ANALYSIS OF EXTREME RAINFALL AND STREAM FLOW EVENTS FROM UPPER EWASO NG'IRO DRAINAGE BASIN IN KENYA

By

Kennedy Kitonga Mwetu

A thesis submitted to the Graduate School in partial fulfillment for the requirements of the Masters of Science Degree in Agricultural Engineering of Egerton University

May, 2004

DECLARATION

I declare that this thesis is my original work and has not been submitted to any other University for an award of a Degree.

Signed………………………………… Date…… 2/6/2004……

Kennedy Kitonga Mwetu

BMII/0442/01

RECOMMENDATION

The thesis has been submitted to the Graduate School for examination with our recommendation as the University supervisors.

Signed………………………………… Date…… 2/6/2004……

J.O. ONYANDO, Ph.D
Department of Agricultural Engineering
Egerton University
P.O. BOX 536
Njoro, Kenya.

Signed………………………………… Date…… 2/6/2004………

M.C.CHEMELIL, Ph.D
Department of Agricultural Engineering
Egerton University,
P.O. BOX 536
Njoro, Kenya.

COPYRIGHT

DEDICATION

First, to GOD who makes all things possible. Secondly, to my dear parents for their affection, love, support and guidance they offered to me.

ACKNOWLEDGEMENT

Several people provided me with special assistance both in field work and thesis writing. I would like to acknowledge their help. First, it is with gratitude that I acknowledge Dr. J.O.Onyando of Department of Agricultural Engineering for his assistance, guidance, criticisms and comments on this work. He shared his vast knowledge in hydrology and journals willingly. Secondly, my thanks and acknowledgement goes to Dr. M.C. Chemelil of Department of Agricultural Engineering. His contribution to this work cannot be under estimated. He created an enabling environment in both financial and academic resources for data analysis and thesis writing. He provided useful advice and guidance throughout thesis writing. I also acknowledge the contributions of Dr. D. M. Nyaanga of the same Department who accepted to read and comment on my work. His comments were indeed worthwhile.

Thirdly, I sincerely thank all my lecturers and colleagues in the Faculty of Engineering and Technology for their generous contribution and support. I wish to thank UNESCO for their financial assistance through Dr. Onyando. This arrangement enabled me acquire data and prepare my thesis. Also I would like to thank Center for Training, Research and Development (CETRAD) staffs who agreed to take off time from their busy programmes and attend to my research needs. Finally, I cannot forget to thank all members of my family and friends who offered what they could both morally and materially to make my Masters of Science studies a success.

ABSTRACT

This study was adopted with three major problems at hand, namely changes the catchment has undergone which are greatly affecting the trends of stream flow and rainfall distribution patterns, series of hydraulic structure failures likely to be connected to inadequate design parameters and ongoing research activities in the basin which need good quality data. In order to provide solutions to the above-mentioned problems, the study was categorized into three major sections as follows, data quality control, progressive trend analysis, design events derivation by frequency probability model. To verify the validity of the applied models, Intensity Duration Frequency curves estimations and comparison of observed and derived values was carried out. The results of this comparison were found to be reasonable. Water control systems and management structures are designed based on hydrologic data measured or estimated.

In the present study measured data were used to calibrate frequency models for predicting design events. Observed rainfall and stream flow data were first extracted from ten gauging stations in upper Ewaso Ng'iro drainage basin in Kenya. Data quality were tested by double mass curve technique, Bivariate model with 2-tailed and coefficient of determination. In all the above tests, the data were found satisfactory for use. The missing rainfall data were estimated by simple proportion and weighting factor methods. These estimated data compared well with observed data and deviations between the two sets were up to 8%. The statistical analysis of all rainfall data revealed that, a part from Nanyuki station which had decreasing mean rainfall trend with time all other stations had increasing trends but at different scales and magnitudes. Extreme Value Type I distribution fitting was confirmed by coefficient of determination and the relationship was perfect with R^2 of 1.0. The Extreme Value Type I functions were derived and used to obtain design storms of 2, 5, 10, 50 and 100 years return periods. The deviations between derived and observed storm values were in the range of 10.2% and 11.7% which was within the reasonable limit. Storms were estimated using Intensity-Duration-Frequency curves for same return periods for comparison and it was found that, they were less accurate especially as return periods increase. Log Pearson Type III fitness on stream flow data were verified by coefficient of determination which

which ranged from 0.94 to 0.97. Fitted Log Pearson Type III distributions were used to simulate design stream flows for specified return periods. Finally, long term flow coefficient K_{95} was derived and regression analysis done with catchment characteristics. Regression equations for determining 95% discharge percentile in ungauged catchments were established.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

LIST OF ABBREVIATIONS

Aden – Adencaple

CETRAD – Center of Training and Development

Cum – Cummulative

Elev – Elevation

EVI –Extreme Value Type I Distribution

FAO – Food and Agriculture Organization

ID – Identification

IDF Curves – Intensity Duration Frequency Curves

K_{95} - Long Term Flow Coefficient at 95 % Percentiles

K_{99} – Long Term Flow Coefficient at 99 % Percentiles

LP3 – Log Pearson Type III Distribution

Max – Maximum

P-value – Probability value

Q – Discharge

R – Pearson linear correlation coefficient

R^2 – Coefficient of determination

RGS – River gauging station

Sy – Standard deviation

UNESCO- United Nations Environmental Socio – Cultural Organization

USDA- United States Development Agency

WMO – World Meteorological Organization

1 INTRODUCTION

1.1 Background

The government of Kenya is committed to increasing agricultural production which depends on water resources among other factors. The upper zones of Ewaso Ng'iro drainage basin have got intensive horticultural activities especially along the river banks. Thus, there is need for proper use of water resources in terms of water availability, possible storage capacities, permissible utilization rates and techniques for preserving existing hydraulic structures. Also there is rapid population growth in the upper zones of the Ewaso Ng'iro drainage basin (Decurtins, S., 1992), leading to increased forest encroachment and water use crisis. In addition to this, water users residing in the upper reaches of the drainage basin divert water continuously during the low flow season. This causes inadvertent water shortage down streams where pastoralists depend on the river water. The impacts of these human interventions then, lead to over exploitation and depletion of water resources. This has greatly affected catchment hydrology and biodiversity.

Finally, there are a number of ongoing research activities in the basin that require reliable data. Moreover, the available data have missing gaps due to inconsistent reading or failure of automatic recorders especially during extreme rainfall and stream flow events. These gaps need to be filled to create continuous and reliable data record for the design of water resources and management system structures. On the basis of this background, the study addressed the hydrologic regime of the upper basin and used frequency models to estimate rainfall and stream flow events.

1.2 Problem and Justification

The upper catchments of Ewaso Ng'iro drainage basin have in recent past experienced destruction of hydraulic structure due to extreme hydrologic events. The rainfall and stream flow data available were short and with gaps due to failures of measuring instruments. The destruction of hydraulic structures could have been as a result of using inadequate data for design purposes. To curb this problem of using inadequate design data leading to destruction of hydraulic structures, frequency models were adopted to generate reasonably reliable design data.

The design of water management structures requires reliable hydrologic data. Under-design leads to failures of the structures by even less frequent events and over-design is more expensive and leads to wastage of resources. Thus, it is only through the use of reliable hydrologic data that appropriate design of various catchment management structures can be achieved. These catchment management structures are designed to accommodate future extreme rainfall and stream flow events whose precise magnitudes are unknown at the time of the design and construction. Through the analysis of available data, estimates of extreme future rainfall and stream flow events can be reasonably estimated. This way, the structures designed and constructed on the basis of such data, stand less risks of failure and therefore are more reliable in managing extreme events (Onyando et. al., 2004). This research took into consideration only the extreme rainfall and stream flow events because they are more destructive. Thus, the present work attempted to derive design storms and stream flows through frequency models that are site specific. This was preferred to the Intensity Duration Frequency (IDF) curves, which are general in nature and regionally based (Onyando and Chemelil, 2004).

1.3 Objectives

The main objective was derivation of design data using frequency models derived from the catchments data. The following specific objectives were suggested:

1. To test data quality analysis and estimate the missing data.
2. To derive maximum annual daily rainfall trends.
3. To derive frequency model parameters.
4. To determine design storms and stream flows for specified return periods.
5. To develop the link between coefficient k_{95} and catchment characteristics.

1.4 Hypotheses

The following null hypotheses were suggested and tested in this study.

- Data quality analysis and filling of missing data is not possible.
- Derivation of maximum annual daily rainfall trends is not possible.
- Frequency model parameters can not be derived from catchment data.
- Design storms and river flows for specific return periods can not be determined.
- Relationship between k_{95} and some watershed characteristics can not be established.

2 LITERATURE REVIEW

This chapter consists of three main sections. The first section presents a review of what has been done generally on frequency analysis. The objective of this section is to establish knowledge gaps on frequency analysis. It also reveals the application and use for coefficient k_{95} with some watershed characteristics. The second section gives the literature related to extreme annual daily rainfall analysis. It aims at finding out how various frequency and statistical methods have been applied, investigates what is to be modified so as to suit various applications depending on the need at hand. The third section outlines what has been done in relation to extreme annual daily flood analysis.

2.1 General Frequency Analysis

Studies by Sharma and Makhoalibe (1987) suggested that, frequency analysis of extreme rainfall and flood processes are of value to hydrologists and water resources engineers for the design of water conveyance structures and water resources management systems. The authors reported that, although the frequency modeling of these sequences has been the subject of extensive investigations, analyses have been mostly restricted to point sites. Analyses based on point data may serve useful purposes for specific design activities. Nevertheless their use precludes the overall understanding of the behavior of these processes in a region. The regional behavior is important in that it provides information on the general tendency of a process in relation to its frequency, temporal and spatial distribution. The parameters based on the regional analyses can therefore, be safely used for synthetic generation of data and determination of statistics needed at a particular site in a drainage basin.

Further, there is fundamental need to develop frequency models that can be used to assess the hydrology during times of low flow and high flow (Mc Mahon and Nathan, 1991). Where stream flow data exist, a variety of frequency models can be used to analyze the flow characteristics. But ungauged catchments always pose a difficult problem due to lack of suitable data. This has been the case in most developing countries such as Kenya where majority of the catchments are not gauged. Appropriate design of water resources management structures in such catchments therefore poses a big problem. In the present study, attempts have been made to develop methodologies which can be used to estimate hydrologic data in ungauged catchments.

Studies by Loukas *et. al.*, (1995) supported the suggestion of practical application of hydrology in the estimation of extreme flood events, because the planning and design of water resource projects and flood – plain management depend on the frequency and magnitude of peak discharges. The authors analyzed extreme events using purely statistical, simulation and derived distribution techniques. The purely statistical methods attempted to fit extreme value probability distributions to measured peak flows but this procedure requires observed data and can therefore be applied only in gauged watersheds. Their work provided hydrologists with improved methods to estimate design flood parameters for mountainous and rural areas, with specific reference to coastal British Columbia. This present work however will investigate the possibility of using similar procedures in selected ungauged catchments in Ewaso Ng'iro basin in Kenya.

The works of Mati *et. al.,* (2000) presented existence of a serious soil erosion problem in the upper Ewaso Ng'iro drainage basin. Through the use of Geographical Information System (GIS) techniques runoff rates were found to be of very high levels during the time of extreme events. Onyando (2000) suggested that, the design of water resources systems is of increasing importance due to the increasing demand for efficient land and water management in a river basin. The need to control water movement is very crucial especially under extreme conditions so as to reduce the effect of extreme events.

Studies by Sharma and Makhoalibe (1987) postulated that peak rainfalls did not necessarily produce peak discharges in their extreme event analysis contrary to previous assumptions in other hydrology studies. According to these researchers, peak rainfall events have been used to give peak discharges in ungauged catchments in Lesotho. This implies that the use of extreme rainfall events for design purposes may not necessarily reflect extreme discharge which a water management structure is designed to control. Therefore analysis of extreme discharge is important as well.

Molnar (1988) suggested that, in cases where long-term mean daily discharges were not available, they could be estimated by means of coefficient k_{99}, which is a function of discharge at ninety nine percentiles and mean daily discharge. This approach worked well in areas with little variability of runoff characteristics in the drainage area but in this work it was tested in areas of wide variability.

Wilson (1975) reported that, it frequently happens when assembling rainfall data that some areas are inadequately recorded. This has been particularly due to mechanical problems of the recording instruments. Also he pointed out that, it was possible to estimate the missing data of one station if the data for the surrounding stations were available and are of the same kind and frequency. This approach was applied to some rainfall stations in upper Ewaso Ng'iro basin to fill up the data gaps and to extend data record.

Hydrology is highly data dependent and requires good data of the catchment in both time and space. The data must portray a good representation of the entire catchment and in addition needs to be checked for homogeneity before use. The precise meaning of the words "representation" and "homogeneity" are obscured by the fact that the requirements are dependent on the purpose of the data and the method of analysis employed (Linsley and Franzini, 1979). With respect to time, "representative" means that the data period must be long enough to include an adequate range of the information to be used. For instance, a flood frequency analysis using only the highest flood each year, twenty years record is minimal (Linsley and Franzini, 1979). The concept of homogeneity therefore implies that the records should have a common meaning throughout the period of record. The same understanding of representation and homogeneity was implied in this study.

Before any analysis is carried out, the data should be checked for temporal and spatial homogeneity. The simplest and most reliable technique for testing homogeneity in hydrologic time series is the double mass curves, which is considered together with the differences in the actual time of occurrence of selected events. Under these circumstances time consistency is important since certain temporal parameters can be compared (Wilson, 1975; and Rohr and Killingtveit, 2003). The double mass curve was used in this study to analyze the homogeneity and consistency of the data.

The double mass curve analysis technique is popular and is commonly used to test rainfall and stream flow records for non homogeneity (Linsley and Franzini, 1979). Non homogeneities are most commonly introduced into rainfall record by moving the station. In mountainous terrain or in urban centers, even a small movement may cause a marked change in rainfall catch. A change in observer or type of equipment may also cause a shift of several percentages.

