	20	4.6	26	6.4	28.9	95.5	57	67	31	3.3	0	95.5
2005	1.7	0	16.2	4.9	22.59	40.82	58.64	52.62	39.83	26.78	6.87	2.88	58.6
2006	3.1	0.2	0.1	4.1	34.6	68.2	85.7	72	36	25.1	0.3	2.6	85.7
2007	0	0	1.1	24.2	4.1	43.2	57.7	74.1	45	38.7	4.1	0	74.1
2008	1.2	0	0	36.1	20.2	24.6	68.3	39.7	33.3	30.9	12.9	0	68.3
2009	0	0	3.6	0	3.8	19.8	47.3	64.5	29.7	20.6	2.6	0	64.5

2010	8.3	0	0	27.5	26	19.1	49.1	58.1	38.6	21.5	1.5	0	58.1
2011	0	0	8.3	9.8	41.3	72.8	72.3	27	31.2	14	10.5	0	72.8
2012	0	0	0	0	9.7	31.2	61.8	96.7	55.7	4.8	1	10.5	96.7
2013	0	0	1.4	1	40.1	36.3	86.4	34.5	26.9	44.4	9.4	0	86.4
2014	0	0	25.6	17.3	22.59	35	44.1	48.2	34.5	47.2	0	0.4	48.2
2015	0	0.8	0.4	0	30.6	33.7	46.4	39.7	66.7	34.5	10.4	22.5	66.7
Total	67.4	46.7	186.3	508.7	926.2	1673.8	2404.1	2157.5	1633.2	1098.1	281.7	118.0	
Avg	1.64	1.14	4.54	12.41	22.59	40.82	58.64	52.62	39.83	26.78	6.87	2.88	

In the course of the 41 years of data recording from the national metrological station, it was noted that there was 3 months(March-May) data missing in 1991,8 months(May-December) in 2005,and 1 month (May) in 2014. In general, it was noted that over the course of 41 years period of data collection, there were 3 times data missing for the month of May and other months were partly experiencing only a singular time of data missing. This was clearly mentioned in the table above in order to clarify the discussion here. The missing data filling technique was mainly based on taking the average value of a month over the period of 41 years and fill the missed rainfall data to the specific month to make sure that all months have got their own maximum monthly rainfall data from the annual maximum daily data. And it was really practical to do this way as it was noted that some years got their annual maximum rainfall data specifically from this average method of maximum monthly rainfall data calculation. From the table above, it was clear to note that 25% of the rainfall data was missed in 1991,66.67% of the rainfall data missed in 2005 and 7.14% of the data was missed in 2014.The overall missing data from the 41 years daily rainfall data was 2.44% (12 months missed data in 1991,2005 and 2014) divided by the total months of over 41 years (492 months).

4.2. Rainfall data and data sources

4.2.1. Annual maximum daily rainfall data

Table 3:Annual maximum daily rainfall data

Years	Annual maximum daily rainfall(mm)
1975	64.3
1976	73.4
1977	133.2
1978	61.1
1979	56.2
1980	79.9
1981	108.7
1982	35.9
1983	64.3
1984	70.4
1985	83.3
1986	58.3

1987	47.4
1988	57.1
1989	61.1
1990	114.1
1991	90.5
1992	53.8
1993	58.3
1994	39
1995	58.6
1996	51.1
1997	69.2
1998	51.5
1999	64.6
2000	105.3
2001	124.7
2002	89.5
2003	76.9
2004	95.5
2005	58.64
2006	85.7
2007	74.1
2008	68.3
2009	64.5
2010	58.1
2011	72.8
2012	96.7
2013	86.4
2014	48.2
2015	66.7

It was also possible to understand that it is only in 2005 that no annual maximum daily rainfall data as compared to the other years. This was because of the filled data was only managed be selected in 2005 while none in other years. This leads us to a conclusion that the annual maximum daily missing data was 2.44%.This implies that 97.56% of the data was fully recorded and it is very good data recording when it comes to the annual maximum daily data.

4.2.2. Hourly rainfall data

Table 4:annual maximum sub hourly and hourly rainfall data(2012-2015)

Annual maximum sub hourly and hourly rainfall data(2012-2015)							
Year	15min RF(mm)	30min RF(mm)	45min RF(mm)	60min RF(mm)	120min-RF(mm)	180min-RF(mm)	24hr RF(mm)
2012	28.6	39.4	46.8	58	87.4	93.6	128.8
2013	35	54.2	59.2	61.6	66.6	67	88
2014	23	30.6	36.8	41.6	46.2	46.6	82.2
2015	20.2	31.2	38	41.8	55.2	58	72.8
Avg	**26.70**	**38.85**	**45.20**	**50.75**	**63.85**	**66.30**	**92.95**

However, the same data recording efficiency was not noted in the automatic sub hourly and hourly rainfall data recording. The study was only basing on 4 years fully recording sub hourly and hourly rainfall data. In this regards, it was purposefully selection made to these years out of the rainfall data that was recorded since 2009.The main reason was that the sub hourly and hourly rainfall data are the driving factors in order to reach to develop the IDF curves and relationships for Bahir dar city. In most literature review, it was normal and recommended to consider number of years with full sub hourly and hourly rainfall recorded data. The intention was mainly to reach to a reliable results and findings at the development of the IDF relationships and curves of the study area, or Bahir dar city. Therefore, the was no any missing sub hourly and hourly rainfall data in the 12 months of the 4 years mentioned above(2012-2015).

4.3. Tests for daily rainfall data

4.3.1.1. Test for Independence and Stationarity

Using the Wald-Wolfowitz(1943)(W-W) test, the Bahir Dar annual maximum daily rainfall data can be calculated and the values be determined. Hence, it follows as below.

Table 5:Statistic R calculation

SN	Years	Max annual daily rainfall(mm)	R calculation
1	1975	64.3	4719.62
2	1976	73.4	9776.88
3	1977	133.2	8138.52
4	1978	61.1	3433.82
5	1979	56.2	4490.38
6	1980	79.9	8685.13
7	1981	108.7	3902.33
8	1982	35.9	2308.37

9	1983	64.3	4526.72
10	1984	70.4	5864.32
11	1985	83.3	4856.39
12	1986	58.3	2763.42
13	1987	47.4	2706.54
14	1988	57.1	3488.81
15	1989	61.1	6971.51
16	1990	114.1	10326.05
17	1991	90.5	4868.9
18	1992	53.8	3136.54
19	1993	58.3	2273.7
20	1994	39	2285.4
21	1995	58.6	2994.46
22	1996	51.1	3536.12
23	1997	69.2	3563.8
24	1998	51.5	3326.9
25	1999	64.6	6802.38
26	2000	105.3	13130.91
27	2001	124.7	11160.65
28	2002	89.5	6882.55
29	2003	76.9	7343.95
30	2004	95.5	5600.12
31	2005	58.64	5025.448
32	2006	85.7	6350.37
33	2007	74.1	5061.03
34	2008	68.3	4405.35
35	2009	64.5	3747.45
36	2010	58.1	4229.68
37	2011	72.8	7039.76
38	2012	96.7	8354.88
39	2013	86.4	4164.48
40	2014	48.2	3214.94
41	2015	66.7	4288.81
		R	**219,747.39**

The statistic R can be calculated as below like.

$$R = \sum_{i=1}^{41-1} x_i x_{i+1} + x_1 x_{41}$$

$$R = x_1x_2 + x_2x_3 + x_3x_4 + ... + x_{40}x_{41} + x_1x_{41}$$

$R = 64.3*73.4 + 73.4*133.2 + 133.2*61.1 + ... + 48.2*66.7 + 64.3*66.7$

$R = 219,747.39$

Accordingly, R follows the mean and variance of the following values.

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

As mentioned above, the values of $s_r = Nm_r^{'}$,where $m_r^{'}$ is the rth moment of the sample about the origin.

Hence, the sample moments about the origin, which implies $\alpha = 0$, $X_i = 0$=raw score.

First Moment(M_1)-Mean=$\overline{X}$

$$M_1 = \frac{1}{N}\sum_{i=1}^{N} X_i = \frac{1}{N}\sum_{i=1}^{N}(freq * X)$$

Second Moment(M_2)

$$M_2 = \frac{1}{N}\sum_{i=1}^{N} X_i^2 = \frac{1}{N}\sum_{i=1}^{N}(freq * X^2)$$

Third Moment(M_3)

$$M_3 = \frac{1}{N}\sum_{i=1}^{N} X_i^3 = \frac{1}{N}\sum_{i=1}^{N}(freq * X^3)$$

Forth Moment(M_4)

$$M_4 = \frac{1}{N}\sum_{i=1}^{N} X_i^4 = \frac{1}{N}\sum_{i=1}^{N}(freq * X^4)$$

(Spring, 2000)

Table 6:Statistic parameters calculations

Column A	Column B	Column C	Column D	Column E	Column F	Column G
SN	Years	Max annual daily rainfall(mm)	R calculation	Mean	C-E	F*F
1	1975	64.3	4719.62	72.62	-8.32	69.19
2	1976	73.4	9776.88	72.62	0.78	0.61
3	1977	133.2	8138.52	72.62	60.58	3670.17
4	1978	61.1	3433.82	72.62	-11.52	132.67
5	1979	56.2	4490.38	72.62	-16.42	269.55
6	1980	79.9	8685.13	72.62	7.28	53.03

7	1981	108.7	3902.33	72.62	36.08	1301.91
8	1982	35.9	2308.37	72.62	-36.72	1348.22
9	1983	64.3	4526.72	72.62	-8.32	69.19
10	1984	70.4	5864.32	72.62	-2.22	4.92
11	1985	83.3	4856.39	72.62	10.68	114.10
12	1986	58.3	2763.42	72.62	-14.32	205.01
13	1987	47.4	2706.54	72.62	-25.22	635.95
14	1988	57.1	3488.81	72.62	-15.52	240.81
15	1989	61.1	6971.51	72.62	-11.52	132.67
16	1990	114.1	10326.05	72.62	41.48	1720.75
17	1991	90.5	4868.9	72.62	17.88	319.76
18	1992	53.8	3136.54	72.62	-18.82	354.12
19	1993	58.3	2273.7	72.62	-14.32	205.01
20	1994	39	2285.4	72.62	-33.62	1130.17
21	1995	58.6	2994.46	72.62	-14.02	196.51
22	1996	51.1	3536.12	72.62	-21.52	463.03
23	1997	69.2	3563.8	72.62	-3.42	11.68
24	1998	51.5	3326.9	72.62	-21.12	445.97
25	1999	64.6	6802.38	72.62	-8.02	64.29
26	2000	105.3	13130.91	72.62	32.68	1068.11
27	2001	124.7	11160.65	72.62	52.08	2712.53
28	2002	89.5	6882.55	72.62	16.88	285.00
29	2003	76.9	7343.95	72.62	4.28	18.34
30	2004	95.5	5600.12	72.62	22.88	523.58
31	2005	58.64	5025.448	72.62	-13.98	195.39
32	2006	85.7	6350.37	72.62	13.08	171.14
33	2007	74.1	5061.03	72.62	1.48	2.20
34	2008	68.3	4405.35	72.62	-4.32	18.65
35	2009	64.5	3747.45	72.62	-8.12	65.90
36	2010	58.1	4229.68	72.62	-14.52	210.77
37	2011	72.8	7039.76	72.62	0.18	0.03
38	2012	96.7	8354.88	72.62	24.08	579.94
39	2013	86.4	4164.48	72.62	13.78	189.94
40	2014	48.2	3214.94	72.62	-24.42	596.24
41	2015	66.7	4288.81	72.62	-5.92	35.02
		R	**219,747.39**		sum	19832.06
		Mean	**72.62**		**Variance**	**495.80**
					St dev	22.27

Table 7:Average R calculation

SN	Years	X	X^2	X^3	X^4
1	1975	64.3	4134.49	265847.71	17094007.56
2	1976	73.4	5387.56	395446.9	29025802.75
3	1977	133.2	17742.24	2363266.4	314787080.2
4	1978	61.1	3733.21	228099.13	13936856.9
5	1979	56.2	3158.44	177504.33	9975743.234
6	1980	79.9	6384.01	510082.4	40755583.68
7	1981	108.7	11815.69	1284365.5	139610530.2
8	1982	35.9	1288.81	46268.279	1661031.216
9	1983	64.3	4134.49	265847.71	17094007.56
10	1984	70.4	4956.16	348913.66	24563521.95
11	1985	83.3	6938.89	578009.54	48148194.43
12	1986	58.3	3398.89	198155.29	11552453.23
13	1987	47.4	2246.76	106496.42	5047930.498
14	1988	57.1	3260.41	186169.41	10630273.37
15	1989	61.1	3733.21	228099.13	13936856.9
16	1990	114.1	13018.81	1485446.2	169489413.8
17	1991	90.5	8190.25	741217.63	67080195.06
18	1992	53.8	2894.44	155720.87	8377782.914
19	1993	58.3	3398.89	198155.29	11552453.23
20	1994	39	1521	59319	2313441
21	1995	58.6	3433.96	201230.06	11792081.28
22	1996	51.1	2611.21	133432.83	6818417.664
23	1997	69.2	4788.64	331373.89	22931073.05
24	1998	51.5	2652.25	136590.88	7034430.063
25	1999	64.6	4173.16	269586.14	17415264.39
26	2000	105.3	11088.09	1167575.9	122945739.8
27	2001	124.7	15550.09	1939096.2	241805299
28	2002	89.5	8010.25	716917.38	64164105.06
29	2003	76.9	5913.61	454756.61	34970783.23
30	2004	95.5	9120.25	870983.88	83178960.06
31	2005	58.64	3438.65	201642.41	11824311.07
32	2006	85.7	7344.49	629422.79	53941533.36
33	2007	74.1	5490.81	406869.02	30148994.46
34	2008	68.3	4664.89	318611.99	21761198.71
35	2009	64.5	4160.25	268336.13	17307680.06
36	2010	58.1	3375.61	196122.94	11394742.87
37	2011	72.8	5299.84	385828.35	28088304.03
38	2012	96.7	9350.89	904231.06	87439143.79

39	2013	86.4	7464.96	644972.54	55725627.8
40	2014	48.2	2323.24	111980.17	5397444.098
41	2015	66.7	4448.89	296740.96	19792622.23
		S_1	S_2	S_3	S_4
		2977.34	236040.7	20408733	1912510916

$$\overline{R} = 215{,}712.82$$

$$Var(R) = 8{,}990{,}556.16$$

Standard deviation of R is calculated below.

$$s_R = 2{,}998.43$$

$$u = \frac{(R - \overline{R})}{(\text{var}(R))^{1/2}}$$

$$u = \frac{(219{,}747.39 - 215{,}712.82)}{2998.43}$$

$$u = \frac{(219{,}747.39 - 215{,}712.82)}{2998.43}$$

$$u = 1.35$$

The test value $u = 1.35$ is less than the critical value at 5% significance level $u_{0.025} = 1.96$. Thus we can accept the hypothesis of independent and stationarity. The Bahir dar maximum annual daily rainfall data are concluded to be independent and stationary at the 5% significance level.