Non homogeneities in stream flow data are most commonly caused by construction of a dam, levees, or diversion upstream of the station. In small watersheds, urbanization and other major changes in land use, or forest fires may be significant. Corrections for changes should be made before the record is used or alternatively, the analysis could be limited to the portion of the period before or after the change (Linsley and Franzini, 1979; and Wilson, 1975).

The double mass curve can identify the periods of changing trends of hydrological variables. The approach involves plotting the accumulating totals of one time series against one or more station data series. The data series being tested should, if possible be compared to at least 4 to 5 base station records of the data. One of the series is assumed to be homogenous, and if the plot is an acceptable straight line, the other series are also assumed to be homogenous (Linsley and Franzini, 1979).

For consistency, the plot is expected to maintain a constant slope throughout. A change in slope and an identified clear break in the data series indicate inconsistency (Chemelil and Smout, 2000). Establishing the gradient and extending the plot of the consistent part could correct variations causing inconsistency.

2.2 Extreme Annual Rainfall Analysis

This subsection reviews the methods of fitting the data in a given type of probability distributions, testing the goodness of fit and extreme value type I distribution. Chow et. al., (1988) fitted the distribution of maximum annual daily rainfall from College Station in Texas (1911-1979) using the above method. The feasible range of rainfall variable was divided into discrete intervals and then the number of observations falling into each interval was counted. The relative frequency function and cumulative frequency function were determined by equations (2.1) and (2.2) respectively.

$$f_s(x_i) = \frac{n_i}{n} \tag{2.1}$$

Where subscript s for the function f was the sample size of the rainfall distribution.

Subscript i for the variable x (rainfall amount) was the number of the intervals and n_i was the number of observations in the interval (i).

$$F_s(x_i) = \sum_{j=1}^{n} f_s(x_j) \tag{2.2}$$

The standard normal variates corresponding to the upper limit of each of the data intervals were calculated by the equation (2.3)

$$Z = \frac{x - \mu}{\sigma} \tag{2.3}$$

Where z = standard normal variate

μ = mean of daily rainfall events

x = measured variable (daily rainfall amount)

σ = standard deviation

Relative frequency and cummulative frequency functions were plotted against maximum annual daily rainfall events and the type of distribution identified (Chow et. al., 1988). From the resulting probability distribution, the suitable probability distribution was selected based on the best fit of regression determined by coefficient of determination (R^2).

The same method of moments was used by Clarke (1973) to fit maximum daily mean discharge for 28 year period to two–parameter gamma distribution from Brenig basin. The same researcher found the method of moments to be easier and more accurate than the method of maximum likelihood in fitting the distribution, which justified its adoption in the current study. To check the goodness of fit chi-square test denoted by equation (2.4) is mostly used. Shahin et. al. (1993) revealed that this procedure has got little power in justifying which probability distribution fits the data and hence coefficient of Pearson and determination were used to check the goodness of fit in this work.

$$\chi_c^2 = \sum_{i=1}^{m} n[f_s(x_i) - p(x_i)]^2 \tag{2.4}$$

Where m = number of intervals

$nf_s(x_i)$ = Observed number of occurrences in interval (i)

$np(x_i)$= Corresponding expected number of occurrences in interval (i) and also from chi-square test equation (2.5) could be derived.

$$v = m - p - 1 \tag{2.5}$$

Where v = degrees of freedom

m = number of intervals

p = number of parameters used in fitting the proposed distribution.

In their work, if calculated chi-square was less than limiting chi-square value, then it was agreed that the data fits the distribution. The maximum daily rainfalls events have been mostly modeled by the Extreme Value Type I distribution (Chow et. al., 1988; Sheng Yue, 2000; and Lee *et. al.*, 2001). With regard to maximum annual daily rainfall, Extreme Value Type I was preferred to analyze the rainfall data because of good performance as revealed by Chow et. al. (1988). The Extreme Value Type I probability distribution function as documented in hydrologic texts is given by equation (2.6):

$$F(x) = \exp\left[-\exp\left(-\frac{\chi-\mu}{\alpha}\right)\right] \qquad \infty \le x \le \infty \tag{2.6}$$

The parameters α and μ were estimated by equations (2.7) and (2.8).

$$\alpha = \frac{\sqrt{6}}{\pi}s \tag{2.7}$$

$$\mu = \bar{x} - 0.5772\alpha \tag{2.8}$$

Where μ is the mode of distribution or point of maximum probability density and y is the reduced variate defined by the equation (2.9).

$$y = \frac{x-\mu}{\alpha} \tag{2.9}$$

Type of Extreme Value distribution asymptotic was obtained by plotting variate x against reduced variate y as Type I, Type II or Type III. Substituting equation (2.9) in equation (2.6), equation (2.10) was derived. By making y the subject of the formula in equation (2.10), gives equation (2.11).

$$F(x) = \exp[-\exp(-y)] \tag{2.10}$$

$$y = -\ln\left[\ln\left(\frac{1}{F(x)}\right)\right] \tag{2.11}$$

Since $F(x) = \dfrac{T-1}{T}$, then equation (2.11) could also be written as equation (2.12).

$$y_T = -In\left[In\left(\frac{T}{T-1}\right)\right] \tag{2.12}$$

For the EVI distribution χ_T is related to y_T by equation (2.13).

$$\chi_T = \mu + \alpha y_T \qquad (2.13)$$

Where x_T is the maximum annual daily rainfall event of return period T. Equation (2.13) then yields the maximum annual rainfall event of specified return period (T). This method was used in this study to estimate design storms.

2.3 Extreme Annual Flood Analysis

This section discusses techniques used to analyze annual daily discharge. It focuses on Log Pearson Type III distribution which is commonly used in flood frequency analysis (Linsley and Franzini, 1979). Maximum annual daily flow analysis has been mostly modeled by Log Pearson Type III distribution while drought flows by the Weibull Type III distribution (Chow et. al., 1988; Shahin et. al., 1993; and Meunier, 2001). The procedure for the analysis as presented by the authors involved converting the discharges into logs and deriving the statistical parameters, namely mean, standard deviation and coefficient of skewness. The parameters are used to derive probability models for a given catchment. The maximum annual daily discharges are obtained by taking the antilog of the equation (2.14).

$$y_T = \bar{y} + K_T S_y \qquad (2.14)$$

Where, K_T is frequency factor read from Log Pearson Type III Distribution Tables presented in Table A6 (Appendix A). The Table 2.1 summarizes the distributions adopted in this study and includes Extreme Value Type I and Log Pearson Type III distributions. They have been preferred due to their good performance (Chow et. al., 1988).

Table 2.1. Summary of the frequency probability distributions applied.

Distribution	Probability function	parameters
1. Extreme Value Type I.	$F(X) = \exp\left[-\exp\left(\frac{-x-\mu}{\alpha}\right)\right]$	$\alpha = \frac{\sqrt{6}}{\pi}s$ $\mu = \bar{x} - 0.5772\alpha$
2. Log-Pearson Type III.	$f(x) = \frac{\lambda^B (y - \in)^{B-1} \ell^{-\lambda(x-\in)}}{x\Gamma(B)}$	$\lambda = \frac{S_y}{\sqrt{B}}$ $B = \left[\frac{2}{C_s(y)}\right]^2$ $\in = \bar{y} - S_y\sqrt{B}$

Source: Chow et. al., 1988

3 METHODOLOGY

3.1 Study Area

The study area covers the upper catchments of the Ewaso Ng'iro drainage basin with an altitude range of 1780 m to 5199 m above mean sea level (Decurtins, 1992). It lies between latitudes $0^0 20$" South and $1^0 15$" North, Longitudes $36^0 10$" East and $38^0 00$" East and situated on the leeward side of Mt.Kenya and Nyandarua Range. The study area is shown in Figure 3.1, together with rivers selected for detailed study.

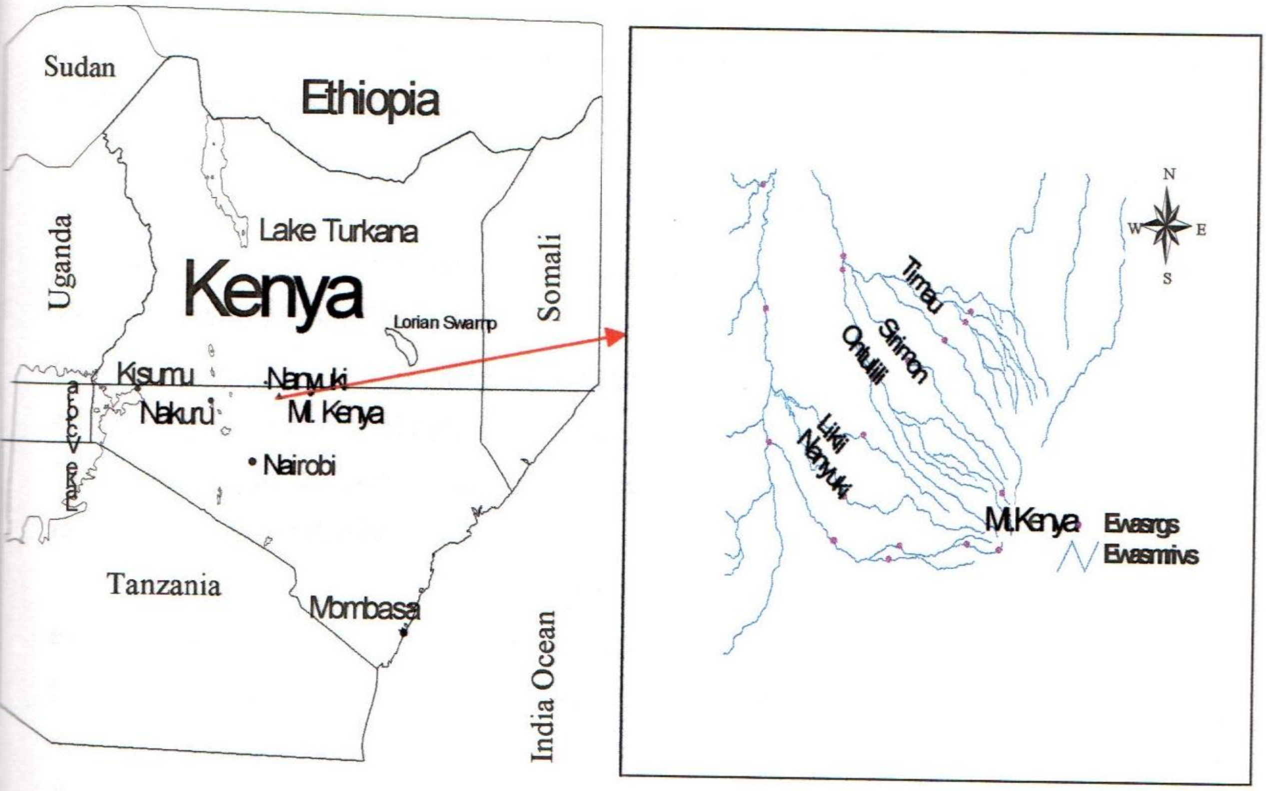

Figure 3.1. Study area located on the upper parts of Mt. Kenya.

Soils in study catchments are broadly classified into four major classes based on the FAO-UNESCO method. These were then converted into the USDA system of classification. The top most part of the basin consists of loam to clay soils (*Cambisols*), which are imperfectly drained and shallow to moderately deep with rock outcrops. Next to this zone is clay loam to clay (*Andosols*), which are well drained and very deep. Clay soils (*Luvisols*), comes after andosols, which are well drained and moderately deep to very deep. Finally, there is clay (*Phaeozems*), which is well drained and very deep (Sombroek et. al., 1980).

About 70% of the basin comprises of rangelands especially the lower parts. The central plateau has large – scale commercial ranches while northern region of the basin is dominated by pastoral communities (Decurtins, 1992). Although wild animals are found in most parts of the basin, there are several game reserves on the upper parts of the basin run privately or by public institutions. The small proportions of the upper basin occupied by croplands are used as large-scale wheat and barley farms in thin strips around the slopes of Mt. Kenya and the Nyandarua range. Small - scale settlements are operated as mixed farms (Liniger, 1998). In these small-scale operations, maize is the predominant crop. A lot of horticultural farming using irrigation systems is practiced along the river banks.

The upper basin experiences two rainy seasons, long rains occurring between April and June and short rains between November and December. The annual rainfall amounts vary significantly with the central part receiving between 600 mm and 750 mm per annum (Berger, 1989). Close to Mount Kenya the value increases rapidly with altitude. On the western side of the mountain (Naro Moru track), the maximum annual daily rainfall approximately 1,500 millimeters is reached at an altitude of about 3,200 meters above mean sea level (Liniger, 1998). The beginning of the rainy season varies greatly and is thus difficult to predict the extreme rainfall event, a fact that has a major impact on the agricultural productivity of the area.

The human settlement regimes in upper Ewaso Ng'iro basin (Mount Kenya region) can be explained within the context of four significant environmental phenomena (Liniger, 1998). Firstly, the evolution of settlements, which involved major ethnic groups of Eastern Bantu, Southern Cushitics and subsequent interactions between these communities (Kimambo, 1989). These early spatial movements laid the foundation for the regional distribution of the contemporary communities of Kikuyu, Embu, Meru, Tharaka and Nderebo. These communities introduced different types of land-use activities characterized by subsistence farming, small-scale mixed farming, animal husbandry, hunting and logging. These activities were influenced by colonial settlement, mostly in Nyeri District, Laikipia District and Timau District. The settlers introduced commercial agriculture, involving large-scale and small-scale farming in coffee, tea, rice and animal husbandry.

The second factor relates to the significant role of Mount Kenya Forest as a suitable habitat for launching the Mau Mau Rebellion against White farmer's alienation of native land. The struggle against land annexation caused establishment of squatter settlements in the region. These settlements prominent in Nyeri District, still sustain some degree of population mobility within Mount Kenya Region. It should be noted that there is a tendency for out-migration from squatter settlements in search of virgin lands for settlement (Muruiki, 1989).

Thirdly, Mount Kenya and its immediate environment are richly endowed with diverse natural and human resources. Distinct spatial variations in rainfall, water resources, soil fertility, forestry resources, wildlife and income per capita, have lead to different land-use activities. These have caused population mobility due to spatial variations in perception of environmental utility and risks. The region is therefore, becoming a focus of economic investments, especially in tourism and commercial agriculture.

3.2 Data Quality Analysis and Estimation Methods

This section outlines the procedures taken to accomplish the specific objective on data quality analysis and estimation of missing data.

3.2.1 Homogeneity and Consistency Tests

A transect survey was conducted in October 2002, to locate the rainfall-gauging stations used in the study area. The stations were selected on the basis of data consistency and length of their record. However, due to lack of long data record in many rainfall-gauging stations, some were chosen with less than thirty years record. These were accepted to ensure the entire catchment was well represented. Some of the rainfall gauging stations had incomplete data for the selected period of this study. This was due to irregularities and failures of the recording instruments during the period selected. Adencaple station had complete data and thus was accepted as the base station in the homogeneity tests.

The reconnaissance survey made possible the selection of streams with long time data record, and spatially represents the entire catchment. Some data had gaps especially in areas where flow measuring instruments had been washed away by floods. The periods of continuous records were used in quality analysis. Despite the fact that some river gauging stations had some years without complete data, all stations had daily flow data between 1970 and 2001. Timau station had complete data and thus was taken to be the base station in homogeneity tests.

3.2.2 Variability Test between the Stations

Variability tests were aimed at establishing whether the data collected had any significant variation and might require special treatment. Consequently the coefficient of determination (R^2) was used to check whether the data had any significant variation. Coefficient of determination (R^2) in this case means the proportion of total sum of squares attributable to another source of variation or the degree of closeness of data of one station to another. If the coefficient of determination was close to 1.0, then the two stations being compared were close in variability otherwise there was no close variability. A similar approach was applied to stream flow data to check their variability in different streams.

3.2.3 Correlation Tests

To investigate whether the data were significantly correlated, Bivariate Model with 2-tailed test was used because there was no control of the measured variable (Steel and Torry, 1981). Probability value (p-value) from the model was used to compare the results from the correlation analysis. If the p-values from the analysis were less than the model values, it indicated that the data had no relationship at all otherwise there was some relationship. Correlation in this case implies association or relationship between two or more station data. The correlation analysis was performed for all rainfall and stream flow data.

3.2.4 Methods of Estimating the Missing Data

Two methods were used to estimate the missing data, namely simple proportion and weighting factor methods. Both methods were used because of the resulting spatial variations of the rainfall stations considered as revealed in data quality analysis. Sirimon Station had significant variation from other stations as indicated in the results (Table 4.7 section 4.1.3) and thus it was treated separately by weighting factor method, which utilized regression analysis. The station had an explained variance of 29.3%, which meant that the estimated values did not represent the original values with sufficient quality if simple proportion was used. The coefficient of variation in other stations was 99.9%, and therefore a simple proportion method was adopted.

Simple Proportion Method

Simple proportion method was applied in rainfall gauging stations; Adencaple, Ontulili, Gathiuru and Nanyuki, which typically had little spatial variation in precipitation as revealed in Table 4.7 (section 4.1.3). It involved calculation of the annual arithmetic mean values from surrounding rainfall-gauging stations and using them to fill up gaps in the rainfall station with missing data for the corresponding years.

Weighting Factor Method

This method was used to fill up missing data based on the assumption that, gauges that were spatially close to each other tended to depict similar rainfall characteristics. The steps involved were: (i) Identification of neighbouring stations (ii) Correlation and linear regression analysis (iii) Identification of the base station and (iv) Computation of missing data by regression.

Identification of Neighbouring Stations

The stations with missing data and the surrounding stations that were likely to yield similar decadal characteristics were selected manually. During the selection, the stations with similar spatial variation were preferred instead of close neighbours because they gave almost similar data. The presence of higher quality data in terms of consistency and complete time series was another criterion applied.

Correlation and Linear Regression Analysis

By means of statistics (correlation coefficient, mean values and linear regression curve) the rainfall gauging stations that were assumed to yield similar characteristics were compared with the rainfall gauging station with missing data. These statistics were calculated for each neighbouring rainfall station (denoted as x) and the gauge for which the data needed to be filled (denoted as y). The Pearson linear correlation coefficient (R) was calculated to check the relationship between the station data. The correlation coefficient (R) normally ranges from –1.0

to 1.0 with R value of unity (1.0) indicating a complete positive dependency between two variables, say x and y.

Coefficients of determination in Table 4.3 (section 4.1.1) were above 0.99 in all rainfall stations. From this kind of relationship, it was thus concluded that, complete positive dependency existed among the stations. Based on coefficient of determination criteria, equations of linear regression for estimating missing data were derived.

Identification of the Best fit / Base Station

The station with complete consistent data was selected by visual inspection and used as base station in the regression analysis. In this case Adencaple station was selected as the base station.

Determination of Missing Data based on Regression

Estimation of missing data based on regression was applied at Sirimon rainfall station where variation in rainfall data could not be assumed. Ratios of longer –time period were calculated from regression analysis among means of adjacent gauges and the Sirimon rainfall station gauge. These ratios were used as weighting factors to estimate the missing data.

3.3 Derivation of Rainfall Trends

Three year moving averages of annual maximum daily rainfall events were worked out and plotted against time in years. The average catered for any significant maximum annual daily rainfall variation and smoothed up the trend line. The nature of the trend line was the indicator of what was happening in the sub catchment. The trends were described as either increasing or decreasing with time.

3.4 Extreme Value Type1 (EV1) Distribution

Statistical analyses were carried out to derive frequency model parameters and their probability distribution for the collected data. It involved determination of statistical parameters (mean and standard deviation) and model parameters as outlined in equations (3.1) and (3.2).

$$\alpha = \frac{\sqrt{6}}{\pi} S \tag{3.1}$$

$$\mu = \bar{x} - 0.5772\alpha \tag{3.2}$$

where α is significance level, μ is mode of the distribution, $\bar{x}$ is mean and S standard deviation of maximum annual daily rainfall events. Prediction of extreme annual daily rainfall events was performed by Extreme Value Type I model presented in equation (3.3).

$$F(x) = \exp\left[-\exp\left(-\frac{x-\mu}{\alpha}\right)\right] \qquad \infty \leq x \leq \infty \tag{3.3}$$

The equation (3.3) above was used to derive maximum annual daily rainfall events of specific return periods. The rainfall stations considered were: Ontulili Forest Station, Sirimon Gate, Gathiuru, Nariyuki and Adencaple.

3.5 Log Pearson Type III Distribution

The stream flow statistical analysis was carried out based on Log Pearson Type III distribution from five river gauging stations. These included Timau, Ontulili, Likii, Sirimon and Nanyuki. The distribution was found to be appropriate in analyzing stream flows as indicated in literature review section 2.3 and was expressed as;

$$f(x) = \frac{\lambda^{B}(y-\in)^{B-1}\ell^{-\lambda(x-\in)}}{x\Gamma(B)} \tag{3.4}$$

The parameters of Log Pearson Type III distribution were determined as in equations 3.5 to 3.7.

$$\lambda = \frac{S_y}{\sqrt{B}} \tag{3.5}$$

$$B = \left[\frac{2}{C_s(y)}\right]^2 \tag{3.6}$$

$$\in = \bar{y} - S_y\sqrt{B} \tag{3.7}$$

where y = logarithm of maximum annual daily stream flow events, S_y = standard deviation of maximum annual daily stream flow events. Maximum annual daily stream flow events for specified return periods were derived using equation 3.8.

$$Logx = \overline{Logx} + K\sigma_{\log x} \tag{3.8}$$

where $\overline{Logx}$ is logarithmic mean, σ_{LogX} is logarithmic variance of the maximum annual daily stream flow events and K is frequency factor obtained from Tables of K values for Log Pearson Type III distribution through interpolation between coefficient of skewness and return period.

3.6 Comparison between Derived and Observed Values

To compare the derived and observed values, Adencaple and Gathiuru rainfall stations were chosen because they had data for thirty years period even though Gathiuru had some years without complete data. The data were divided into two portions of fifteen years time series and EVI probability model established for the first portion for each rainfall station. Using derived EVI models, extreme annual daily rainfall data were derived and compared with observed rainfall data in the second portion of the data period. Also annual daily rainfall data were derived by IDF curves of Nanyuki and compared with rainfall data derived by established EVI models.

3.7 Long-Term Flow Coefficient (K_{95})

Investigation was carried out to find the relationship between long-term flow coefficient and some catchment characteristics namely; elevation, slope and maximum annual daily rainfall. The stream flows were ranked in descending order of magnitude and discharge at ninety nine percentiles recorded. However, some stream flow stations could not give ninety nine percentiles flow and thus highest flows were recorded. Ninety nine percentile discharge (Q_{99}) or highest flow in that manner was divided by average discharge (Q_{Avg}) to obtain long term flow coefficient. This process was done to all stream flow data and their respective long term flow coefficients were obtained.

Common long term flow coefficient in all stream flow data was selected and used to correlate and regress with catchment characteristics. The catchment characteristics chosen were based on the influence they had on stream flows. The long-term flow coefficient achieved was (K_{95}) expressed as in equation (3.9).

$$K_{95} = \frac{Q_{95}}{Q_{Avg}} \tag{3.9}$$

where Q_{95} is discharge at ninety five percentile and Q_{Avg} is mean maximum annual daily discharge. Using correlation and regression analysis, the coefficient of determination was determined and used as measure of their relationship. The ratio (Q_{95}/Q_{Avg}) can be used to estimate discharge at ninety five percentiles at ungauged sub catchments if their mean discharges are available.

4 RESULTS AND DISCUSSIONS

4.1 Data Quality Analysis

Data quality analysis is essential before data application. This is because during data collection, errors may be introduced and hence correction may be required. The data quality analysis gives a data set which is homogenous and also indicates where data need correction. The analysis could give data of the same kind and frequency (Shahin et. al., 1993). Thus during data application, data of the same kind and frequency can be analyzed by the same technique. Table 4.1 gives the selected rainfall-gauging stations, their identification number, data period and years with missing data.

Table 4.1. Rainfall-gauging stations selected from upper Ewaso Ng'iro basin for the study.

RGS	Station number	Data period recorded	Years with incomplete data
Adencaple	2	1970-2001	None
Ontulili	75	1970-2001	1993 and 1994
Sirimon	116	1983-2001	1990 and 1993
Nanyuki	40	1989-2001	1992
Gathiuru	17	1970-2001	2000 and 2001

Adencaple station had complete data and thus was used as the base station in the homogeneity and consistency tests. Gathiuru rainfall station had data for all the years but years listed in Table 4.1 were not complete and their missing data were estimated. The derivation of frequency probability models was confined on continuous data period and estimated data values were not included because their validation was not done.

4.1.1 Double Mass Curve

The cumulated maximum annual daily rainfall data used in plotting double mass curve are presented in Table 4.2 and from it the following observations were made: Adencaple rainfall station had complete data, Gathiuru rainfall station had incomplete data in year 2000 and 2001, Ontulili rainfall station had missing data from 1993 upto 2001, Sirimon had missing data from

1970 to 1982 and finally Nanyuki had missing data from 1970 to 1988. Dashes in the same table indicate years without or with incomplete data in general.

Table 4.2. Cummulated annual daily rainfall (mm) used in double mass curve.

Year	Adencaple	Ontulili	Gathiuru	Sirimon	Nanyuki
1970	375.7	616	1153.6	-	-
1971	961.5	1568.2	2147.7	-	-
1972	1576.9	2306.6	3247.9	-	-
1973	2029.9	3021.1	3991.6	-	-
1974	2531.9	3864.8	4629	-	-
1975	3270.7	4817	5421.2	-	-
1976	3857.1	5169.8	6313	-	-
1977	4764.6	6439.3	7193.9	-	-
1978	5468.7	6984.2	7863.9	-	-
1979	5992.3	7867.3	8784.1	-	-
1980	6286.7	8366	9465.5	-	-
1981	7138.3	9362.5	10075.4	-	-
1982	7910.7	10520.6	11159.8	-	-
1983	8549.5	11567.8	12041.4	1061.7	-
1984	8835	12134.2	12956.1	1609	-
1985	9327.1	12938.6	13473.1	2659.5	-
1986	10215.8	13956.2	14299.2	3667.9	-
1987	10740.3	14635.7	15154.2	4536.4	-
1988	11637.5	15632.7	16024.3	5897.8	-
1989	12216	17106.2	17107.7	7030.3	977.9
1990	12928.8	18319.5	18247.1	7383.5	2045.5
1991	13431.8	19066.7	19222.3	8199.2	2761.8
1992	14013.6	19727.8	19920.3	8794.2	3590.9
1993	14488.2	-	20861.1	9462.6	4541.4
1994	15329.5	-	21767.9	10455.4	5583.7
1995	16027.3	-	22720.4	11378.1	6819.1
1996	16593.1	-	23839.1	11963.1	7668.9
1997	17753.4	-	24453	13394	9127.2
1998	18520.6	-	25745.5	14617.1	10131.8
1999	18921	-	26984.7	15291.4	10736.6
2000	19305	-	-	15908.3	11167.7
2001	19826.2	-	-	16650.1	11852.8

The double mass curve technique investigated whether collected rainfall data were homogenous and consistent throughout selected period of study and revealed that no correction of data was needed. The graphical representation in Figure 4.1, indicate that the rainfall data are homogeneous and consistent since a straight-line plot through multiple regression analysis was obtained. The coefficient of determination (R^2) in Table 4.3 of 0.99 close to one and the positive slope confirms the good quality data (Chemelil and Smout, 2000).