4.3.1.2. Test for Homogeneity and Stationarity

Using the Mann-Whitney(1947)(M-W) test, we can determine the parameters for the test as shown below.

Table 8: Homogeneity and stationary calculations
Sub divide the data in to p and q data series with p=20 and q=21

p	Annual maximum daily RF data	q	Annual maximum daily RF data	N=P+q	Annual maximum daily RF data
1	64.3	1	58.6	1	35.9
2	73.4	2	51.1	2	39
3	133.2	3	69.2	3	47.4
4	61.1	4	51.5	4	48.2
5	56.2	5	64.6	5	51.1
6	79.9	6	105.3	6	51.5
7	108.7	7	124.7	7	53.8
8	35.9	8	89.5	8	56.2
9	64.3	9	76.9	9	57.1
10	70.4	10	95.5	10	58.1
11	83.3	11	58.64	11	58.3
12	58.3	12	85.7	12	58.3
13	47.4	13	74.1	13	58.6
14	57.1	14	68.3	14	58.64
15	61.1	15	64.5	15	61.1
16	114.1	16	58.1	16	61.1
17	90.5	17	72.8	17	64.3
18	53.8	18	96.7	18	64.3
19	58.3	19	86.4	19	64.5
20	39	20	48.2	20	64.6
		21	66.7	21	66.7
				22	68.3
				23	69.2
				24	70.4
				25	72.8
				26	73.4
				27	74.1
				28	76.9
				29	79.9
				30	83.3
				31	85.7
				32	86.4
				33	89.5
				34	90.5
				35	95.5
				36	96.7

	37	105.3
	38	108.7
	39	114.1
	40	124.7
	41	133.2
R	380	
V	170	
W	250	
U	170	

The calculation for the parameters are shown below.

By sub dividing the original data series of size $N = p + q = 41$ in to $p = 20$ and $q = 21$,we can rank the combined data set series in to increasing order as described above. Following this, it is possible to determine R,V,W and U as follows.

R is the sum of the ranks of the elements of the first sample(size p) in combined series(size N),the yellow highlighted ranks in the above table.

$$R = 1+2+3+7+8+9+11+12+15+16+17+18+24+26+29+30+34+38+39+41 = 380$$

From the above formulae, we can determine the values of V and W as follows.

$$V = R - \frac{(p(p+1))}{2}$$

$$V = 380 - \frac{(20(20+1))}{2} = 380 - 210 = 170$$

Again, W can also be determined from the above formula as well.

$$W = pq - V$$

$$W = 20*21 - 170 = 420 - 170 = 250$$

Accordingly, we can also determine U from the approaches given above. In the concept above, U is the smaller value of V and W.

This leads that $U = 170$.

In the same fashion, we can also determine $\overline{U}$ and $\text{var}(U)$ accordingly.

$$\overline{U} = \frac{pq}{2}$$

$$\overline{U} = \frac{20*21}{2} = 210$$

$$\text{var}(U) = \left[\frac{pq}{N(N-1)}\right]\left[\frac{N^3 - N}{12} - \sum T\right]$$

The summation of T is zero as there are no J values observed as tied observations at a given rank. This guide us to:

$$\text{var}(U) = \left[\frac{20*21}{41(41-1)}\right]\left[\frac{41^3 - 41}{12} - \sum 0\right]$$

$$\text{var}(U) = \left[\frac{420}{1640}\right]\left[\frac{68880}{12} - \sum 0\right]$$

$$\text{var}(U) = 1470$$

As a result, it is possible to determine the statistic u from the formula below.

$$u = (U - \overline{U}) / \left[\text{var}(U)\right]^{1/2}$$

$$u = (170 - 210) / \left[1470\right]^{1/2}$$

$$u = (170 - 210) / \left[1470\right]^{1/2} = -1.04$$

Since $|u| = |-1.04| = 1.04$ is less than the critical value $u_{0.025} = 1.96$,the Bahir Dar maximum annual daily rainfall data can be considered to be homogeneous and stationary at 5% level of significance.

4.3.1.3. Test for Outliers

Using the G-B test we, can determine the lower and upper outliers limits as shown below.

Table 9: Testing for outliers

Years	X=Max daily annual rainfall(mm)	LNX	Mean	LnX-Mean	(LnX-Mean)2
1975	64.3	4.16	4.24	-0.08	0.01
1976	73.4	4.30	4.24	0.05	0.00
1977	133.2	4.89	4.24	0.65	0.42

1978	61.1	4.11	4.24	-0.13	0.02
1979	56.2	4.03	4.24	-0.21	0.05
1980	79.9	4.38	4.24	0.14	0.02
1981	108.7	4.69	4.24	0.45	0.20
1982	35.9	3.58	4.24	-0.66	0.44
1983	64.3	4.16	4.24	-0.08	0.01
1984	70.4	4.25	4.24	0.01	0.00
1985	83.3	4.42	4.24	0.18	0.03
1986	58.3	4.07	4.24	-0.18	0.03
1987	47.4	3.86	4.24	-0.38	0.15
1988	57.1	4.04	4.24	-0.20	0.04
1989	61.1	4.11	4.24	-0.13	0.02
1990	114.1	4.74	4.24	0.49	0.25
1991	90.5	4.51	4.24	0.26	0.07
1992	53.8	3.99	4.24	-0.26	0.07
1993	58.3	4.07	4.24	-0.18	0.03
1994	39	3.66	4.24	-0.58	0.33
1995	58.6	4.07	4.24	-0.17	0.03
1996	51.1	3.93	4.24	-0.31	0.10
1997	69.2	4.24	4.24	-0.01	0.00
1998	51.5	3.94	4.24	-0.30	0.09
1999	64.6	4.17	4.24	-0.07	0.01
2000	105.3	4.66	4.24	0.41	0.17
2001	124.7	4.83	4.24	0.58	0.34
2002	89.5	4.49	4.24	0.25	0.06
2003	76.9	4.34	4.24	0.10	0.01
2004	95.5	4.56	4.24	0.32	0.10
2005	58.64	4.07	4.24	-0.17	0.03
2006	85.7	4.45	4.24	0.21	0.04
2007	74.1	4.31	4.24	0.06	0.00
2008	68.3	4.22	4.24	-0.02	0.00
2009	64.5	4.17	4.24	-0.08	0.01
2010	58.1	4.06	4.24	-0.18	0.03
2011	72.8	4.29	4.24	0.05	0.00
2012	96.7	4.57	4.24	0.33	0.11
2013	86.4	4.46	4.24	0.22	0.05
2014	48.2	3.88	4.24	-0.37	0.13
2015	66.7	4.20	4.24	-0.04	0.00
	Sum total	173.93		Total	3.48
	Mean	4.24		St.dev-S	0.30

$$\bar{x} = 4.24$$

and the standard deviation is $s = 0.3$

The k_N value can be calculated as below

$$k_N = -3.62201 + 6.28446N^{1/4} - 2.49835N^{1/2} + 0.491436N^{3/4} - 0.037911N$$

and where N=41 in our case as it is possible to check from above.

$$k_N = -3.62201 + 6.28446 * 41^{1/4} - 2.49835 * 41^{1/2} + 0.491436 * 41^{3/4} - 0.037911 * 41$$

$$k_N = 2.69$$

Now, it is quite possible to calculate the upper and low outliers as below.

1. Upper outlier limit

$$x_H = \exp(\bar{x} + k_N s)$$

$$x_H = \exp(4.24 + 2.69 * 0.3)$$

$$x_H = \exp(5.047)$$

$$x_H = 155.56mm$$

2. Lower outlier limit

$$x_L = \exp(\bar{x} - k_N s)$$

$$x_L = \exp(4.24 - 2.69 * 0.3)$$

$$x_L = \exp(3.433)$$

$$x_L = 30.97mm$$

Therefore, from above analysis of the outliers (upper or lower),there is no any single data that falls under the range of the outliers as we can observe from the maximum daily annual rainfall data presented above. Hence, there is **no any annual maximum daily data with outliers.**

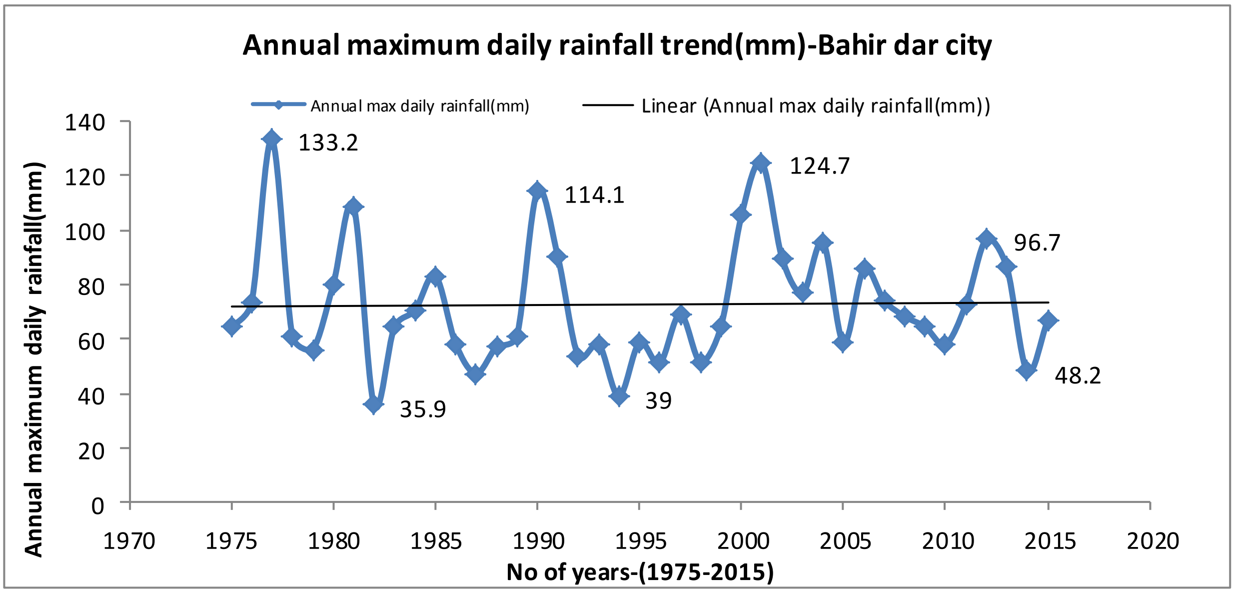


Figure 4:Annual maximum daily rainfall distribution trend to Bahir dar city

a. Identifying best fit distribution functions using easy fit software

Using the easy fit software, it was possible to determine the best fitting probability density function from the above mentioned distribution functions. Hence, the generalized extreme value function was selected as the best probability density function that works for Bahir dar maximum annual daily rainfall data.

i. Goodness of Fit - Summary

Table 10:Goodness of fit summary

#	Distribution	Kolmogorov Smirnov		Anderson Darling		Chi-Squared	
		Statistic	Rank	Statistic	Rank	Statistic	Rank
1	Exponential	0.4306	9	9.7407	9	65.25	9
2	Exponential (2P)	0.25397	8	3.9098	8	6.0802	7
3	Gen. Extreme Value	0.0729	1	0.21605	1	0.56342	1
4	Gumbel Max	0.07757	5	0.23894	3	0.84495	5
5	Log-Pearson 3	0.07584	2	0.24384	4	0.83869	4
6	Lognormal	0.08783	6	0.29759	6	2.3557	6
7	Lognormal (3P)	0.07642	4	0.24735	5	0.81798	2
8	Normal	0.132	7	0.94985	7	8.7965	8
9	Pearson 5 (3P)	0.07624	3	0.23861	2	0.82737	3

Using easy fit software analysis, it was quite possible to identify the best fit distribution function for Bahir dar city annual maximum daily rainfall data. Therefore, as it was mentioned or listed in the table above, one can easily identify that the best fitting distribution function to the study area annual maximum daily rainfall data is generalized extreme value distribution function. Using all the formulae or methods of identifying the best fitting distribution function which includes Kolmogorov Smirnov, Anderson Darling, and Chi-Squared test methods.