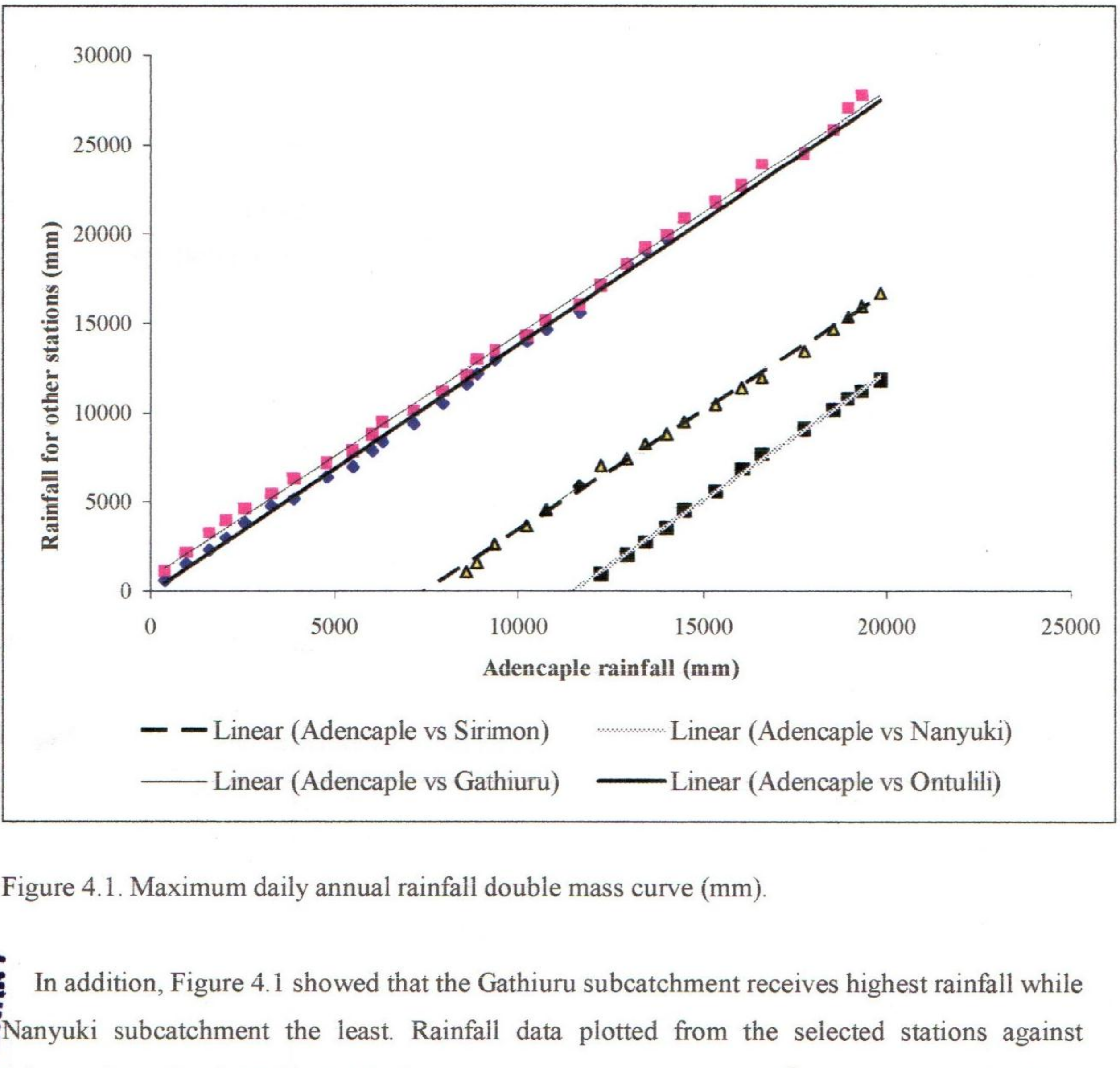

Figure 4.1. Maximum daily annual rainfall double mass curve (mm).

In addition, Figure 4.1 showed that the Gathiuru subcatchment receives highest rainfall while Nanyuki subcatchment the least. Rainfall data plotted from the selected stations against Adencaple station in Table 4.3 had very strong relationship since R^2 approached unity (Steel and Torry, 1981).

Table 4.3. Slopes and coefficient of determination of rainfall double mass curve.

Rainfall gauging station	Double mass curve slope	Coeff. of determination (R^2)
Ontulili vs Adencaple	+1.38	0.99
Sirimon vs Adencaple	+1.32	0.99
Gathiuru vs Adencaple	+1.36	0.99
Nanyuki vs Adencaple	+1.43	0.99

Positive gradient in all stations revealed that the data were of the same kind and frequency (Shahin et. al., 1993). From the results therefore, it can be concluded that the rainfall data were homogeneous and good for use.

River Gauging Stations selected in the study were presented in Table 4.4. All stations had daily stream flow data between 1970 and 2001 but some had years without complete data. However, they were selected to ensure that the entire catchment was represented. Timau station had complete data for all the years and thus was chosen as the base station in homogeneity and consistence tests. Due to breaks or gaps in the data for the period selected, regression lines could not be continuous throughout the period chosen. Thus, the homogeneity test was confined to the period with continuous data.

Table 4.4. River Gauging Stations chosen for the study

River	RGS	Station code	Data Period Record	Years of incomplete data	catchment area (km^2)
Timau	5BE6	AE	1970-2001	None	58.30
Likii	5BE7	AA	1970-2001	1991,1992,1993,1995	172.00
Ontulili	5BE2	AB	1970-2001	1990,1991,1996 and1997	61.10
Sirimon	5BE4	AC	1970-1983	1984,1985,1986 and1987	59.50
Nanyuki	5BE1	A9	1970-2001	1993,1994 and 1995	67.80

Similarly, the stream flow data were subjected to the homogeneity tests using the double mass curve technique. The cumulated annual daily stream flows used in plotting mass curve were as in Table 4.5. The Table indicated that, Sirimon stream flow data were taken from 1970 to 1983 and used in the quality analysis because had complete data. For Likii and Nanyuki stream flow stations, periods used in quality analysis are 1970-1998 and 1970-1984 respectively.

Table 4.5. Cumulated annual daily flow (m^3/s) used in double mass curve.

Year	Timau	Sirimon	Likii	Nanyuki
1970	53.05	126.42	344.11	189.78
1971	118.63	355.34	953.55	412.05
1972	187.21	483.46	1348.16	603.69
1973	237.68	616.09	1672.04	828.78
1974	293.24	798.60	2221.89	994.97
1975	403.75	1059.22	3030.25	1245.26
1976	480.47	1195.44	3603.64	1485.07
1977	593.13	1394.28	4362.22	1863.09
1978	697.79	1521.19	4956.95	2162.64
1979	776.54	1659.09	5510.35	2402.03
1980	835.42	1834.21	5725.94	2489.34
1981	936.59	2094.49	6282.82	2746.51
1982	1001.90	2396.25	6933.31	3027.50
1983	1069.55	2645.81	7574.47	3282.24
1984	1120.52	-	7841.19	3391.23
1985	1167.98	-	7934.34	-
1986	1243.15	-	8521.81	-
1987	1292.91	-	8837.19	-
1988	1346.87	-	9934.79	-
1989	1419.32	-	10582.36	-
1990	1470.78	-	11270.89	-
1991	1512.27	-	11555.26	-
1992	1542.24	-	12017.38	-
1993	1584.91	-	12525.02	-
1994	1653.35	-	13206.81	-
1995	1708.99	-	13691.41	-
1996	1777.02	-	14076.44	-
1997	1878.38	-	14897.68	-
1998	2035.16	-	15811.79	-
1999	2077.94	-	-	-
2000	2098.29	-	-	-
2001	2135.97	-	-	-

The results of homogeneity analysis for stream flows (Figure 4.2) indicated that, all the data were homogenous and consistent since straight line plots were obtained through regression. Coefficients of determination were derived to confirm the regression fitness. It was also discovered that, Likii had highest flow while Sirimon the least, this could be explained by the fact that Likii had a largest drainage subcatchment than the other drainage subcatchments selected.

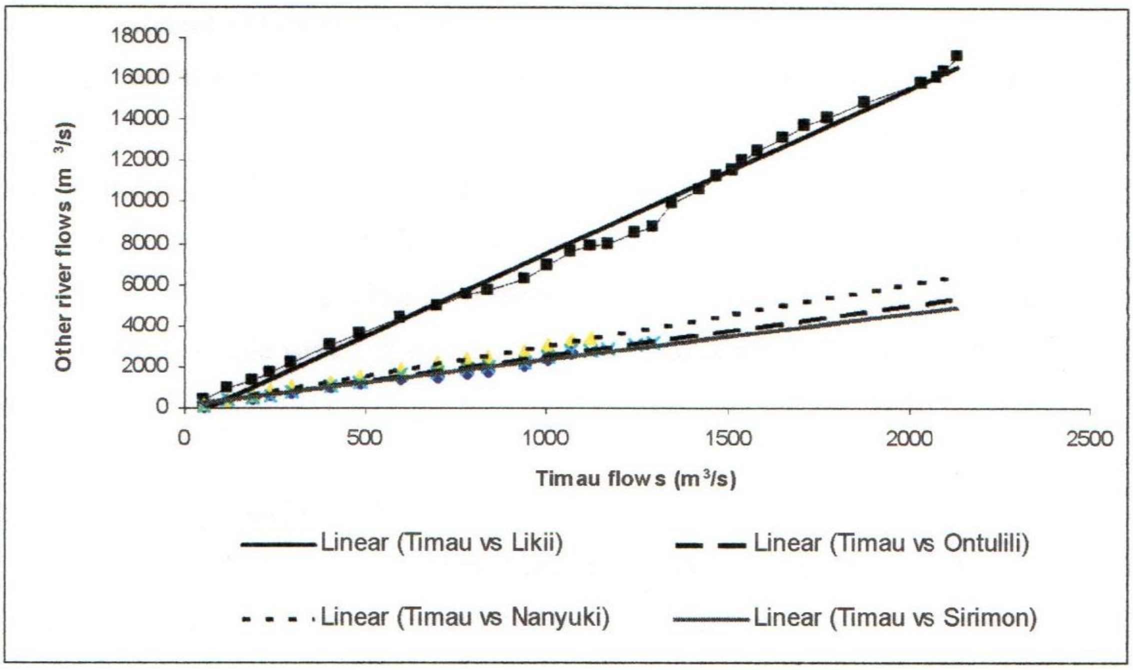

Figure 4.2. Double mass curve of annual daily stream flows (m^3/s).

Apart from Likii stream flow station which had highest gradient, all other streams had almost the same flow gradient as indicated in Table 4.6. In all cases R^2 was above 0.99.

Table 4.6. Slopes and coefficients of determination for stream flow double mass curve.

River gauging station	Slope	(R^2)
Ontulili	2.42	1.00
Likii	7.98	0.99
Sirimon	2.24	0.99
Nanyuki	2.95	1.00

The Likii gradient was higher than other stations because of high flow. The other stations had slight variations in general as shown by varying coefficients of determination. They however seem to move together as depicted by R^2 values close to perfect relationship although at different scales and magnitudes. The uniformity in gradient positivity confirms good quality data for use.

4.1.2 Coefficient of Determination

Rainfall and stream flow gauging stations were located far apart and thus variations in terms of trend or frequency in collected data were queried and consequently coefficient of determination was calculated to confirm the same. In all cases rainfall data gave R^2 of 99.69% indicating variations were not significantly different and 98.59% for stream flow data hence variations were also not significantly different as in Table 4.3 and 4.6. Based on coefficient of determination, the data were still accurate for use but to confirm the same multiple correlations were applied.

4.1.3 Bivariate Model with 2-Tailed

Correlation analysis was done by bivariate model with 2-tailed because of its compatibility with testing of two variables occurring at the same time and results were as shown in Table 4.7

Table 4.7. Rainfall Multiple Correlations.

Station	Statistics	Adencaple	Ontulili	Sirimon	Gathiuru	Nanyuki
	R	1	0.997	0.293	0.999	0.999
	Sig. (2-tailed)	.	0	0.223	0	0
Adencaple	N	32	30	19	32	12
	R	0.997	1	0.995	0.996	0.992
	Sig. (2-tailed)	0	.	0	0	0
Ontulili	N	30	30	17	30	10
	R	0.293	0.995	1	0.297	-0.019
	Sig. (2-tailed)	0.223	0	.	0.217	0.954
Sirimon	N	19	17	19	19	12
	R	0.999	0.996	0.297	1	0.997
	Sig. (2-tailed)	0	0	0.217	.	0
Gathiuru	N	32	30	19	32	12
	R	0.999	0.992	-0.019	0.997	1
	Sig. (2-tailed)	0	0	0.954	0	.
Nanyuki	N	12	10	12	12	12

Correlation is significant at 0.01 level (2-tailed) and dot means no correlation.

Bivariate model with 2-tailed test analyzed whether rainfall gauging stations were either positively or negatively or not correlated at all by setting Pearson correlation coefficient (R). Significant level given by p-value indicated percentage value beyond which the variables could not be related to each other. The correlation enables us to know which data need to be grouped together in the quality analysis because they are related and which ones are not related. The correlation was significant at 0.01 or 1% level as shown in Table 4.7 of rainfall multiple correlations which meant that, p-value above 0.01 or 1% indicated rainfall events were actually different in terms of kind and frequency. The following conclusions were drawn from rainfall multiple correlations:

- Adencaple and Ontulili rainfalls were positively correlated (0.997) but there was no significance difference since p-value of 0.000 was less than set level of 0.01.
- Adencaple and Sirimon rainfalls were positively correlated (0.293) but significantly different since p-value of 0.223 was higher than set value of 0.01.
- Adencaple and Gathiuru rainfalls were positively correlated (0.999) but there was no significance difference since p-value of 0.000 was less than set value of 0.01.
- Adencaple and Nanyuki rainfalls were positively correlated (0.999) but not significantly difference since p-value of 0.000 was less than set value of 0.01.
- Sirimon and Nanyuki rainfalls were the only ones with negative correlation of -0.190 and also significantly different.

Correlation analyses for the stream flows were conducted and results are in Table 4.8. The significant level (p-value) from bivariate model was 0.01 or 1% and above. The stream flow correlation analysis was carried out with null hypothesis that all stream flows were significantly different since they belong to different sub catchments but in same catchment and with alternative hypothesis that they were not significantly different. From Table 4.8 the following observations were made:

- Timau and Ontulili stream flow data were positively correlated (0.998) but not significantly different since calculated p-value (0.000) was less than set value (0.01)
- Timau and Likii stream flow data were positively correlated (0.996) but not significantly different.
- Timau and Sirimon stream flow data were positively correlated (0.993) but not significantly different.

- Timau and Nanyuki stream flow data were positively correlated (0.992) but not significantly different.