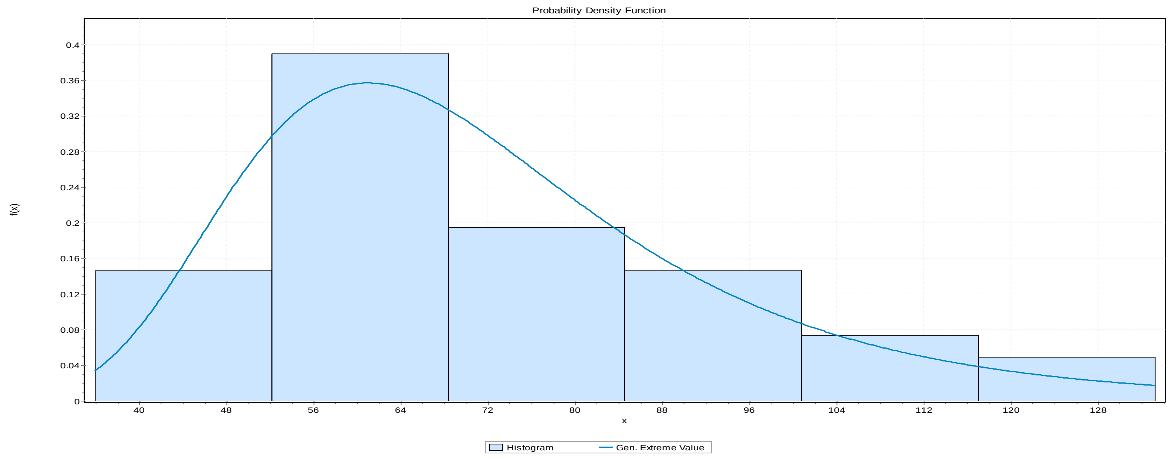


Figure 5:Probablity distribution function to the annual maximum daily rainfall data

The probability density function of the annual maximum daily rainfall data was also shown above in order to show the distribution of the rainfall data looks like. In addition, the cumulative distribution function also shown below as well.

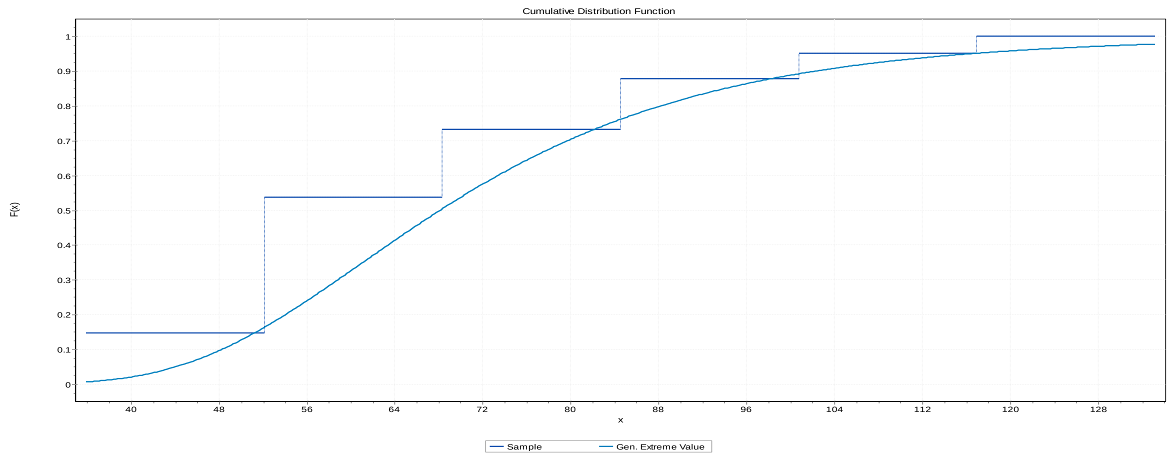


Figure 6: Cumulative distribution function to the annual maximum daily rainfall data

b. Determination of parameters of the best fit distribution function(GEV)

From the analysis report of the easy fit software, one can determine the parameters of the generalized extreme value(GEV) distribution function. The location, shape and scale parameters of the selected distribution, which is the generalized extreme value distribution function was determined from the easy fit software output. Hence, the results were shown here below.

- The shape parameter(k), $k = 0.05876$

- The scale parameter(σ), $\sigma = 16.748$
- The location parameter(μ), $\mu = 61.921$

From the easy fit software analysis, we can determine the following values of the maximum annual daily rainfall data. The mean, variance, standard deviation, coefficient of variation, skew ness and kurtosis are determined.

- Mean =72.618
- Variance=545.41
- Standard deviation=23.354
- Coefficient of variation=0.3216
- Skewness=-1.5414
- Kurtosis=4.8067

Fitting Result

Table 11:Annual maximum daily rainfall data fitting

#	Distribution	Parameters
1	Gen. Extreme Value	k=0.05876 σ=16.748 μ=61.921------selected

Table 12:Details of general extreme value distribution function

Gen. Extreme Value [#3]					
Kolmogorov-Smirnov					
Sample Size Statistic P-Value Rank	41 0.0729 0.97019 1				
□	0.2	0.1	0.05	0.02	0.01
Critical Value	0.16349	0.18687	0.2076	0.23213	0.24904
Reject?	No	No	No	No	No
Anderson-Darling					
Sample Size Statistic Rank	41 0.21605 1				
□	0.2	0.1	0.05	0.02	0.01

Critical Value	1.3749	1.9286	2.5018	3.2892	3.9074
Reject?	No	No	No	No	No
Chi-Squared					
Deg. of freedom Statistic P-Value Rank	5 0.56342 0.98962 1				
□	0.2	0.1	0.05	0.02	0.01
Critical Value	7.2893	9.2364	11.07	13.388	15.086
Reject?	No	No	No	No	No

4.4. Simple scaling and time scale invariant characteristics

4.4.1. Scale invariance analysis for annual maximum daily rainfall data for $1hr \le d \le 24hr$ duration

Using the sub hourly rainfall data collected for the last 4 years, we can be able to determine the scaling factor exponent. It helps to make generation of the rainfall intensity in the sub hourly duration. From the analysis of the data, the linearity of the graphs behaved in two ways of durations $15\min \le d < 1hr$ and $1hr \le d \le 24hr$.

The rainfall data extracted from the duration of 2012-2015 years was shown below here.

Table 13: Annual maximum sub hourly and hourly calculation

Year	15min(mm)	30min(mm)	45min(mm)	60min(mm)	120min-RF(mm)	180min-RF(mm)	24hr(mm)
2012	28.6	39.4	46.8	58	87.4	93.6	128.8
2013	35	54.2	59.2	61.6	66.6	67	88
2014	23	30.6	36.8	41.6	46.2	46.6	82.2
2015	20.2	31.2	38	41.8	55.2	58	72.8
Avg	**26.70**	**38.85**	**45.20**	**50.75**	**63.85**	**66.30**	**92.95**

The analysis was made based on two categories of duration for the annual maximum daily rainfall data

Table 14: Moment calculation from sub hourly rainfall data

Year	**15min(mm)**	I(mm/hr)	I*I	I*I*i*		Year	**30min(mm)**	I(mm/hr)	I*I	I*I*i*
2012	28.60	190 .67	36353.78	6931453.63		2012.00	39.40	131.33	17248.44	2265295.70
2013	35.00	233.33	54444.44	12703703.70		2013.00	54.20	180.67	32640.44	5897040.30
2014	23.00	153.33	23511.11	3605037.04		2014.00	30.60	102.00	10404.00	1061208.00
2015	20.20	134.67	18135.11	2442194.96		2015.00	31.20	104.00	10816.00	1124864.00
	1st Moment	178.00					1st Moment	129.50		
	2nd Moment		33111.11				2nd Moment		17777.22	
	3rd Moment			6420597.33			3rd Moment			2587102.00
Year	**45min(mm)**	I(mm/hr)	I*I	I*I*i*		Year	**60min(mm)**	I(mm/hr)	I*I	I*I*i*
2012	46.80	104z.00	10816.00	1124864.00		2012.00	58.00	58.00	3364.00	195112.00
2013	59.20	131.56	17306.86	2276814.13		2013.00	61.60	61.60	3794.56	233744.90
2014	36.80	81.78	6687.60	546897.47		2014.00	41.60	41.60	1730.56	71991.30
2015	38.00	84.44	7130.86	602161.87		2015.00	41.80	41.80	1747.24	73034.63
	1st Moment	100.44					1st Moment	50.75		
	2nd Moment		10485.33				2nd Moment		2659.09	
	3rd Moment			1137684.37			3rd Moment			143470.71
Year	**120min(mm)**	I(mm/hr)	I*I	I*I*i*		Year	**180min(mm)**	I(mm/hr)	I*I	I*I*i*
2012	87.40	43.70	1909.69	83453.45		2012.00	93.60	31.20	973.44	30371.33
2013	66.60	33.30	1108.89	36926.04		2013.00	67.00	22.33	498.78	11139.37
2014	46.20	23.10	533.61	12326.39		2014.00	46.60	15.53	241.28	3747.95
2015	55.20	27.60	761.76	21024.58		2015.00	58.00	19.33	373.78	7226.37
	1st Moment	31.93					1st Moment	22.10		
	2nd Moment		1078.49				2nd Moment		521.82	
	3rd Moment			38432.61			3rd Moment			13121.26

Table 15: The log transformed values of the moments of the intensities and also the duration.

	Intensity log(I) values		
log(d)	1st Moment	2nd Moment	3rd Moment
-1.39	5.18	10.41	15.68
-0.69	4.86	9.79	14.77
-0.29	4.61	9.26	13.94
0.00	3.93	7.89	11.87
0.69	3.46	6.98	10.56
1.10	3.10	6.26	9.48
3.18	1.35	2.76	4.22

From the above table, one can generate the graph shown below in order to understand the scale invariance property of the rainfall data at hourly and sub hour durations.

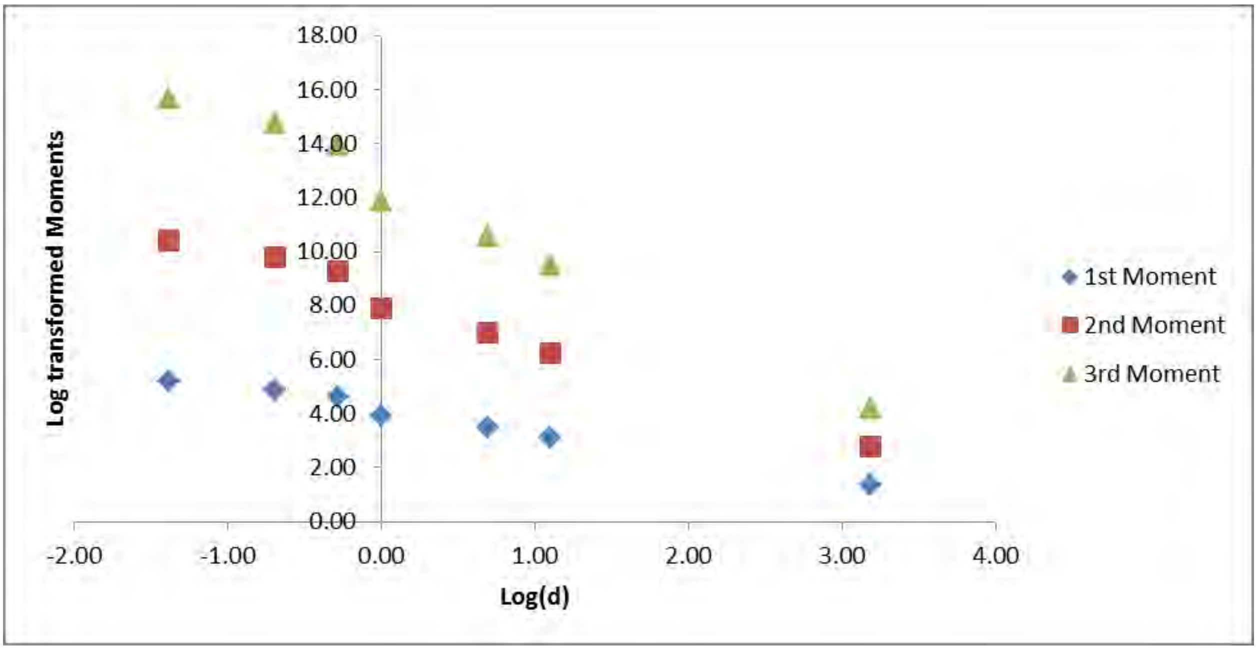


Figure 7:Time scale invariance of Bahir dar city rainfall data

From the above graph, one can easily detect that there is a distinction of the linearity of the slope for two duration of a) $1hr \le d \le 24hr$ and b) $15\min \le d < 1hr$ exactly at duration of 1hour.Hence,we treat the scaling in two different durations in order to calculate the scale invariance factor as shown below.