Table 4.8. Stream Flow Multiple Correlations.

Rivers	Statistics	Timau	Ontulili	Likii	Sirimon	Nanyuki
Timau	Pearson Correlation	1	0.998	0.996	0.993	0.992
	Sig. (2-tailed)	.	0	0	0	0
	N	32	18	32	14	31
Ontulili	Pearson Correlation	0.998	1	0.999	0.994	0.999
	Sig. (2-tailed)	0	.	0	0	0
	N	18	18	18	14	17
Likii	Pearson Correlation	0.996	0.999	1	0.994	0.999
	Sig. (2-tailed)	0	0	.	0	0
	N	32	18	32	14	31
Sirimon	Pearson Correlation	0.993	0.994	0.994	1	0.994
	Sig. (2-tailed)	0	0	0	.	0
	N	14	14	14	14	14
Nanyuki	Pearson Correlation	0.992	0.999	0.999	0.994	1
	Sig. (2-tailed)	0	0	0	0	.
	N	31	17	31	14	31

Correlation is significant at 0.01 level (2-tailed) and dot means no correlation.

Generally from Table 4.8 the null hypothesis was rejected because all the mean flows were not significantly different and alternative hypothesis accepted. All p-values were less than 0.01 or 1%. This meant that the stream flows chosen were not significantly different and despite of breaks or gaps they were accurate for use.

4.2 Estimation of Missing Rainfall Data

In order to choose the appropriate method for filling the missing rainfall data, multiple correlation and regression analysis was done to find out whether rainfall scenarios were of the same kind and frequency. The following rainfall stations were found to be of the same kind and frequency namely Adencaple, Ontulili, Gathiuru and Nanyuki and hence the method of simple proportion was used to estimate the missing data. Two stations, Nanyuki and Ontulili were selected and missing data estimated by simple proportion method and results were given in Table 4.9.

Table 4.9. Estimated missing Annual Rainfall Data

Rainfall station	Year	Estimated value in (mm)
Nanyuki	1992	829.1
Ontulili	1993	539.3
	1994	1025.3

The estimated values for Sirimon rainfall station were computed by weighting factor method, whereby Adencaple station was taken as the base station. This method was applied in this station because its data was not of the same kind and frequency as other stations as revealed in the correlation analysis. The estimated maximum annual daily rainfall values of sirimon station were 46.1 mm and 46.3 mm for 1990 and 1993 years respectively. Through comparison of estimated and observed values for the same years in Table 4.10, it was found that percentage deviation from the observed values was 8%. This value lies within acceptable range for maximum annual daily rainfall estimation (Wilson, 1975). The actual recorded maximum annual daily rainfall values in Table 4.10 were lower than estimated values, since they were only the maximum values for the available data and not for the whole year. In the estimation analysis, the estimated values were more reliable as they account for the whole year.

4.3 Trends of maximum annual daily Rainfall

The progressive averages of three consecutive years of record were plotted at the mid point of three years period in the rainfall trends analysis. Three year moving averages were selected because the data records were not long enough and this could give more points to plot since many points have higher accuracy. The result of the analysis was presented graphically in Figure 4.3 for Ontulili rainfall station.

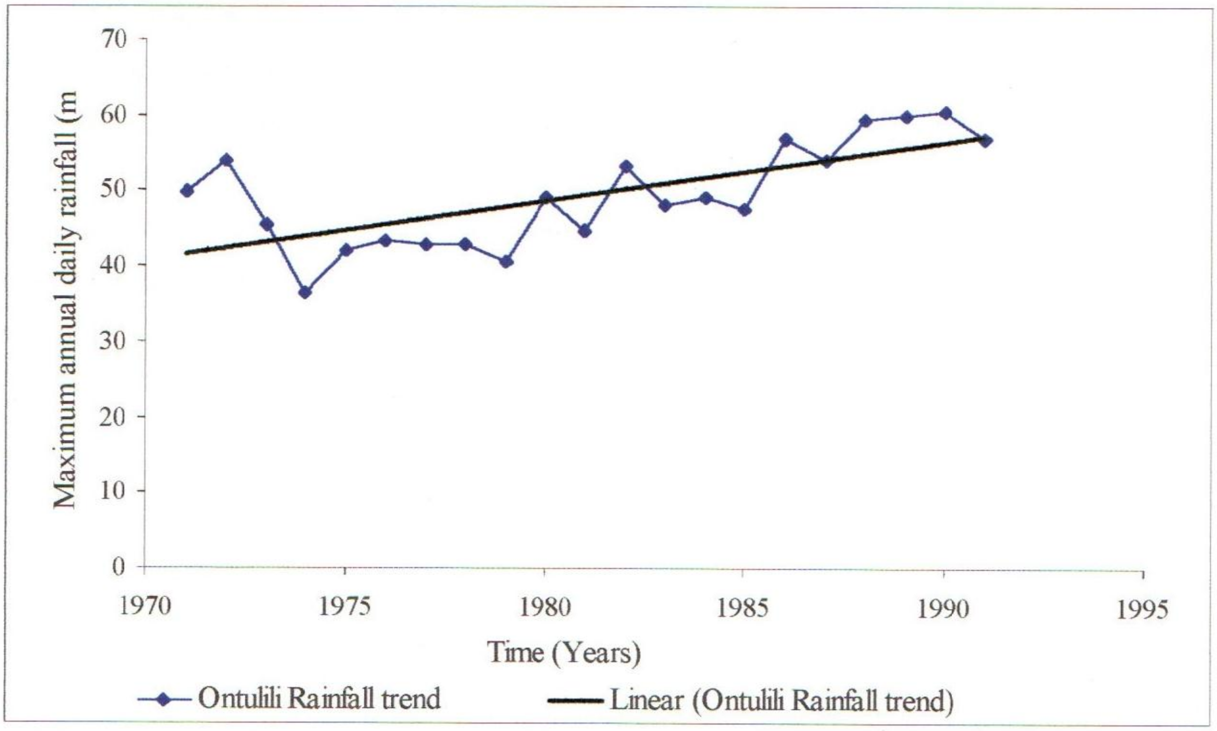

Figure 4.3. Ontulili maximum annual daily rainfall trend.

The averages of three years were preferred so that erratic variation of particular years could be minimized. From Figure 4.3 it was observed that Ontulili had increasing trend and this meant that Ontulili sub catchment was not under serious degradation affecting rainfall regime or instrument operation was not interfered with. Abrupt change of the trend line may indicate a problem at the gauging station or change in rainfall regime in that region in summary. The statistical analysis of all rainfall stations revealed that, apart from Nanyuki station which had decreasing maximum annual daily rainfall trend all other stations had increasing trend but at different scales and magnitude. The results of maximum annual daily rainfall trends of other stations were shown in Figure DI-D4 (Appendix D).

4.4 Maximum Annual Daily Rainfall Analysis

Extreme Value Type I distribution parameters were computed from data in Table 4.10 of maximum annual daily rainfall events. The parameters were fitted in EVI models and used to predict rainfall events for specified return periods. Dashes in Table 4.10 below indicated years with missing rainfall data. Some stations had shorter data periods than required thirty years for frequency analysis (Linsley and Franzini, 1979). However, they were within accepted limits since quality analysis results confirmed their accuracy. Ontulili Forest Station and Adencaple standard deviations of 12.79 and 12.80 respectively were slightly higher than Gathiuru and Sirimon. This was because the former two stations are located on lower semi-arid areas of the basin and storms are highly variable especially extreme events. The latter two stations are located on higher areas of the basin and rainfalls are more modest with less variable extreme events.

Table 4.10. Maximum Annual Daily Rainfall (mm) for selected five stations.

Year	Ontulili FS	Sirimon	Gathiuru	Nanyuki	Adencaple
1970	30.0	-	71.9	-	28.1
1971	61.0	-	42.4	-	36.1
1972	58.4	-	29.9	-	32.7
1973	42.4	-	37.0	-	28.0
1974	35.9	-	39.9	-	41.8
1975	31.1	-	49.3	-	71.0
1976	59.0	-	40.2	-	33.8
1977	40.0	-	40.3	-	50.4
1978	30.0	-	30.1	-	56.5
1979	58.4	-	69.0	-	50.3
1980	32.9	-	54.0	-	30.6
1981	56.0	-	48.6		61.2
1982	45.3	-	55.0	-	37.7
1983	58.5	56.5	49.1	-	52.5
1984	40.7	34.7	35.6	-	39.1
1985	48.0	43.7	39.2	-	41.6
1986	54.0	45.2	38.8	-	66.0
1987	68.5	52.4	54.0	-	42.2
1988	39.6	55.6	55.7	-	75.0
1989	70.0	63.2	56.0	36.0	33.1
1990	70.0	45.1	43.3	53.0	50.4
1991	41.6	39.5	72.6	25.1	43.4
1992	58.5	41.1	30.4	62.2	49.6
1993	-	44.2	44.5	43.0	48.2
1994	-	60.4	41.2	58.0	51.6
1995	-	54.8	40.4	54.7	38.2
1996	-	40.1	33.0	37.0	51.1
1997	-	46.0	44.0	57.0	68.4
1998	-	63.5	64.0	40.0	70.1
1999	-	54.0	40.3	32.5	38.4
2000	-	54.0	-	28.5	47.6
2001	-	39.2	-	45.4	39.4
Mean	**49.12**	**49.12**	**45.16**	**44.03**	**47.00**
Std. dev	**12.79**	**8.46**	**10.38**	**11.61**	**12.80**

The Extreme Value distribution has got three asymptotic forms and to select which form fitted the data, plotting was done. Rainfall events (variate, x) was plotted against reduced variate (y) calculated from rainfall data. The Type I distribution is unbounded in x, while the Type II distribution has lower bound and the Type III distribution has an upper bound (Chow et. al., 1988). The Extreme Value Type I emerged to fit the data best as in Figure 4.4.

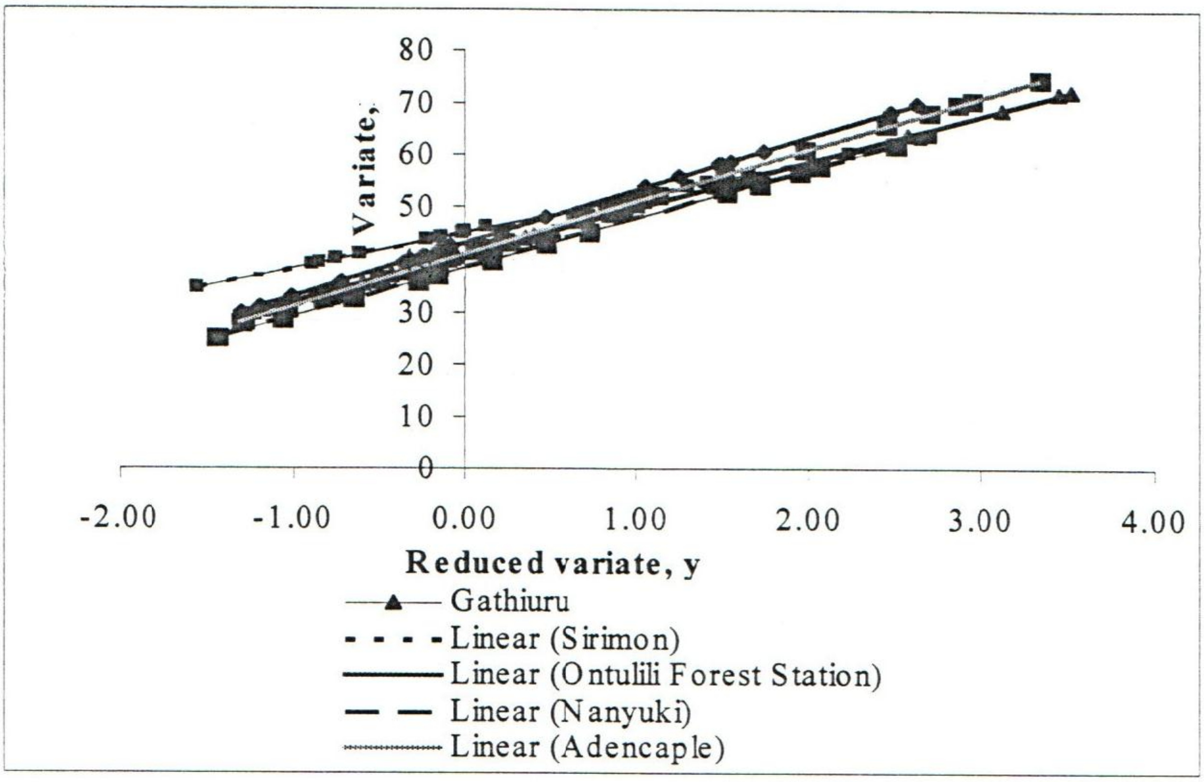

Figure 4.4 Extreme Value Type I distribution

Table 4.11 presented statistical parameters derived from Figure 4.4, linear equations with R^2 of unity indicating curves fitted were not bounded in x. The linear equations had varying gradients but this was not discussed because the focus was on bounding aspect of fitted curve.

Table 4.11 Statistical parameters from Extreme Value Type I fitted.

Station	Resulting Equation	R^2
Sirimon	$y=6.8x+45.2$	1.00
Ontulili	$y=10.2x+43.2$	1.00
Gathiuru	$y=8.9x+40.9$	1.00
Nanyuki	$y=9.4x+38.6$	1.00

The means and standard deviations of maximum annual daily rainfall in Table 4.10 were used to derive Extreme Value Type I models. The rainfall data collected were of different time periods but sufficient to compute the required model parameters. The results of the frequency models so derived are given in Table 4.12.

Table 4.12. Extreme Value Type I derived models.