a)duration of $1hr \le d \le 24hr$

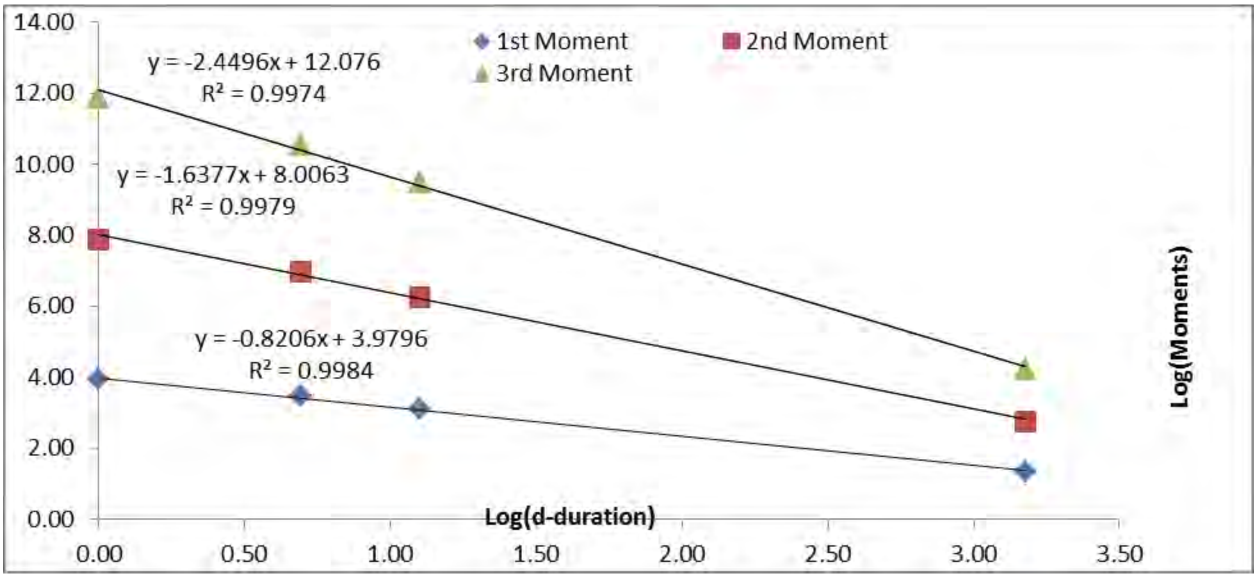


Figure 8:Scale invariant factors calculation 1

b) duration of $15\min \le d < 1hr$

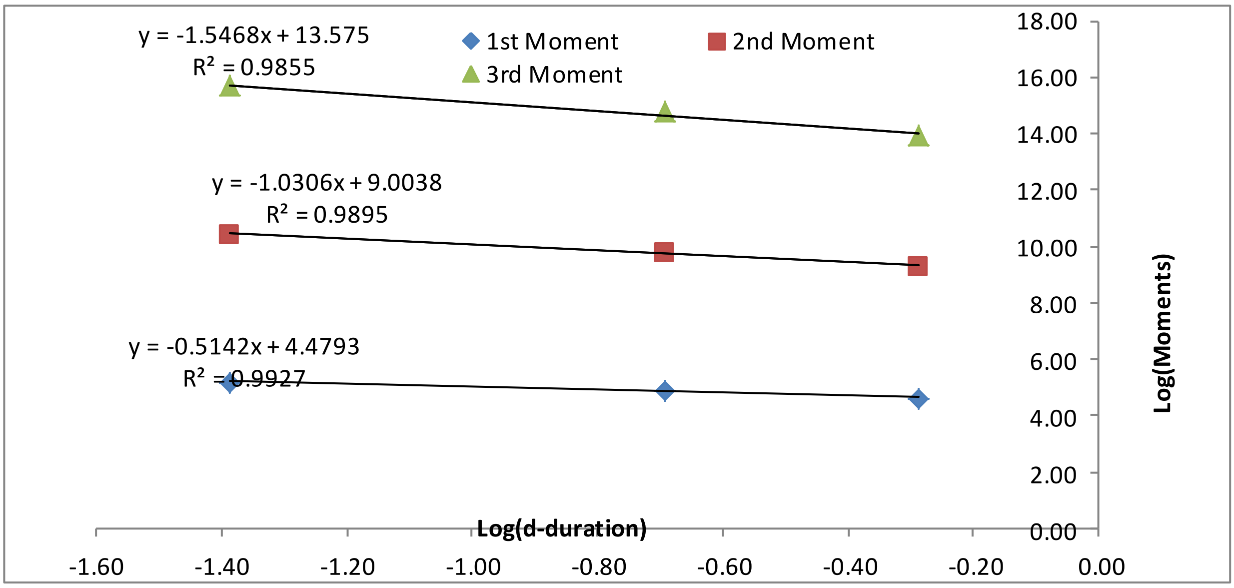


Figure 9: Scale invariant factors calculation 2

The slopes of the best fit straight lines, Hq, are then be plotted with the moment order of 3.From the plot below, it is quite easy to understand that the relationship of the Hq and q is straight lines for two different durations with the R square value exactly 1.This justifies that the simple scale holds true. The slopes obtained by using linear regression, are the values of the scale exponent H for the two durations, a) $15\min \le d < 1hr$ and b) $1hr \le d \le 24hr$ with it‟s the values of scaling exponents of γ as 0.5163 and 0.8145 respectively.

4.4.2. Estimating the IDF relationship for Bahir dar city

Using the easy fit software, it was possible to identify the best fit probability distribution function what works for the annual maximum daily rainfall recorded at Bahirdar city.

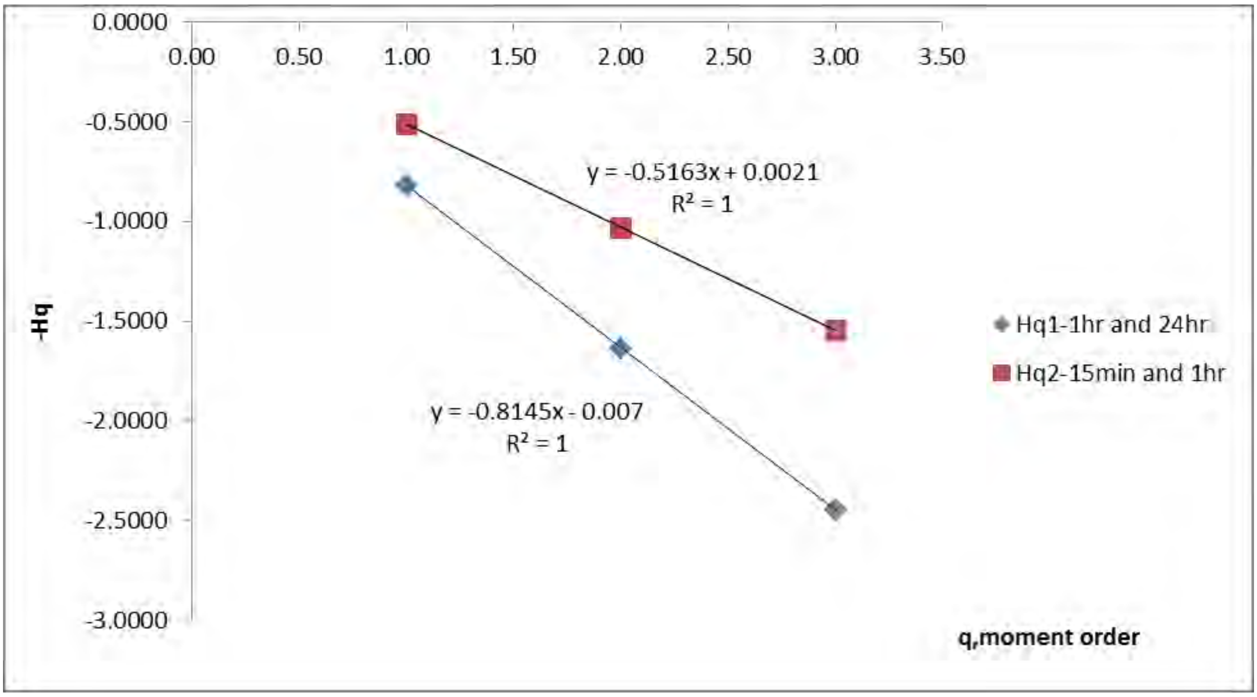


Figure 10:Determination of the Scale invariant factors exponents

The generalized extreme value distribution function is of the form in equation below.

$$F(x) = \exp\left\{-\left[1 + k\left(\frac{x-u}{\alpha}\right)\right]^{-1/k}\right\} \qquad 4.4.2.1$$

for $1 + \frac{k(x-\mu)}{\sigma} > 0$,where $\mu \in \mathbb{R}$ is the location parameter, $\sigma > 0$ the scale parameter and $k \in \mathbb{R}$ the shape parameter.

The density function is, consequently

$$f(x;\mu,\sigma,k) = \frac{1}{\sigma}\left[1 + k\left(\frac{x-\mu}{\sigma}\right)\right]^{-1/(k-1)} e^{\left\{-\left[1+k\left(\frac{x-\mu}{\sigma}\right)\right]^{-1/k}\right\}} \qquad 4.4.2.2$$

again, for $1 + \frac{k(x-\mu)}{\sigma} > 0$, From the concept of $i_{d,T} = \frac{a(T)}{b(d)}$,we can determine the IDF curve relationship from the inverse of the cumulative distribution function and duration relations.

The inverse of the generalized extreme value function is given below as:

$$F^{-1}(1-1/T) = \mu + \frac{\alpha}{k}(1-(-\ln(1-1/T))^{k}) \quad 4.4.2.3$$

From the output of the easy fit software analysis, the parameters were determined as shown below.

The shape parameter(k), $k = 0.05876$

The scale parameter(σ), $\sigma = 16.748$

The location parameter(μ), $\mu = 61.921$

Hence for the duration of $1hr \le d \le 24hr$,we generated a relationship between the return period and durations. The IDF relationship for this period is here below. $\gamma = 0.8145$

$$i_{d,T} = \frac{61.921 + 285.024*(1-(-\ln(1-1/T))^{0.05876})}{d^{0.8145}}$$

In the same approach, we can also determine the relationship for the duration of $15\min \le d < 1hr$ .Again, using the location and scale parameters computed above and using the scale invariance factor, we can determine the IDF relationship.

Using the same scale and location factors, we can generate the IDF relationship for the duration of $15\min \le d < 1hr$ .Hence, it is shown here below. But in this case, we use only the scale invariance factor of $\gamma = 0.5163$.

$$i_{d,T} = \frac{61.921 + 285.024*(1-(-\ln(1-1/T))^{0.05876})}{d^{0.5163}}$$

Similarly, for the duration of $15\min \le d < 1hr$,we generated a relationship between the return period and durations. The IDF relationship for this period is here below.

$$i_{d,T} = \frac{61.921 + 285.024*(1-(-\ln(1-1/T))^{0.05876})}{d^{0.5163}}$$

Table 16: Intensity Duration Frequency Values-for Bahir dar city

	Intensity(mm/hr)					
Duration(min)	2-years	5-years	10-years	25-years	50-years	100-years
15	139.096	175.86	198.894	226.576	246.14	264.776
30	97.25	122.957	139.059	158.414	172.09	185.12
45	78.88	99.733	112.793	128.49	139.586	150.155
60	67.994	85.967	97.225	110.757	120.32	129.43
120	38.66	48.88	55.283	62.977	68.41	73.59
180	27.79	35.133	39.734	45.264	49.17	52.896

5.3.4.Development of IDF curves by using simple scaling method to Bahir dar city

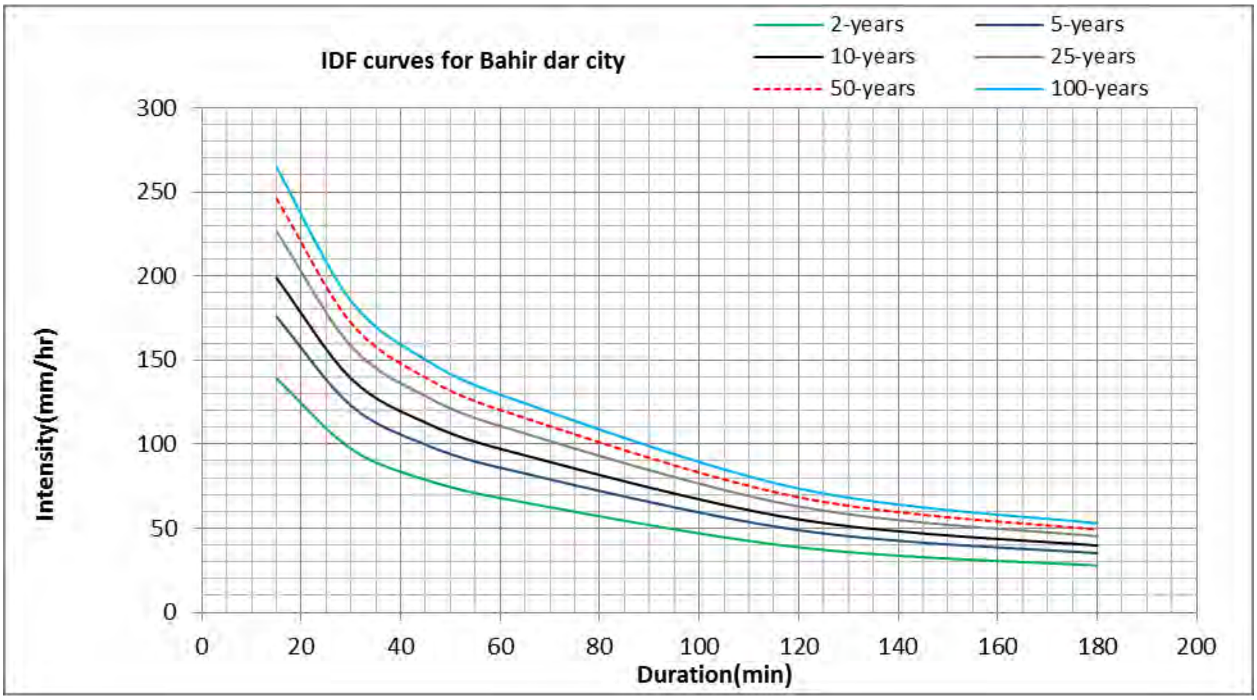
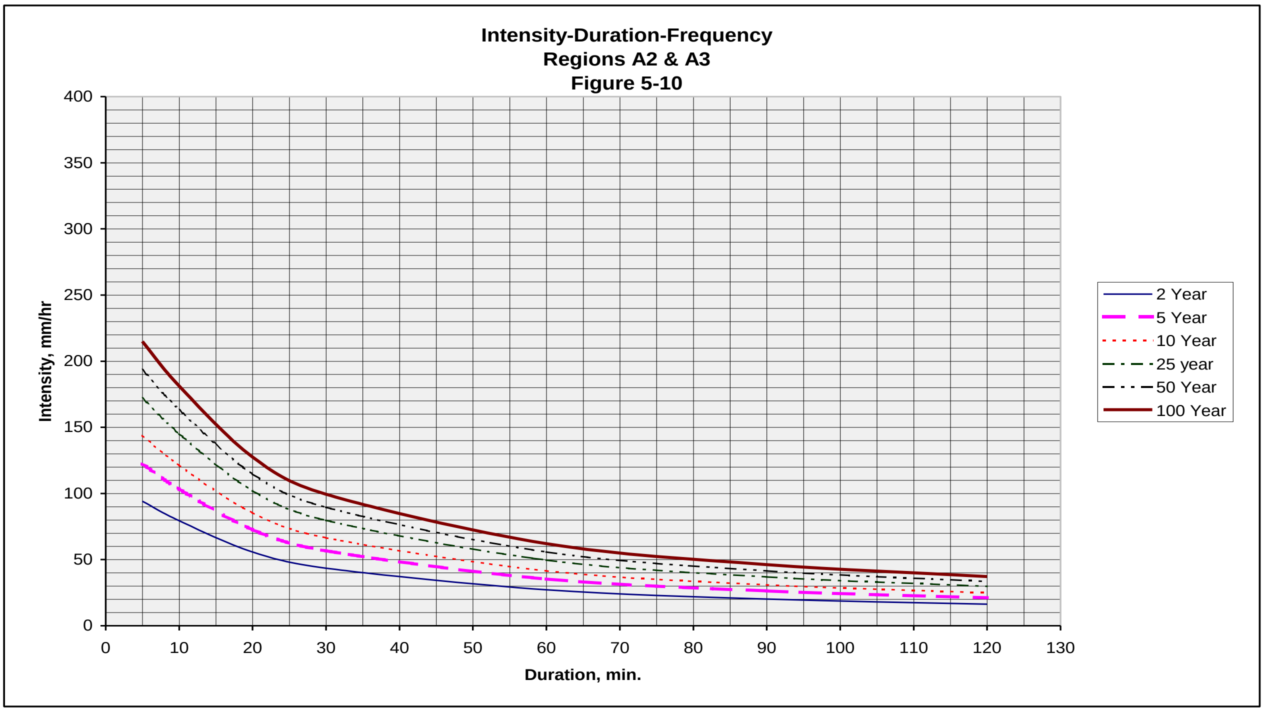


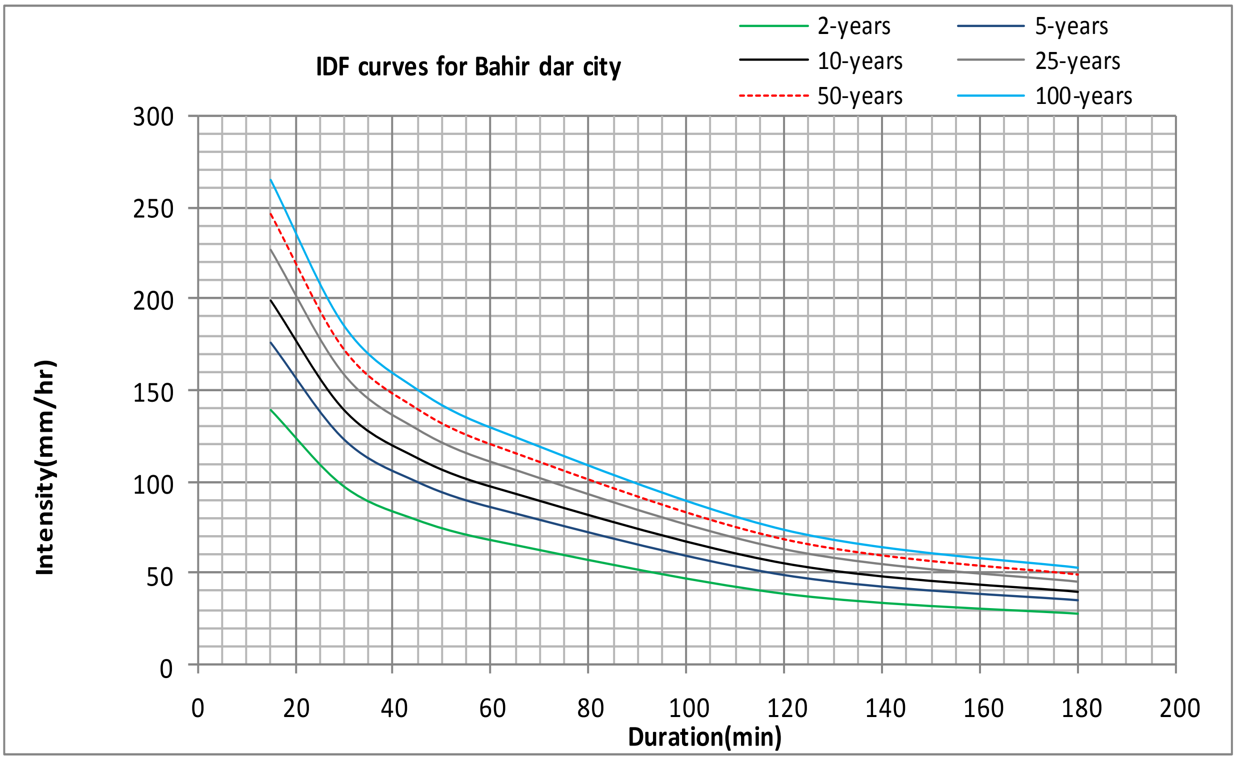
Figure 11:IDF curves using simple scaling

Regional IDF curve

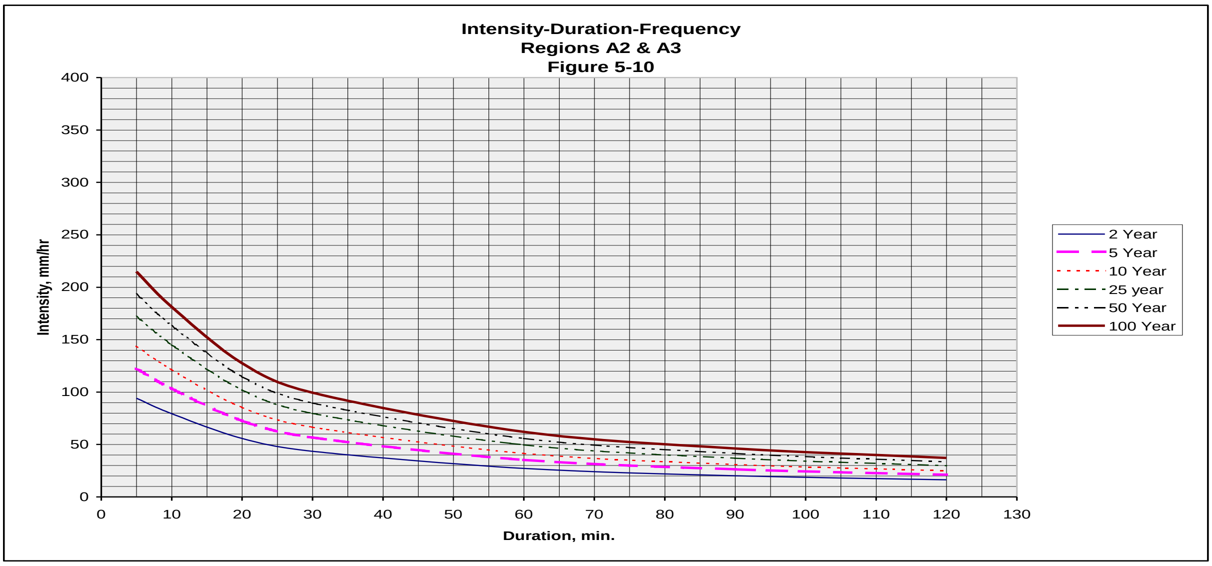


4.4.3. **Comparison of Simple scaling and** Regional IDF curves for Bahir dar city

- **IDF curves using simple scaling method**



- Regional IDF curves for Bahir dar city



From the above regional and simple scaling method developed IDF curves for Bahir dar city, the relationship for both curves were developed using the logarithmic fitting analysis .Accordingly, the analysis of the logarithmic correlation for the specific return period, the multiplier factor for relating the intensities developed from the simple scaling method and regional method determined.

Table 17:Intensities for simple scaling method to Bahir dar city

Duration		Intensity(mm/hr)					
Min	Hrs	2-years	5-years	10-years	25-years	50-years	100-years
15	0.25	139.096	175.86	198.894	226.576	246.14	264.776
30	0.5	97.25	122.957	139.059	158.414	172.09	185.12
45	0.75	78.88	99.733	112.793	128.49	139.586	150.155
60	1	67.994	85.967	97.225	110.757	120.32	129.43
120	2	38.66	48.88	55.283	62.977	68.41	73.59
180	3	27.79	35.133	39.734	45.264	49.17	52.896

Table 18: Intensities from Regional IDF curves for Bahir dar city

Duration		Intensity(mm/hr)					
Min	Hrs	2-years	5-years	10-years	25-years	50-years	100-years
15	0.25	67.00	87.00	100.00	120.00	178.00	180.00
30	0.5	43.00	57.00	75.00	89.00	100.00	110.00
45	0.75	35.00	45.00	50.00	62.00	75.00	90.00
60	1	28.00	35.00	40.00	50.00	57.00	62.00
120	2	18.00	20.00	23.00	30.00	37.00	40.00

4.4.3.A. For return period, T=2 years, the logarithmic data fitting was shown below. Let

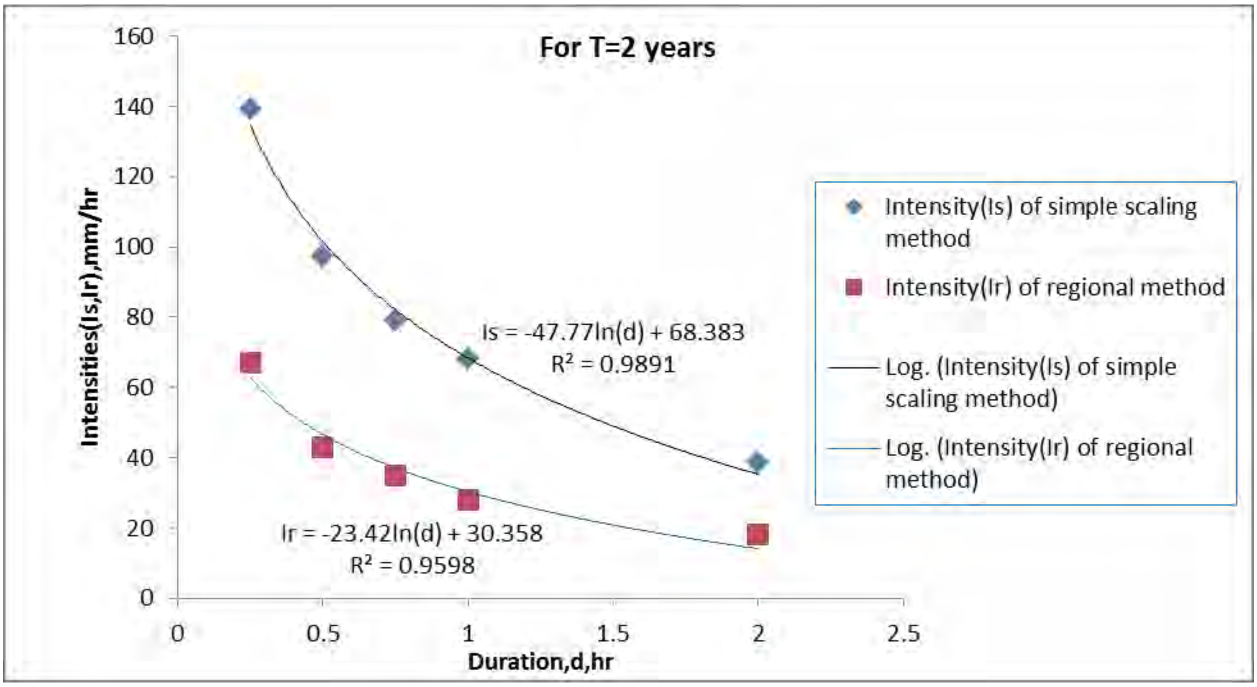


Figure 12: Logarithmic data fitting for 2 years

I_s be intensities from simple scaling method and I_d ,hence, the relationship can be developed between the regional and simple scaling method developed intensities for Bahir dar city.

From above graphs, the logarithmic correlation of the intensities a s function of duration in hours, one can developed the following relationships.

$$\frac{I_s}{I_r}=\frac{-47.77\ln(d)+68.383}{-23.42\ln(d)+30.358}$$

$$I_s=\left(\frac{-47.77\ln(d)+68.383}{-23.42\ln(d)+30.358}\right)*I_r$$

4.4.3.B. For return period T=5 years, the same procedure followed as well.