Rainfall Gauge Station	Probability Frequency Model	Data Period in years
Sirimon Gate	$F(x)=\exp\left[-\exp\left(-\frac{x-45.3099}{6.5937}\right)\right.$	19
Ontulili	$F(x)=\exp\left[-\exp\left(-\frac{x-43.3663}{9.9712}\right)\right.$	23
Gathiuru	$F(x)=\exp\left[-\exp\left(-\frac{x-40.4897}{8.0957}\right)\right.$	32
Nanyuki	$F(x)=\exp\left[-\exp\left(-\frac{x-38.8041}{9.0551}\right)\right.$	13
Adencaple	$F(x)=\exp\left[-\exp\left(-\frac{x-41.2423}{9.9807}\right)\right.$	32

The derived Extreme Value Type I (EVI) distributions in Table 4.12 were used to obtain extreme rainfall events. Model parameters, significance level (α) and point of maximum probability density (μ) were calculated by formulae $\sqrt{6}S/\pi$ and $\bar{x}$ - 0.5772α respectively and substituted in probability frequency models in Table 4.12 directly. Table 4.13 gives estimated design storms in millimeters using derived frequency probability (EVI) models for specified return periods.

Table 4.13. Design storms in mm estimated using derived frequency probability models based on EVI.

Rainfall Station	2year	5year	10year	50year	100year
Ontulili	47.02	58.32	65.81	82.27	89.24
Sirimon Gate	47.73	55.20	60.15	71.04	75.64
Gathiuru	43.46	52.63	58.71	72.09	77.73
Nanyuki	42.12	52.39	59.18	74.14	80.46
Adencaple	44.90	56.21	63.70	80.19	87.16

The deviations from rainfall stations selected for the same return periods were not significantly different. For instance, for 2-year return period, the maximum annual daily rainfall magnitudes estimated were between 42.13 mm and 47.73 mm thus the highest percentage deviation was 11.7% in all the stations. For a 5-year return period maximum annual daily rainfall magnitudes ranged from 52.39 mm to 58.32 mm and highest percentage deviation was 10.2% while 10-year return period highest percentage deviation was 10.8%. Since stations spatially cover the upper catchments of Ewaso Ng'iro drainage basin, then frequency probability models could be used in storm estimation in any other sub catchment in this upper zone of the basin because simulated storms were almost the same.

The return periods of 2, 5, 10, 50 and 100 years in Table 4.13 were selected for estimation because they are normally used in catchment management practices. For example, two-year storm is used for design of soil conservation structures such as contour bands, terraces and tied ridges and other temporary structures, which are ploughed seasonally. Five-year and ten-year storms are used for cut off drains, gully control structures, bench terraces, artificial waterways and culverts. Fifty-year storms are used in the design of bridges while the hundred- year ones are usually used in reservoir design (Onyando, 2002).

The developed frequency probability models were acceptable for effective catchment management since they could be used to predict extreme rainfall events for appropriate design of structural conservation measures in the upper catchments of Ewaso Ng'iro drainage basin. The derived frequency models and hence the estimated maximum annual daily rainfall events were reasonable since they were developed from reasonably reliable historical data revealed in the quality analysis.

To verify whether the derived maximum annual daily rainfall values using established frequency probability models were within acceptable range, two techniques were applied: Intensity Duration Frequency (IDF) curves and comparison between derived and observed values. Nanyuki intensity duration frequency curves were used in this case to simulate annual daily rainfall data for specified return periods which were compared with derived annual daily rainfall data. The results of IDF curves for Nanyuki and estimated annual daily rainfall events are in Table 4.14.

Table 4.14. Maximum annual daily design storms (mm) estimated by both IDF Curves and derived frequency probability models.

Method applied	5 years	10 years	50 years	100 years
Nanyuki IDF curves	60.00	70.00	100.00	106.00
Ontulili derived model	58.32	65.81	82.27	89.24
Percentage deviation (%)	2.80	6.00	17.70	15.80

Ontulili station was chosen because it is nearer to Nanyuki town where IDF curves were available than other stations. It was therefore used to compare maximum annual daily rainfall events simulated using derived frequency probability models and Intensity Duration Frequency curves. In the two stations, the percentage deviation increased from 2.8% to 17.70% as the return periods increased from five years to fifty years because maximum annual daily rainfall events become more stochastic and in all cases the values obtained by IDF curves were higher than frequency probability simulations. The reason could be IDF curves represented infinitely small area of the basin during their derivation as opposed to frequency probability models which represent relatively larger area. The Extreme Value Type I models represented a wider area in their frequency model derivation and therefore could be more reliable than values derived by IDF curves.

In order to compare the derived maximum annual daily rainfall events and the observed values for the corresponding return periods, the data for each rainfall station were divided into two portions. The first portion was used to derive frequency model parameters and frequency model based on Extreme Value Type I and the second portion was used to give observed maximum annual daily rainfall events for comparison.

Table 4.15. Derived and observed maximum annual rainfall events (mm) used in comparison

RGS	Statistical parameters		Period in years	Derived values	Observed values	Percentage deviation
	Mean	Std deviation				
Gathiuru	45.72	12.21	15.00	65.67	72.60	9.50
Adencaple	43.21	12.60	15.00	63.81	75.00	14.90
Ontulili	45.48	11.67	15.00	64.55	70.00	7.80
Sirimon	48.43	9.10	9.00	59.52	63.50	6.30
Nanyuki	46.22	14.15	6.00	58.63	57.00	-2.80

A part from Nanyuki rainfall station with observed values lower than derived values, all other rainfall stations had observed values higher than derived values. The highest percentage deviation of 14.9% at Adencaple rainfall station could not be sufficiently explained because the data were too short and also the station is located in semi arid zone where extreme rainfall events are very erratic. However, the derived values were slightly lower than observed values in a percentage range of 6.3% to 14.9% for the same return periods. This meant that, derived values were nearly same to the observed ones and resulting slight variation could be adequately covered by factor of safety if derived rainfall data were used in design process (FAO, 1986), but otherwise comparison was within reasonable limits. Thus hydraulic structures designed based on the data from derived frequency probability models stand less risks of destruction.

4.5 Stream Flow Analysis

Data from five rivers that represented the entire catchment were used to derive parameters of Log Pearson Type III distribution model. Due to hydraulic structure failures as discussed before, some river flow data were shorter than recommended for flood frequency analysis (Linsley and Franzini, 1979). However, the frequency analysis was confined to the portion where data was continuous and homogeneous as confirmed in quality analysis. The statistical parameters calculated from the stream flow data presented in Tables A2-A5 (Appendix A) were: logarithmic sum, mean, variance and coefficient of skewness.

Several distributions have been suggested as appropriate for stream flow analysis but there is no proof of their adequacy (Linsley et. al., 1992). Based on the argument that the distributions of stream flow analysis were unlimited, Log Pearson Type III was used and resulting graphical representations are presented in Figure C1-C4 (Appendix C). Table 4.16 gives derived Log Pearson Type III distribution models for predicting discharge in upper sub catchments of Ewaso Ng'iro basin and coefficient of determination confirms their fitness.

Table 4.16. Derived Log Pearson Type III Distribution models.

RGS	Derived LP3	Data period used	Coeff. of determination
Likii	Logx= 2.03 + 0.14 k	32	0.97
Timau	Logx= -0.07 + 0.14 k	32	0.94
Nanyuki	Logx= 0.770 + 0.05 k	32	0.97
Ontulili	Logx= 0.48 + 0.11 k	20	0.96
Sirimon	Logx= 0.66 + 0.11 k	17	0.96

Where (x) is the measured stream discharge and (k) is frequency factor selected from Table A6 (Appendix A) for the computed value of coefficient of skewness and desired return period. The coefficients of determination defined as the portion of the data that fits Log Pearson Type III distribution were used to justify LP3 fitness. The logarithmic equations were preferred because of high values of coefficient of determination, ranging between 0.94 and 0.97. The continuous data periods used in the LP3 derivation were as outlined in column three of Table 4.16.

Derived Log Pearson Type III distribution models were used to simulate stream flow data for specific return periods. The estimated stream flows from derived LP3 distribution models are in Table 4.17 for some return periods. Only return periods of interest to catchment management practices were considered in the LP3 simulation.

Table 4.17. Simulated stream flows (m^3/s) using derived LP3 distribution models

RGS-catchment	2-year T	10-year T	50-year T	100-year T
Nanyuki	5.90	6.82	7.41	7.63
Timau	0.85	1.62	1.64	1.78
Ontulili	3.06	4.14	4.87	5.14
Likii	11.73	16.31	18.73	19.48
Sirimon	4.56	6.33	7.76	8.33

From Table 4.17 it can observed that, for short return period the estimated flow by derived LP3 is almost the same as observed flow for the same return period Tables A1-A5 (Appendix A). Also as the return period increases the estimated value by derived LP3 models decreases in all stream flow data. This implies that, peak discharges for longer periods of time are so erratic and variable hence their prediction is not reasonable. This phenomenon can be controlled by using mean annual daily discharges that minimizes the effect of peak daily discharge variability.

4.6 Long Term Flow coefficient (K_{95})

Long-term flow coefficient (K_{95}) has been used in hydrology studies to establish catchment characteristic functions. It can be used to simulate discharge data where the catchment has not been gauged but some mean catchment characteristics are known (Molner, 1988). Through correlation and regression analysis of long-term flow coefficient (K_{95}) with catchment characteristics namely average elevation, catchment slope and maximum annual daily rainfall regression equations were derived. The equations can be used to determine 95% discharge percentile in ungauged catchment or any other discharge percentile of interest. The catchment characteristics selected were based on the data availability and consistency of the available data.

The average slope was determined by getting slope at different parts of the catchment down the slope. The slope at particular point down the slope was obtained by getting sine of angle formed between differences in contours (vertical interval) with scale distance along the slope. The long term flow coefficient was obtained by dividing discharge at 95% percentiles with average discharge in Tables A2-A5 (Appendix A) ranked in descending order of magnitude. The results of correlation and regression of k_{95} with catchment characteristic elements are indicated in Table 4.18.

Table 4.18. Correlation and Regression Analysis of K_{95} with catchment characteristics.

Station	Slope, (%)	Elevation (m)	Maxi daily annual r/f (mm)	(K_{95}) Values
Sirimon	5.1	2579.2	49.12	1.23
Ontulili	4.4	2516.7	49.12	1.24
Nanyuki	5.5	2586.7	44.03	1.16
Timau	2.7	2266.7	-	1.48
Likii	4.9	2553.3	-	1.24
Correl R	-0.98	-0.98	0.99	
R^2	0.97	0.95	0.99	
Regre-ssion	**$K_{95} = -0.43\ln(S) + 1.9$**	**$K_{95} = -0.01\ln(E) + 3.52$**	**$K_{95} = 0.10(R)^{0.64}$**	

Where (S) is average percentage slope, (E) is elevation in meters and (R) is maximum annual daily rainfall in millimeters. The regression equations were obtained by plotting k_{95} values against catchment characteristics and these graphical representations are shown in Figures B1-B3 (Appendix B). The coefficient of determination for the average slope and long- term flow coefficient was 0.97, indicating a satisfactory relationship between the two variables. The same relationships were obtained for elevation (R^2=95%) and mean annual daily rainfall (R^2=99%).

Regression analysis indicated that, the governing equation for slope was logarithmic while for elevation was linear. Finally, maximum annual daily rainfall emerged to relate by power form of equation as in Table 4.18. Generally, regression equations are functions of discharge at 95 % and average discharge and thus having average discharge for a particular stream it is possible to compute discharge of any percentile of interest. The application of long term flow coefficient can be useful in ungauged sub catchments within the catchment where long-term flow coefficient has been established since generation of discharge at 95% percentile or any other percentile is possible.

5 CONCLUSION AND RECOMMENDATION

5.1 Conclusion

Double mass curve technique revealed homogenous and consistent collected rainfall and stream flow data by showing a straight line plot through multiple regressions. Positive double mass curve gradients and coefficient of determination close to unity confirmed good quality data for engineering purposes. Still double mass curve plots indicated Gathiuru sub catchment receives highest rainfall and Nanyuki sub catchment the least. Also by same technique it was found that, Likii river had highest flow while Sirimon river the least. From multiple rainfall correlations, Sirimon rainfall station was found to correlate negatively (-0.190) with Nanyuki rainfall station and significantly different from all other rainfall stations. Similarly from stream flow multiple correlations, all probability values (p-values) from stream flow data were less than set probability values by Bivariate model with 2-tailed test. This meant that in stream flow stations, data were positively correlated and not significantly different.

Nanyuki and Ontulili rainfall station data with p-value of zero were found to be of the same kind and frequency and hence simple proportion method was applied to calculate missing data. Sirimon missing data were computed by weighting factor method since its rainfall data were significantly different from others. Apart from Nanyuki rainfall station, all other stations had increasing rainfall trend with time. The Extreme Value distribution fitting gave a straight line plot not bounded in x-axis, indicating rainfall data adequately suited Extreme Value Type I distribution. The EVI probability models were derived and used to predict some storm events of specific return periods. Two year return period estimates in all rainfall stations gave percentage deviation of 11.7 %. Estimates for 5- year and 10- year return periods were 10.2 % and 10.8 % respectively. This meant that, any of those derived EVI models could be used in ungauged sub catchments in upper Ewaso Ng'iro drainage basin to generate rainfall events of 2, 5 or 10 years return periods.

To verify EVI model predicted values, Intensity Duration Frequency curves of Nanyuki were used to derive maximum annual daily rainfall events. The comparison between the two indicated that IDF curve estimated values were higher than derived EVI model estimates and percentage deviation with time increased from 2.8 % to 17.7 % on 2- year and 50- year return periods respectively. This meant that as return period increased, the maximum annual daily rainfall events became more erratic and unpredictable. However, IDF curve estimation and EVI model estimation comparison with time gave a reasonable result. Also comparison between observed and derived rainfall events using derived EVI models was satisfactory with percentage deviation of 9.5 % in all the rainfall stations a part from Adencaple station located in semi arid zone where extreme rainfall regimes are more variable. The Log Pearson Type III fitness was confirmed by coefficient of determination nearly to unity. The LP3 models were derived and used to derive stream flows for specified return periods.

Correlation and regression analysis of long term flow coefficients with catchment characteristics were reasonable with R^2 above 95 % in all the river gauging stations. Regression equations for deriving k_{95} given the catchment characteristic variables were obtained. Finally, through analysis of the available data from the upper catchments of Ewaso Ng'iro drainage basin, all proposed objectives were answered and the results were reasonable. The suggested hypotheses were also tested and in all cases, the null hypotheses were rejected and alternative hypotheses accepted. Finally from the study findings, it was acceptable to use the derived probability models in the upper parts of the basin to estimate rainfall and stream flow data.