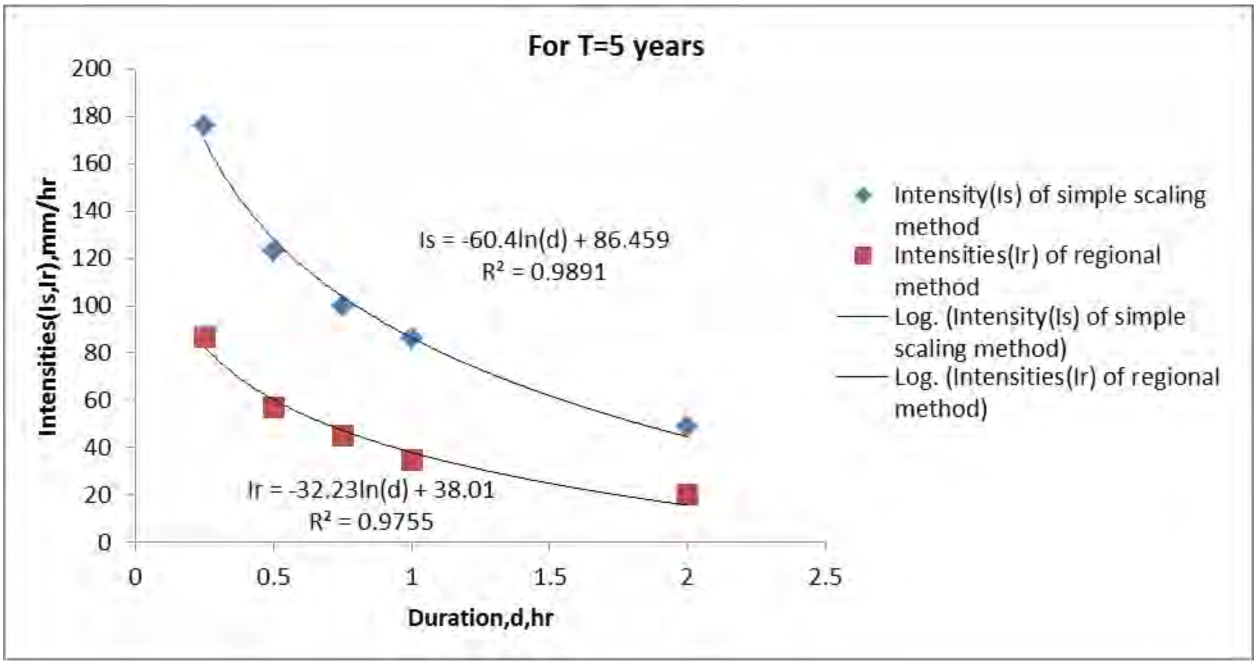


Figure 13: Logarithmic data fitting for 5 years

$$\frac{I_s}{I_r}=\frac{-60.4\ln(d)+86.459}{-32.23\ln(d)+38.01}$$

This implies that:

$$I_s=\left(\frac{-60.4\ln(d)+86.459}{-32.23\ln(d)+38.01}\right)*I_r$$

4.4.3.C. For return period of T=10 years, the relationship between the regional and simple scaling method developed intensities can developed as follows.

$$\frac{I_s}{I_r}=\frac{-68.31\ln(d)+97.783}{-38.52\ln(d)+44.705}$$

$$I_s = \left(\frac{-68.31\ln(d) + 97.783}{-38.52\ln(d) + 44.705} \right) * I_r$$

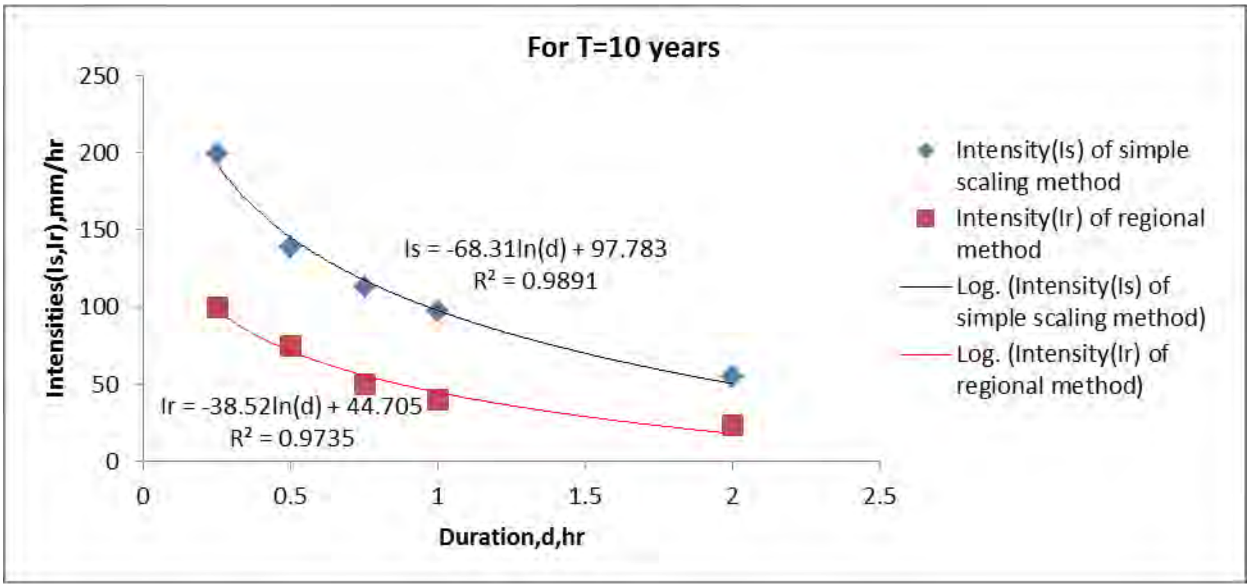


Figure 14: Logarithmic data fitting for 10 years

4.4.3.D. For return period of T=25 years, the relationship between the regional and simple scaling method developed intensities can developed as follows.

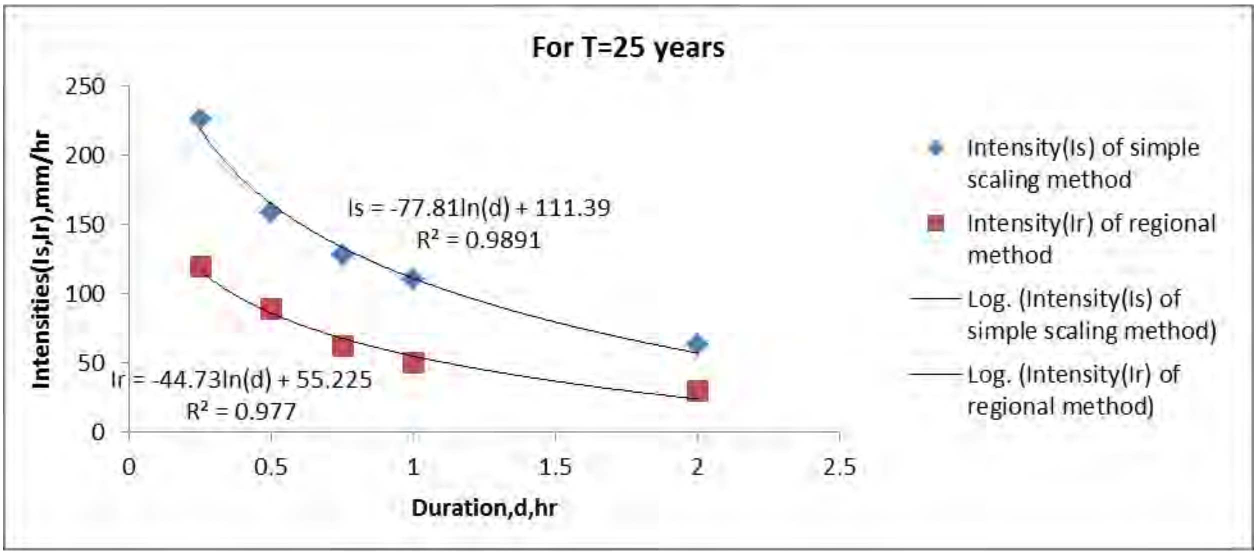


Figure 15: Logarithmic data fitting for 25 years

$$\frac{I_s}{I_r} = \frac{-77.81\ln(d) + 111.39}{-44.73\ln(d) + 55.225}$$

$$I_s = \left(\frac{-77.81\ln(d) + 111.39}{-44.73\ln(d) + 55.225} \right) * I_r$$

4.4.3.E. For return period of T=50 years, the relationship between the regional and simple scaling method developed intensities can developed as follows.

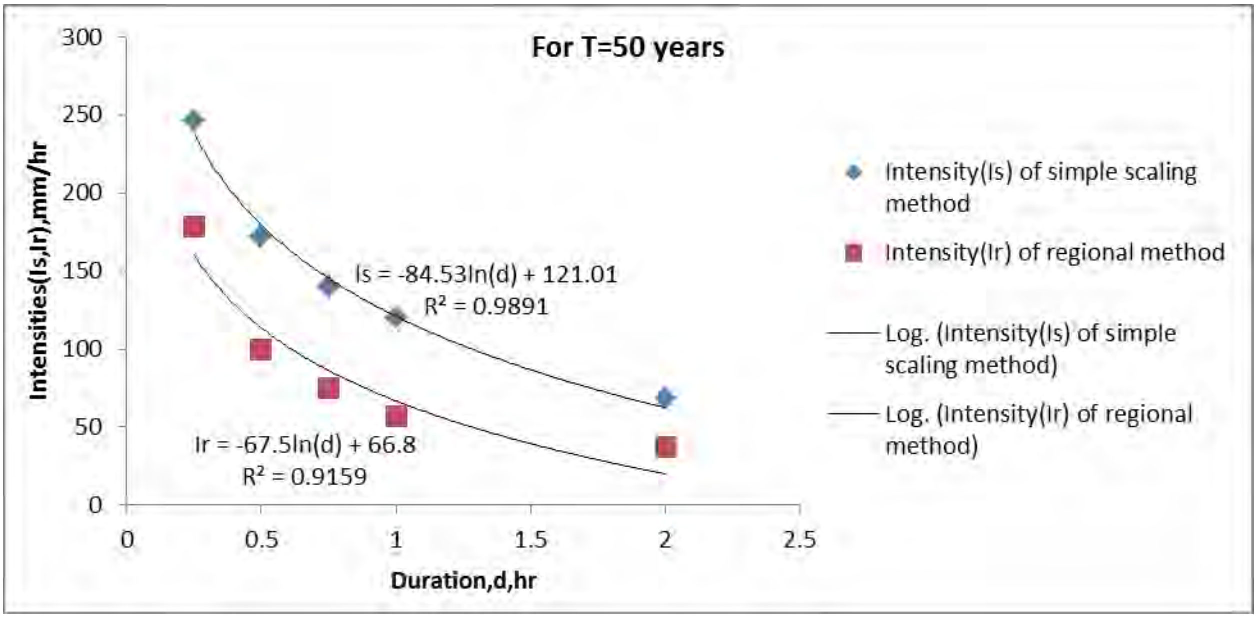


Figure 16: Logarithmic data fitting for 50 years

$$\frac{I_s}{I_r} = \frac{-84.531\ln(d) + 121.01}{-67.5\ln(d) + 66.8}$$

$$I_s = \left(\frac{-84.531\ln(d) + 121.01}{-67.5\ln(d) + 66.8}\right) * I_r$$

4.4.3.F. For return period of T=100 years, the relationship between the regional and simple scaling method developed intensities can developed as follows.

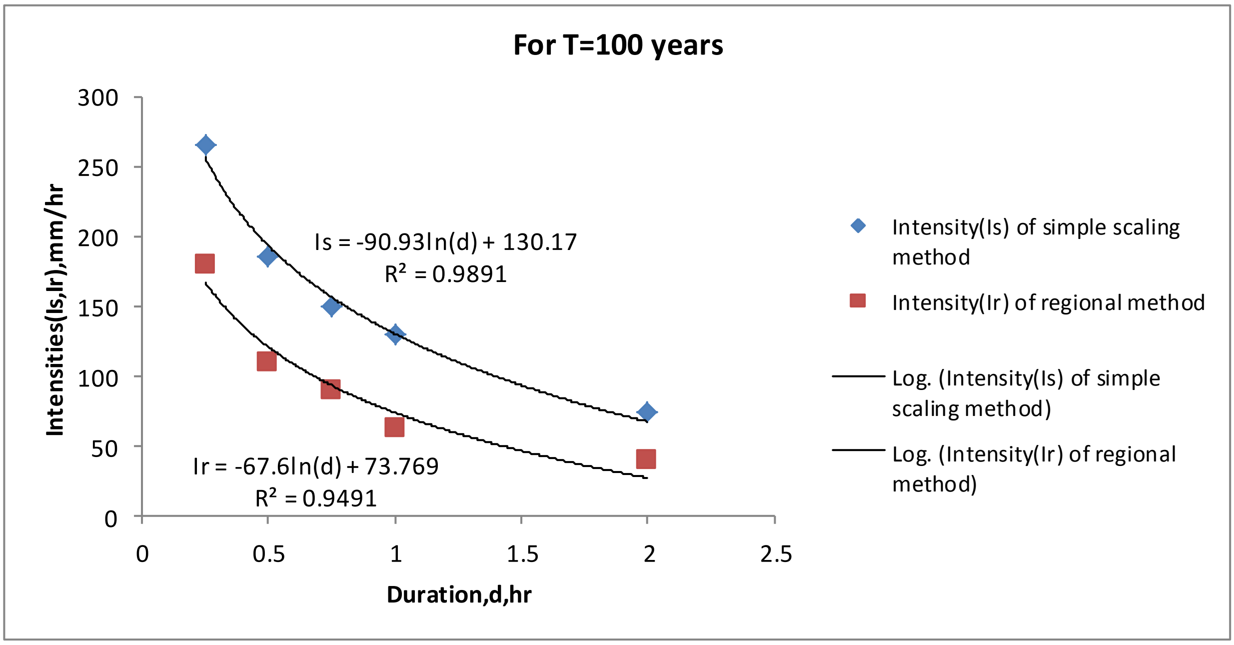


Figure 17: Logarithmic data fitting for 100 years

$$\frac{I_s}{I_r} = \frac{-90.93\ln(d)+130.17}{-67.6\ln(d)+73.769}$$

$$I_s = \left(\frac{-90.93\ln(d)+130.17}{-67.6\ln(d)+73.769}\right) * I_r$$

In order to summarize the above analysis, it is shown below in a tabular way so that one can easily generate intensities comparison data based on the already generated IDF curves and also the formulae below.

Table 19:Summary of regional and simple scaling methods developed intensities and durations

SN	Return period, T (Years)	Intensities (Is, Ir) relationship	Remarks
1	For T=2 years	$I_s = \left(\frac{-47.77\ln(d)+68.383}{-23.42\ln(d)+30.358}\right) * I_r$	Intensities(Is, Ir) should be taken/read from respective IDF curves.
2	For T=5 years	$I_s = \left(\frac{-60.4\ln(d)+86.459}{-32.23\ln(d)+38.01}\right) * I_r$	Intensities(Is, Ir) should be taken/read from respective IDF curves.

3	For T=10 years	$I_s = \left(\frac{-68.31\ln(d)+97.783}{-38.52\ln(d)+44.705}\right)*I_r$	Intensities(Is, Ir) should be taken/read from respective IDF curves.
4	For T=25 years	$I_s = \left(\frac{-77.81\ln(d)+111.39}{-44.73\ln(d)+55.225}\right)*I_r$	Intensities(Is, Ir) should be taken/read from respective IDF curves.
5	For T=50 years	$I_s = \left(\frac{-84.531\ln(d)+121.01}{-67.5\ln(d)+66.8}\right)*I_r$	Intensities(Is, Ir) should be taken/read from respective IDF curves.
6	For T=100 years	$I_s = \left(\frac{-90.93\ln(d)+130.17}{-67.6\ln(d)+73.769}\right)*I_r$	Intensities(Is, Ir) should be taken/read from respective IDF curves.

Note: Duration, d and Intensities(Is, Ir) are in hour and mm/hour respectively.

4.5. Comparison of the regional and simple scaling method IDF curves generated rainfall or intensities data with the observed rainfalls or intensities data.

The rainfall data obtained from the IDF curves developed using the simple scaling method falls above the values of the observed rainfall data or the intensities from IDF curves using simple scaling method falls beyond the values of the observed data. However, the rainfall or intensities from the regional IDF curves were falling below the observed data of Bahir dar city in the sub hourly recording comparison. This showed that the regional IDF curve underestimates the values of the rainfalls or intensities for the Bahir dar city. This will have a consequence or an impact on the under designing of drainage structures in the city, which consequently leads to road side flooding and also damaging of civil works by flooding. But the simple scaling method developed IDF curves have an added advantage in accommodating the flooding generated from the city as it is possible to accommodate the flood generated from the catchment areas. This is mainly because of the values of the rainfalls or intensities using simple scaling generated IDF curves have greater values than the observed values. This difference between the observed and simple scaling method generated IDF curve values of rainfalls or intensity can serve as a factor of safety in designing drainage structures in the city.

The Hazen method of estimating the probability of exceedence was used for determining the respective probabilities. The formula is $\frac{r-0.5}{n}*100$.r is the rank and n is the number of observation, which is 4 in our case.

Table 20: Generation of IDF curves using the observed 4 years rainfall data (2012-2015)

		Duration of 15minutes						Duration of 30minutes			
Rank	Year	Max RF(mm)	probability of exceedence (%)	T(years)	I(mm/hr)	Rank	Year	Max RF(mm)	probability of exceedence (%)	T(years)	I(mm/hr)
1	2012	35	12.5	8.00	140	1	2012	54.2	12.5	8.00	108.4
2	2013	28.6	37.5	2.67	114.4	2	2013	39.4	37.5	2.67	78.8
3	2014	23	62.5	1.60	92	3	2014	31.2	62.5	1.60	62.4
4	2015	20.2	87.5	1.14	80.8	4	2015	30.6	87.5	1.14	61.2
		Duration of 45minutes						Duration of 60minutes			
Rank	Year	Max RF(mm)	probability of exceedence (%)	T(years)	I(mm/hr)	Rank	Year	Max RF(mm)	probability of exceedence (%)	T(years)	I(mm/hr)
1	2012	59.2	12.5	8.00	78.93	1	2012	61.6	12.5	8.00	61.60
2	2013	46.8	37.5	2.67	62.40	2	2013	58	37.5	2.67	58.00
3	2014	38	62.5	1.60	50.67	3	2014	41.8	62.5	1.60	41.80
4	2015	36.8	87.5	1.14	49.07	4	2015	41.6	87.5	1.14	41.60
		Duration of 120minutes						Duration of 180minutes			
Rank	Year	Max RF(mm)	probability of exceedence (%)	T(years)	I(mm/hr)	Rank	Year	Max RF(mm)	probability of exceedence (%)	T(years)	I(mm/hr)
1	2012	87.4	12.5	8.00	43.70	1	2012	93.6	12.5	8.00	31.20
2	2013	66.6	37.5	2.67	33.30	2	2013	67	37.5	2.67	22.33
3	2014	55.2	62.5	1.60	27.60	3	2014	58	62.5	1.60	19.33
4	2015	46.2	87.5	1.14	23.10	4	2015	46.6	87.5	1.14	15.53

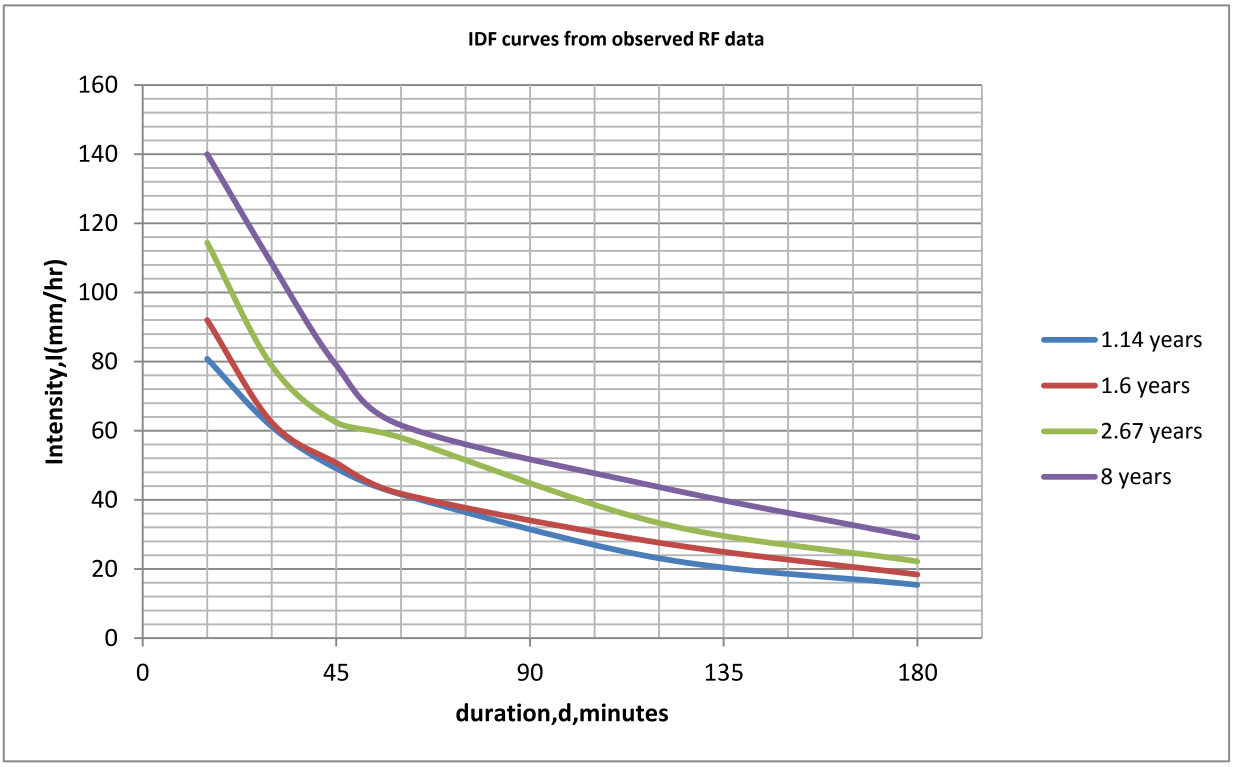


Figure 18:IDF curves from observed rainfall data

4.5.1. Intensity using the observed data

Table 21: Intensities calculated from observed data

Intensity(mm/hr)				
Time(min)	1.14 years	1.6 years	2.67 years	8 years
15	80.8	92	114.40	140
30	61.2	62.4	78.80	108.4
45	49.07	50.67	62.40	78.93
60	41.6	41.8	58.00	61.6
120	23.1	27.6	33.30	43.7
180	15.4	18.4	22.20	29.13

4.5.2. Intensities using simple scaling method

Table 22: Intensities calculated from simple scaling method

Intensity(mm/hr)				
Time(min)	1.14 years	1.6 years	2.67 years	8 years
15	100.74	127.29	151.97	191.66
30	70.45	89.05	106.32	134.08
45	57.13	72.21	86.21	108.73
60	49.24	62.25	74.32	93.72
120	27.98	35.37	42.22	53.25
180	20.1	25.41	30.33	38.25

4.5.3. Intensities from the full data Regional IDF curves

Table 23: Intensities observed from regional IDF curves

Duration		Intensity(mm/hr)					
Min	Hrs	2-years	5-years	10-years	25-years	50-years	100-years
15	0.25	67.00	87.00	100.00	120.00	178.00	180.00
30	0.5	43.00	57.00	75.00	89.00	100.00	110.00
45	0.75	35.00	45.00	50.00	62.00	75.00	90.00
60	1	28.00	35.00	40.00	50.00	57.00	62.00
120	2	18.00	20.00	23.00	30.00	37.00	40.00

4.5.4. Comparison of regional, observed and simple scaling developed IDF curves for selected return period (8-years return period).

Table 24:Intensities from regional, observed and simple scaling curves

8-year return period			
Duration(min)	Regional Intensities	Observed Intensities	Simple Scaling developed Intensities
15	94.8	140	191.66
30	67.8	108.4	134.08
45	48	78.93	108.73
60	38	61.6	93.72
120	21.8	43.7	53.25

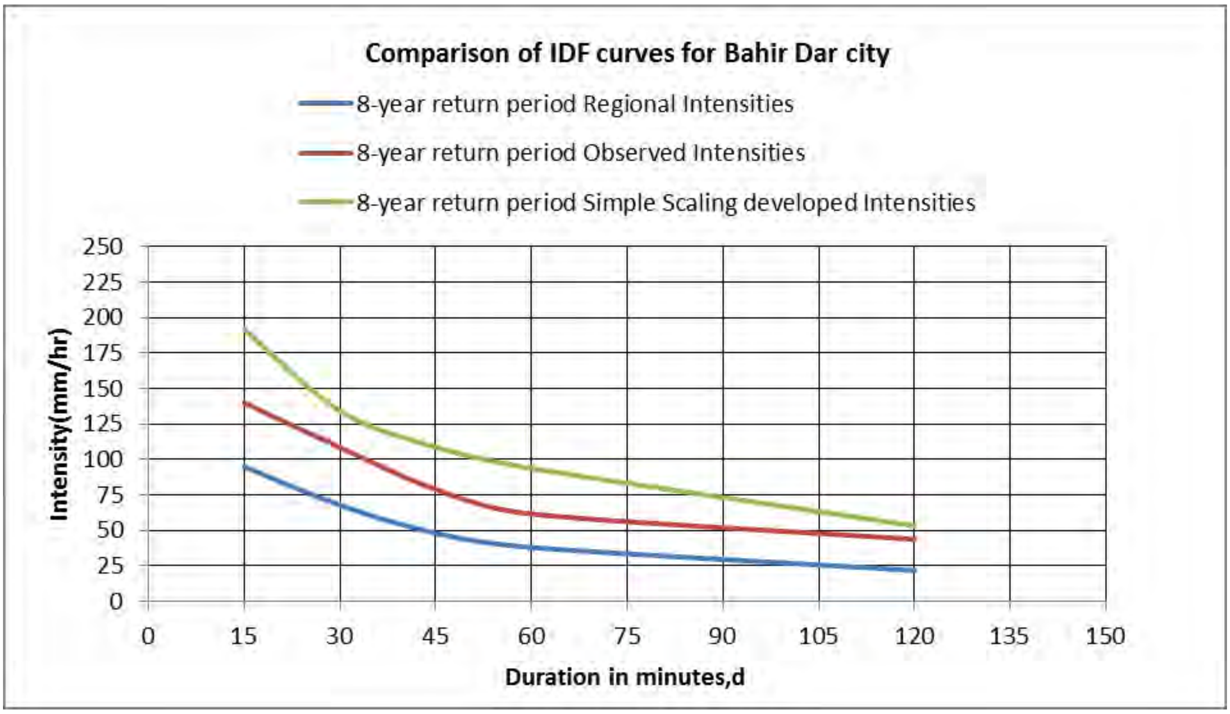


Figure 19: IDF curves from regional, observed and simple scaling rainfall data

4.6. Determination of the Design Rainfall using Frequency Factor

The Determination of the design rainfalls for 2,5,10,25,50 and 100 years of return periods were considered in order have permissible rainfall in any of the specified return periods. The design period calculation was followed through the use of frequency factor method. The formula for this application is shown here with below. Using the extreme value type I (EVI) distribution, we can determine the factors below.

$$x_T = \bar{x} + K_T s \,, \qquad 4.6.1$$

where x_T is the design rainfall to be calculated, $\bar{x}$ is the mean annual maximum daily rainfall, K_T is the frequency factor for extreme distributions, $K_T = -\frac{\sqrt{6}}{\pi}\left\{0.5772 + \ln\left[\ln\left(\frac{T}{T-1}\right)\right]\right\}$ and T -is the respective return period(Chow,1988).