5.2 Recommendation

Based on the results of the study, it seemed there was a lot to be done on these upper sub catchments of Ewaso Ng'iro drainage basin. Therefore, the following recommendations were made for further investigation:

1. Further study is necessary to find the causes of decreasing maximum annual daily rainfall trend in Nanyuki sub catchment.
2. More research is required to explain why percentage deviation between derived and observed maximum annual daily rainfall and stream flow events increased as return period increased.
3. A research can also be conducted to find out why Sirimon maximum annual daily rainfall events were significantly different from other rainfall stations with explained variance of 29.3 %.
4. The use of long term flow coefficient (K_{95}) should be investigated further with more catchment characteristics.

6. REFERENCES

Berger, P., 1989. Rainfall and Agro-climatology of the Laikipia plateau, Kenya. *African studies series* Vol.7. *Geograhica Bernensia and Geophysical Society* of Berne, university of Berne, Switzerland. Pp 19-167.

Chemelil, M.C. and Smout, I.K., 2000. Validation and quality control of Hydrological data series. J. of civil Engineering, JKUAT. Vol.5.

Chow, V.T., Maidment, D.R. and Mays, L.W., 1988. Applied Hydrology. International edition. *McGraw-Hill Book Company* New York. Pp 381-415.

Clark, R.T., 1973. Mathematical models in Hydrology. Published by *Food and Agriculture organization* of the United Nations Rome. Pp 2-18.

Decurtins, S., 1992. Hydrogeographical investigations in the mount of Kenya Subcatchment of the Ewaso Ng'iroRiver. *African studies series* A10. University of Berne, Switzerland. Pp 12-80.

FAO., 1986. Natural Resource and human Environment for food and Agriculture in Africa. Environment and Energy paper No. 6. *FAO*, Rome.

Gichuki, J., 1998. Flood utilization study. Prepared in collaboration with Ministry of Land Reclamation, Regional and Water Development. Laikipia District, Kenya. Unpublished.

Hendriks, M.J., 1983. Stochastic models and calculation of minimum discharges of the Ewaso Ng'iro River. Published by *Wageningen*.Pp 5-10.

Rohr, P.C. and Killingtveit, A., 2003. Rainfall distribution in the slopes of Mt. Kilimanjaro. *HSJ*. 48(1):65-77.

Kimambo, I.N., 1989. The Eastern Bantu people. In: Oggot, B.A(ed): A survey of East Africa history. *East African* publishing House.

Lee, C., Tan, Y., Chen, C. and Yeh, T.J. 2001. Stochastic series lumped rainfall-runoff model for a watershed I Taiwan. Journal of Hydrology 249: 30-45.

Liniger, J., 1998. Development and use of long and short-term precipitation interpolation models in the Ewaso Ng'iro Basin, Kenya. Annual Report.

Linsely, R. k. and Franzini, J.B., 1979. *Water Res. Engineering*. Third Edition. *McGraw-Hill*, New York. Pp 112-134.

Linsley, R.K., Kohler, M.A. and Paulhus, J.H., 1992. Hydrology for Engineers.Third edition. *McGraw-Hill international Book Company*, Singapore. Pp 110-134.

Loukas, A., Michael, C. and Russell, O., 1995. A physically based stochastic-deterministic procedure for the Estimation of Flood Frequency. *Water Res. management.*10: 415-437.

Mati, B.M., Gichuki, F.N., Quinton, J.N., Brewer, T.R. and Liniger, H., 2000. Assessment of Erosion Hazard with the USLE and GIS: A case study of the upper Ewaso Ng'iro North basin of Kenya.*JAG.*Volume 2.

McMahon, T.A. and Nathan, R.J., 1991. Estimating low flow characteristics in ungauged catchments. *Water res.management.*6: 85-100.

Meunier, M. 2001. Regional flow-duration-frequency model for the tropical island of Martinique. Journal of Hydrology 247: 31-53.

Molnar, L., 1988. Hydrology of mountainous areas. *IAHS* publication No. 90.

Muruiki, G., 1989. A History of the Kikuyu to 1900. Ph.D. thesis, *university of Nairobi,* Kenya.

Onyando, J.O., 2000. Rainfall –Runoff Models for ungauged catchments in Kenya. PhD Thesis, *Ruhr University*. Germany.

Onyando, J.O., 2002. Preliminary Hydrological study of Lake Baringo drainage basin. Consultancy report. UNEP/GEF Lake Baringo community based and water management project.

Onyando, J.O., Chemelil, M.C., Awer, M. and Kisoyan, P., 2004. Development of frequency models for predicting design storms in Lake Baringo. Bostwan J. of Technology. 13(1):1-6.

Shahin, M., Van Oorschot, H.J.L. and Delange, S.J. 1993. Statistical Analysis in Water Resources Engineering. Published by *A.A. Belkema*, Netherlands. Pp 210-280.

Sharma, T.C. and Makhoalibe, S., 1987. A regional frequency study of Extremal rainfall and runoff processes. *Water res. Management.*1: 223-235.

Sheng Yue., 2000. The Gumbel Mixed Model Applied to Storm Frequency Analysis. *Water Res, Management.* 14: 377-389.

Sombroek, W.G., Braun, H.M.H. and Van der Pouw, B.J.A. 1980. Exploratory soil map and agro- climatic zone map of Kenya. Kenya soil survey, Ministry of Agriculture, Nairobi, Kenya. Caroprint BV, Soil survey Institute, The Hague, The Netherlands.

Steel, R.G. and Torry, J.H., 1981. Principles and procedure of STATISTICS (Second edition). International edition, *McGraw –Hill Book Company*, Singapore. Pp 311-333.

Wilson, E.M., 1975. Egineering hydrology (second edition). *William Clowes nd Sons* Ltd. London. Pp 18-30.

APPENDICES

NOTES FOR TABLES A1-A5.

The following symbols were used in Tables A1 – A6 and meant the following:

symbol	meaning
X	Discharge in m^3/s
m	Rank of stream flow event in the list ordered by descending magnitude
T	Return interval in years
K	Frequency factor values

Formulae used are as outlined below:

$$\text{Probability (\%)} = \left\lfloor 1-(1-p)^N * 100\% \right\rfloor$$

Where p is exceedence probability calculated as $\frac{1}{T}$, where also T is the return interval.

Return interval is defined as the average interval in years between the occurrence of a stream flow of specified magnitude and an equal or larger stream flow and is computed as follows:

$T = \frac{N+1}{m}$, where N is the number of years in the record and m is the rank of stream flow event in the list ordered by descending magnitude.

Appendix A (Tables)

Table A1: Annual peak flows (m^3/s) for Timau River (1970-2001).

Year	Discarge (X)	m	T	log X	Prob-%
1994	4.91	1	33.00	0.69	3.03
1981	4.41	2	16.50	0.64	6.06
1998	2.69	3	11.00	0.43	9.09
1997	2.17	4	8.25	0.34	12.12
1988	2.11	5	6.60	0.32	15.15
1989	1.67	6	5.50	0.22	18.18
1991	1.55	7	4.71	0.19	21.21
1986	1.49	8	4.13	0.17	24.24
1996	1.46	9	3.67	0.16	27.27
1982	1.41	10	3.30	0.15	30.30
1997	1.39	11	3.00	0.14	33.33
2000	1.38	12	2.75	0.14	36.36
1993	1.3	13	2.54	0.11	39.39
1992	1.19	14	2.36	0.08	42.42
1979	1.06	15	2.20	0.03	45.45
1990	1	16	2.06	0.00	48.48
1987	0.94	17	1.94	-0.03	51.52
2001	0.88	18	1.83	-0.06	54.55
1995	0.8	19	1.74	-0.10	57.58
1975	0.8	20	1.65	-0.10	60.61
1999	0.64	21	1.57	-0.19	63.64
1974	0.63	22	1.50	-0.20	66.67
1978	0.48	23	1.43	-0.32	69.70
1983	0.43	24	1.38	-0.37	72.73
1985	0.38	25	1.32	-0.42	75.76
1971	0.37	26	1.27	-0.43	78.79
1976	0.31	27	1.22	-0.51	81.82
1980	0.28	28	1.18	-0.55	84.85
1972	0.26	29	1.14	-0.59	87.88
1984	0.23	30	1.10	-0.64	90.91
1970	0.21	31	1.06	-0.68	93.94
1973	0.18	32	1.03	-0.74	96.97
sum				**-2.09**	
mean				**-0.07**	
variance				**0.14**	
skewness				**-0.05**	

Table A2: Annual peak flows (m^3/s) for Likii River (1970-2001).

Year	Discharge (X)	m	T	LogX	Prob-%
1986	48.01	1.00	33.00	1.68	3.03
1988	45.62	2.00	16.50	1.66	6.06
1989	41.16	3.00	11.00	1.61	9.09
1998	28.40	4.00	8.25	1.45	12.12
2000	23.59	5.00	6.60	1.37	15.15
1991	22.08	6.00	5.50	1.34	18.18
1997	21.70	7.00	4.71	1.34	21.21
1990	20.87	8.00	4.13	1.32	24.24
1987	20.07	9.00	3.67	1.30	27.27
1996	17.68	10.00	3.30	1.25	30.30
1981	15.34	11.00	3.00	1.19	33.33
1979	15.34	12.00	2.75	1.19	36.36
1994	15.21	13.00	2.54	1.18	39.39
1977	14.92	14.00	2.36	1.17	42.42
1995	14.49	15.00	2.20	1.16	45.45
2001	11.61	16.00	2.06	1.06	48.48
1993	10.38	17.00	1.94	1.02	51.52
1971	9.00	18.00	1.83	0.95	54.55
1992	8.62	19.00	1.74	0.94	57.58
1972	8.60	20.00	1.65	0.93	60.61
1982	8.50	21.00	1.57	0.93	63.64
1975	7.21	22.00	1.50	0.86	66.67
1999	6.77	23.00	1.43	0.83	69.70
1983	6.30	24.00	1.38	0.80	72.73
1973	6.01	25.00	1.32	0.78	75.76
1984	6.01	26.00	1.27	0.78	78.79
1970	5.78	27.00	1.22	0.76	81.82
1978	5.30	28.00	1.18	0.72	84.85
1974	5.23	29.00	1.14	0.72	87.88
1980	4.63	30.00	1.10	0.67	90.91
1976	4.63	31.00	1.06	0.67	93.94
1985	0.67	32.00	1.03	-0.17	96.97
sum				**33.46**	
mean				**1.05**	
variance				**0.14**	
skewness				**-0.83**	

Table A3: Annual peak flows (m^3/s) for Nanyuki River (1985-2000).

Year	Discharge (X)	m	T	LogX	Prob %
1985	15.00	1	33.00	1.18	3.03
1977	13.49	2	16.50	1.13	6.06
1982	11.75	3	11.00	1.07	9.09
1988	11.01	4	8.25	1.04	12.12
1981	10.47	5	6.60	1.02	15.15
1979	10.17	6	5.50	1.01	18.18
1998	9.89	7	4.71	1.00	21.21
1997	9.61	8	4.13	0.98	24.24
1996	8.36	9	3.67	0.92	27.27
1995	7.88	10	3.30	0.90	30.30
1986	7.56	11	3.00	0.88	33.33
1993	7.47	12	2.75	0.87	36.36
1983	6.79	13	2.54	0.83	39.39
1992	6.62	14	2.36	0.82	42.42
1994	6.59	15	2.20	0.82	45.45
1991	6.51	16	2.06	0.81	48.48
2001	5.76	17	1.94	0.76	51.52
1973	5.73	18	1.83	0.76	54.55
1989	5.27	19	1.74	0.72	57.58
1990	5.05	20	1.65	0.70	60.61
1999	4.98	21	1.57	0.70	63.64
1987	4.82	22	1.50	0.68	66.67
1978	4.46	23	1.43	0.65	69.70
1970	4.37	24	1.38	0.64	72.73
1976	3.83	25	1.32	0.58	75.76
1975	3.68	26	1.27	0.57	78.79
1972	3.27	27	1.22	0.51	81.82
1971	3.05	28	1.18	0.48	84.85
1980	2.91	29	1.14	0.46	87.88
1984	2.80	30	1.10	0.45	90.91
1974	2.60	31	1.06	0.41	93.94
2000	2.27	32	1.03	0.36	96.97
sum				**24.72**	
mean				**0.77**	
variance				**0.05**	
skewness				**-0.10**	

Table A4: Annual peak flows (m^3/s) for Ontulili River (1970-1987).

Year	Discharge (X)	M	T	LogX	Prob %
1977	9.5	1	19.00	0.98	5.26
1979	8.84	2	9.50	0.95	10.53
1981	7.34	3	6.33	0.87	15.79
1986	6.23	4	4.75	0.79	21.05
1982	6.01	5	3.80	0.78	26.32
1975	6.01	6	3.17	0.78	31.58
1987	3.57	7	2.71	0.55	36.84
1971	3.47	8	2.38	0.54	42.11
1983	3.15	9	2.11	0.50	47.37
1972	2.83	10	1.90	0.45	52.63
1974	2.4	11	1.73	0.38	57.89
1978	2.14	12	1.58	0.33	63.16
1976	2.01	13	1.46	0.30	68.42
1970	1.92	14	1.36	0.28	73.68
1973	1.67	15	1.27	0.22	78.95
1984	1.37	16	1.19	0.14	84.21
1980	1.27	17	1.12	0.10	89.47
1985	0.56	18	1.06	-0.25	94.74
sum				**8.69**	
mean				**0.48**	
variance				**0.11**	
skewness				**-0.31**	

Table A5: Annual peak flows (m3/s) for Sirimon River (1970-1984).