In order to determine the design rainfalls of the annual maximum daily rainfall for Bahirdar city, we need to follow the following concept of application of the frequency factor.

The frequency factor(K_T) can be determined from the relationship shown below.

$$K_T = -\frac{\sqrt{6}}{\pi}\left\{0.5772 + \ln\left[\ln\left(\frac{T}{T-1}\right)\right]\right\} \qquad 4.6.2$$

The value of the sample mean($\bar{x} = 72.618$) and standard deviation ($s = 23.354$) were calculated from the output of the easy fit software result.

$$\bar{x} = 72.618$$

$$s = 23.354$$

$$x_T = \bar{x} + K_T s \ ,$$

Table 25:Summary of design rainfall results

SN	T(years)	KT	Mean	Standard deviation(S)	XT(mm)
1	2	-0.16427	72.618	23.354	68.78157
2	5	0.71946	72.618	23.354	89.42028
3	10	1.304569	72.618	23.354	103.0849
4	25	2.043854	72.618	23.354	120.3502
5	50	2.592299	72.618	23.354	133.1585
6	100	3.136694	72.618	23.354	145.8723

The design rainfall for the respective return period were calculated and shown above.

CHAPTER 5: CONCLUSIONS AND RECOMMENDATIONS

5.1. CONCLUSIONS

In order to undertake the development of intensity-duration-frequency(IDF) curves for Bahir dar city using simple scaling method, the 41(1975-2015) years of annual maximum daily rainfall data and 4 years (2012-2015) annual maximum sub daily rainfall data were used.

Missing data were observed in the annual maximum daily rainfall data. These accounted to 2.44% of the rainfall data, while 97.56% of the rainfall data were observed rainfall data. The missing data were filled from the analysis resulted from taking the average of every 12 months over 41 years period and filled in the missed 12 months of the data period. In the annual maximum hourly rainfall data, it was not easy to get automatically registered or recorded sub hour or hourly rainfall data. This was because of the absence of automatic recording machine until 2009 in Bahir Dar city. Despite of recording the automatic sub hourly rainfall, the data recorded had missing data to most of the months of the years from 2009-2011.Hence,it was vital to take the non-missing sub hourly and hourly rainfall data for the period of 4 years from 2012-2015.This is also part of the limitation of this study.

In general, Bahir Dar city annual maximum daily rainfall data was having the property of being independent, stationary, and homogenous. When, the rainfall data was said to be independent that it means that no observation in the data has any influence on any following observations. The homogeneity of the annual maximum daily rainfall data showed that all the elements of the data were originated from a single population. The annual maximum daily rainfall was time scale invariant, which it means that the data was stationary. In the other case, from the analysis for outliers, it was noted there were no any outliers in the rainfall data. When it was said that the rainfall data had no any outliers, it means that there were no any data that deviated from the bulk of the rainfall data that were analyzed for this study.

The annual maximum daily rainfall data of the Bair dar city were time scale invariant. This was further verified that the scale exponent of the scale factor of the simple scaling to the rainfall data was constant, which was verifying the scale invariant of the rainfall data in the study area. In the plotting of the scale exponent and order of moment, it was experienced that the scale exponent was analyzed in two parts. This was mainly because of distinction of the linearity of the slope for the two duration of a) $1hr \le d \le 24hr$ and b) $15\min \le d < 1hr$ exactly at duration of 1hour.In this regard, it was possible to generate two scale exponents to the $1hr \le d \le 24hr$ and $15\min \le d < 1hr$ as 0.8145 and 0.5163 respectively.

The generalized extreme value(GEV) distribution function was best fitting to the annual maximum daily rainfall data to the study area. This was identified using the easy fit software that really helped to identify this distribution function in order to further analyze the relationship to IDF and its curves. Using the inverse function of the cumulative generalized extreme value(GEV) function, it was quite possible to drive the relationship among the intensity-duration-frequency for the study area. In the easy fit software, it was also quite possible to determine the parameters of the GEV distribution function.

In this study, the design rainfalls for the return periods of 2 years,5 years,10 years,25 years,50 years and 100 years were determined. In this analysis, it was possible to determine the design rainfall using the frequency factor method which suited to do for the extreme value distribution functions. The design rainfalls for the return periods of 2 years,5 years,10 years,25 years,50 years and 100 years are 69mm, 89mm, 103mm, 120mm, 133mm, and 146mm respectively.

The relationship between the intensities which were developed using the simple scaling method and regional analysis method were calculated with respective formulae for the return period of 2 years,5 years,10 years,25 years,50 years and 100 years with a durations of 15min,30min,45min,60min, and 120min. The development of relationship for the two intensities from simple scaling and regional method were very critical in order to compare and contrast the difference and similarities of the two curves in Bahir dar city. Therefore, the relationship of intensities from the regional IDF curves and simple scaling IDF curves were calculated as a function of durations for each respective return periods.

From annual average rainfall and altitude data of the geographic locations which were considered under the development of A2 and A3 IDF regional curves where Bahir Dar city was located, one can understand that the annual average rainfalls are not the same across the locations or stations on which the regional IDF curves were developed. This led us to a conclusion that there is a significant variation with the comparison of the simple scaling method and regional method IDF curves development. In general, it is very vital to develop independent IDF curves for the respective cities in Ethiopia in order to make sure each and every city has its own specific IDF curves developed that can help in the drainage structures design and construction works. Hence, the annual average rainfall for Bahir Dar city with its altitude of 1797m is 1419mm. In Gondar, the annual average rainfall is 1151mm with its altitude of 2226m.Debre markos has annual average rainfall of 1321mm at its altitude of 2462m.In the same note, Debre tabor has annual average rainfall of 1497mm with its altitude of 2706mm.Fitche has annual average rainfall of 1136mm at altitude of 2802m. In the same fashion, Addis Ababa has annual average rainfall of 1143mmm with its altitude of 2350m.And Adama/Nazareth has annual average rainfall of 808mm with its altitude of 1623m. (climate-data.org).

The rainfall data obtained from the IDF curves developed using the simple scaling method falls above the values of the observed rainfall data or the intensities from IDF curves using simple scaling method falls beyond the values of the observed data. However, the rainfall or intensities from the regional IDF curves were falling below the observed data of Bahir dar city in the sub hourly recording comparison. This showed that the regional IDF curve underestimates the values of the rainfalls or intensities for the Bahir Dar city. This will have a consequence or an impact on the under designing of drainage structures in the city, which consequently leads to road side flooding and also damaging of civil works by flooding. But the simple scaling method developed IDF curves have an added advantage in accommodating the flooding generated from the city as it is possible to accommodate the flood generated from the catchment areas. This is mainly because of the values of the rainfalls or intensities using simple scaling generated IDF curves have greater values than the observed values. This difference between the observed and simple scaling method generated IDF curve values of rainfalls or intensity can serve as a factor of safety in designing drainage structures in the city. Moreover, the accuracy of the regional IDF curves are so limited as the values of the full intensities calculated from the IDF curves are far below even the observed data. This sends a wrong signal to the engineers who are using these curves for the design of drainage engineering structures. It is very wise and vital to use the simple scaling method or re-correct the regional IDF curves data using the relationships developed in table 19.

The rainfall data obtained from the IDF curves developed using the simple scaling method falls above the values of the observed rainfall data or the intensities from IDF curves using simple scaling method falls beyond the values of the observed data. However, the rainfall or intensities from the regional IDF curves were falling below the observed data of Bahir dar city in the sub hourly recording comparison. This showed that the regional IDF curve underestimates the values of the rainfalls or intensities for the Bahir dar city. This will have a consequence or an impact on the under designing of drainage structures in the city, which consequently leads to road side flooding and also damaging of civil works by flooding. But the simple scaling method for developing DF curves has an added advantage in accommodating the flooding generated from the city as it is possible to accommodate the flood generated from the catchment areas. This is mainly because of the values of the rainfalls or intensities using simple scaling generated IDF curves have greater values than the observed values. This difference between the observed and simple scaling method generated IDF curve values of rainfalls or intensity can serve as a factor of safety in designing drainage structures in the city.

5.2. RECOMMENDATIONS

- Annual maximum daily rainfall data and sub hourly and hourly rainfall data recording should be fully recorded in the future.
- Workshops should be organized on the importance of rainfall or other hydrological data to relevant stakeholders working on hydrological data collections so that the fully recorded data can be properly used for the applications of hydraulic engineering works.
- Similar studies should be conducted to most of major cities in Ethiopia.

CHAPTER 6: REFERENNCES

Abubakari,, Sulemana 2013: Development of Intensity-Duration-Frequency Curves for Kumasi, Ghana.

Afrin, S.; Islam, M.M. and Rahman, M.M., 2015: Development of IDF curve for Dhaka City Based on Scaling Theory under Future Precipitation Variability Due to Climate Change, International Journal of Environmental Science and Development,Vol.6,No.5,May 2015.

Bara, M.; Kohnova, S.; Szolgay, J.; Gaal, L., and Hlavcova, H.,2010: Assessing of IDF curves for hydrological design by simple scaling of 1-day precipitation totals.

Chang, C.W. and Hiong, S., 2013:Estimation of sub daily IDF curves in Singapore using simple scaling method,Impact world 2013;International Conference on Climate Change Effects Postdam, May 27-30.

Chow,1988;Applied Hydrology printed by McGraw-Hill Companies in United states of America.

Elshorbagy, Amina A., Panu, U.S., and Simonovic, S.P., 2009: Group-based estimation of missing hydrological data: I. Approach and general methodology.

Ethiopian Road Authority, 2013: Drainage Design Manual.

Gebremariame, Zeryehun Haile: 2009:Assessment of Climate Change Impact on the Net Basin Supply of Lake Tana Water Balance.

Jaleel, Lamia Abdul, and Farwan, Maha Atta, 2013: Developing Rainfall Intensity Duration-Frequency Relationship for Basrah City, Kufa Journal of Engineering(K.J.E),ISSN 2207-5528,Vol.5,Issue 1,Dec.,2013,P.P.105-112.

Merra, Nigatu, 2011:Rainfall Intensity Duration Frequency(RIDF) Relationships under the changing Climate(Case Study on Upper Blue Nile River Basin, Ethiopia; Addis Ababa University, Addis Ababa Institute of technology submitted on September 2011.

Mohymont, B., Demaree, G.R. and Faka, D.N., 2004: Establishment of IDF curves for precipitation in the tropical area of Central Africa-comparison of techniques and results; Natural hazards and Erth System Sciences(2004) 4:375-387 SRef-ID: 1684-9981/ihess/2004-4-375.

Nhat, Le Minh; Tachikawa, Yasuto; Sayama, Takahiro; and Takara, Kaoru; 2008: Estimation of Sub-hourly and Hourly IDF curves using Scaling Prosperities of Rainfall at Guaged Site in Asian Pacific Region;Annuals of Disas.Prev.Res.Inst.,Kyoto Univ.,Bo.51 B,2008.

Nhat, Le Minh, Tachikawa, Yasuto; and Takara, Kaoru;2006: Establishment of Intensity Duration Frequency Curves for Precipitation in the Monsoon Area of Vietnam, Annuals of Disas.Prev.Res.Inst.,Kyoto Univ.,No 49 B,2006.

Nhat, Le Minh; Tachikawa, Y.; Sayama, T. and Takara, K.,2006:Derivation of Rainfall Intensity-Duration-Frequency relationship for Short-Duration Rainfall from Daily Data; Proc. of International Symposium on Managing Water Supply for Growing Demand Bangkok,16-20,October,2006, IHP Technical Documents in Hydrology,no.6,pp.89-96,2006.

Naghettini, Mauro; 2012: Application of Scale Invariance Properties of Rainfall for Estimating the Intensity – Duration - Frequency Relationships at Uberaba, in South-Central Brazil, Aqua -LAC-Vol. 4-No2-Dic.2012.pp 45-60.

Rashid, M.M.; Faruque, S.B. and Alam, J.B., 2012; Modeling of Short Duration Rainfall Intensity Duration Frequency(SDR-IDF) Equation for Sylhet City in Bangladesh, ARNP Journal of Science and Technology, ISSN 2225-7217,Vol.2,No.2,March 2012.

Rao, A.Ramachandra and Hamed, Khaled H.,2000: Flood Frequency Analysis, Printed by CRC Press in the United States of America.

Sabramanya,1984: Engineering Hydrology, printed by Tata McGraw-Hill Publishing Company Limited, New Delhi.

Simonovic, Slobodan P.; Peck, Angela :2009: Updated rainfall intensity duration frequency curves for the city of London under changing climate.

Wondie, Tenagne Addisu, 2009: The impact of urban storm water runoff and domestic waste effluent on water quality of lake Tana and local ground water near the city of Bahir Dar, Ethiopia.