Year	Discharge (X)	M	T	LogX	Prob %
1982	14.91	1	16.00	1.17	6.25
1981	14.65	2	8.00	1.17	12.50
1977	10.52	3	5.33	1.02	18.75
1974	7.29	4	4.00	0.86	25.00
1975	7.08	5	3.20	0.85	31.25
1971	6.93	6	2.67	0.84	37.50
1978	5.68	7	2.29	0.75	43.75
1976	4.01	8	2.00	0.60	50.00
1972	3.59	9	1.78	0.56	56.25
1973	3.45	10	1.60	0.54	62.50
1970	2.93	11	1.45	0.47	68.75
1983	2.49	12	1.33	0.40	75.00
1980	2.23	13	1.23	0.35	81.25
1979	1.53	14	1.14	0.18	87.50
1984	1.37	15	1.07	0.14	93.75
Sum				**9.90**	
Mean				**0.66**	
Variance				**0.11**	
skewness				**0.06**	

Table A6: Values of K for use with the Log Pearson Type III Distribution.

	Recurrence interval, Years					
	2	10	25	50	100	200
Skew coefficient	Chance, %					
g	50	10	4	2	1	0.5
2.0	-0.31	1.30	2.22	2.91	3.61	4.29
1.8	-0.28	1.32	2.19	2.85	3.50	4.15
1.6	-0.25	1.33	2.16	2.78	3.39	3.99
1.4	-0.23	1.34	2.13	2.71	3.27	3.83
1.2	-0.19	1.34	2.09	2.63	3.15	3.66
1.0	-0.16	1.34	2.04	2.54	3.02	3.49
0.9	-0.15	1.34	2.02	2.49	2.96	3.40
0.8	-0.13	1.34	1.99	2.45	2.89	3.31
0.7	-0.12	1.33	1.97	2.41	2.82	3.22
0.6	-0.09	1.33	1.94	2.36	2.75	3.13
0.5	-0.08	1.32	1.91	2.31	2.69	3.04
0.4	-0.07	1.32	1.88	2.26	2.62	2.95
0.3	-0.05	1.31	1.85	2.21	2.54	2.86
0.2	-0.03	1.30	1.82	2.16	2.47	2.76
0.1	-0.02	1.29	1.79	2.11	2.40	2.67
0	0	1.282	1.751	2.054	2.326	2.576
-0.1	0.017	1.270	1.716	2.000	2.252	2.482
-0.2	0.033	1.258	1.680	1.945	2.178	2.388
-0.3	0.050	1.245	1.643	1.890	2.104	2.294
-0.4	0.066	1.231	1.606	1.834	2.029	2.201
-0.5	0.083	1.216	1.567	1.777	1.955	2.108
-0.6	0.099	1.200	1.528	1.720	1.880	2.016
-0.7	0.116	1.183	1.488	1.663	1.806	1.926
-0.8	0.132	1.166	1.448	1.606	1.733	1.837
-0.9	0.148	1.147	1.407	1.549	1.660	1.749

Appendix B (Regression Figures of K_{95} and Catchment characteristics)

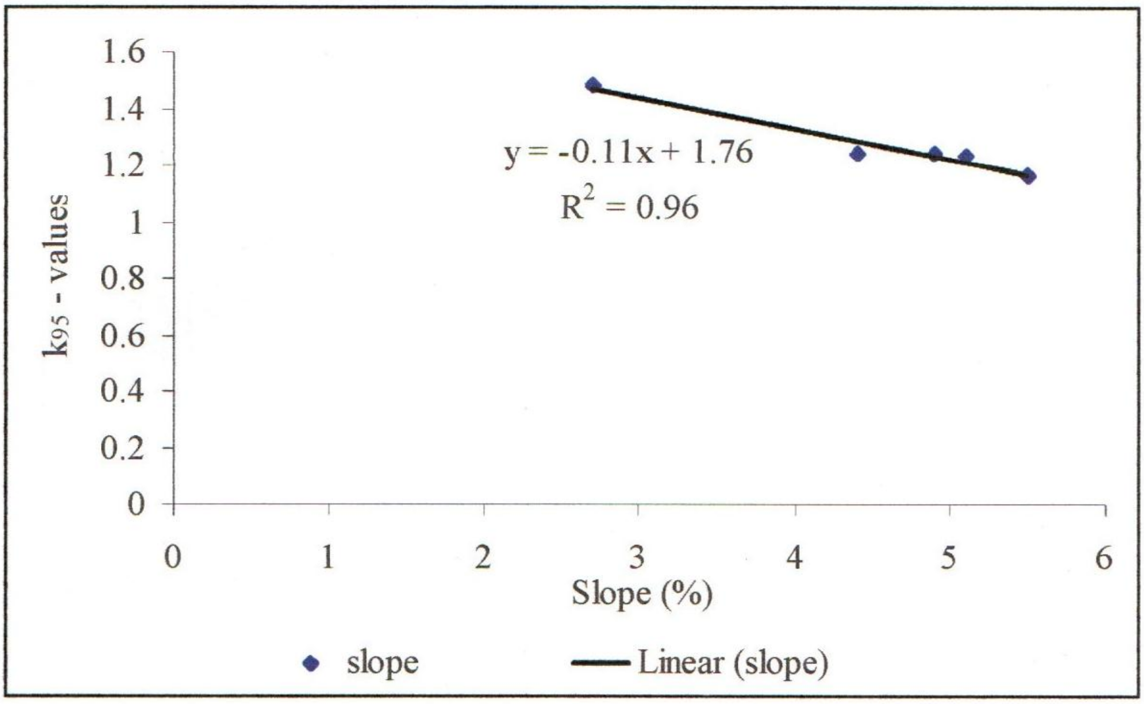

Figure B1: Regression graph of percentage slope against k_{95} values.

NOTES FOR FIGURES B1-B3.

From Tables B1-B3, symbol y represents long term flow coefficient (k_{95}) and x the catchment characteristic (slope, elevation or maximum annual daily rainfall). The regression equations were derived by plotting long term flow coefficient against catchment characteristic variable.

Long term flow coefficient (k_{95}) $=Q_{95}/Q_{Avg}$, where Q_{95} is the discharge at 95 % percentiles got by ranking discharges in ascending order and Q_{Avg} is the average discharge for the selected period. These two discharges were derived form Tables A1-A5 (Appendix A). The criterion of selecting the type of regression equation to be used for a particular catchment characteristic was based on highest value of coefficient of determination (R^2).

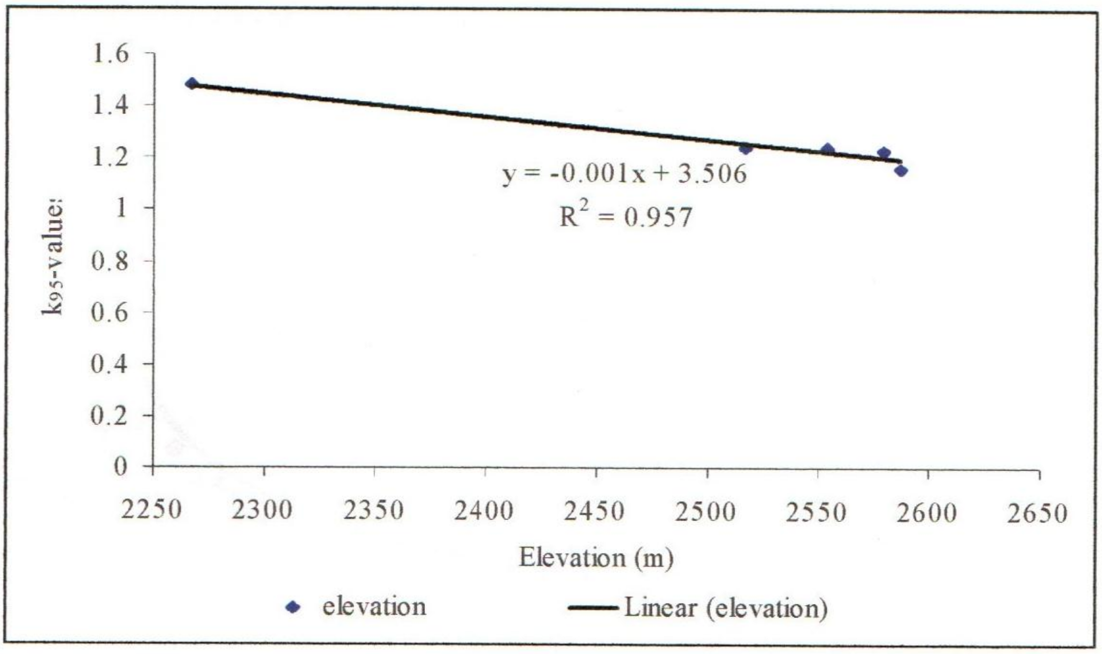

Figure B2: Regression graph of watershed elevation against k_{95} values.

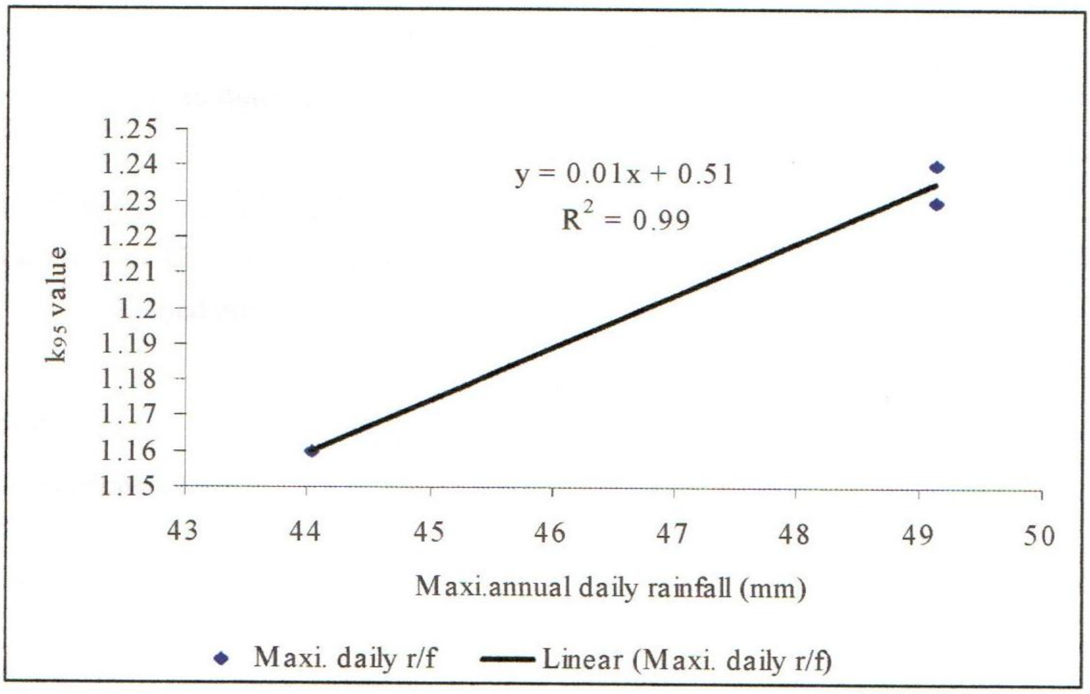

Figure B3: Regression graph of maximum annual daily rainfall and k_{95} values.

Appendix C: (Figures of fitted Log Pearson Type III distribution)

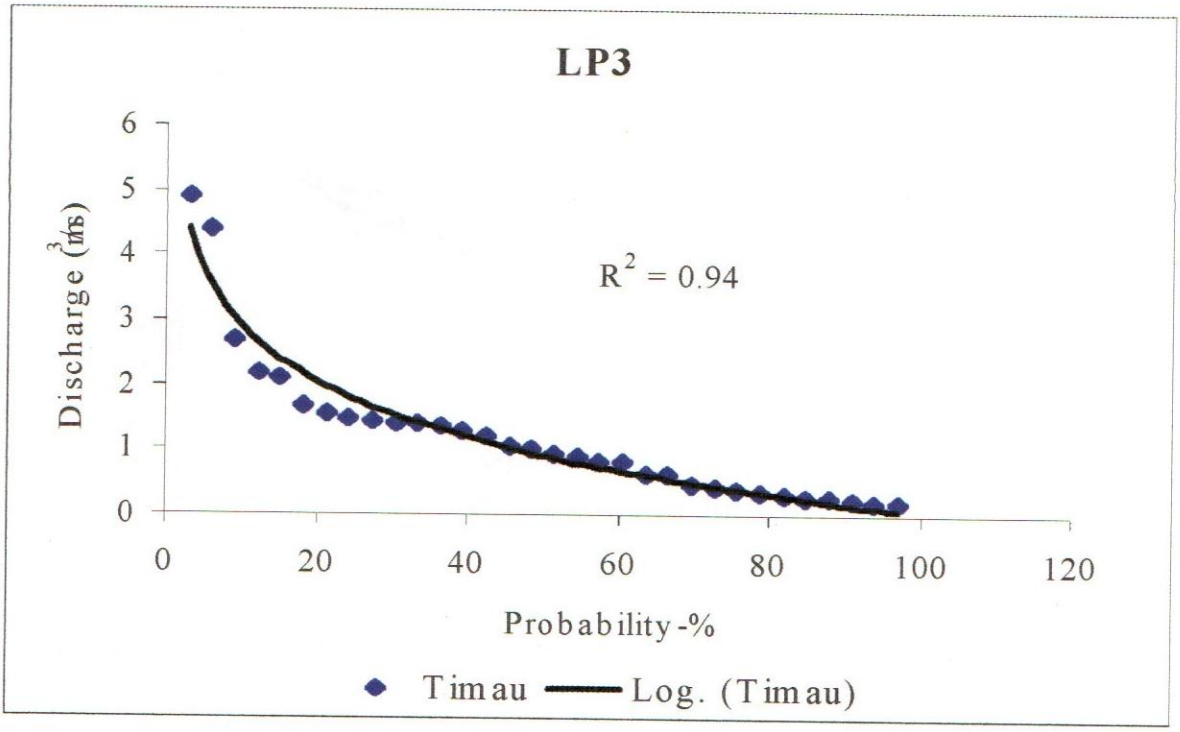

Figure C1: Timau flows in Kenya, 1970-2001.

NOTES FOR FIGURES C1-C4.

Log Pearson Type III model (LP3) was obtained by plotting discharge (m^3/s) against percentage probability as outlined previously.

Probability (%) = $\left\lfloor 1-(1-p)^{N} * 100\% \right\rceil$

The values of percentage probabilities are shown in the last columns of Tables A1-A5.

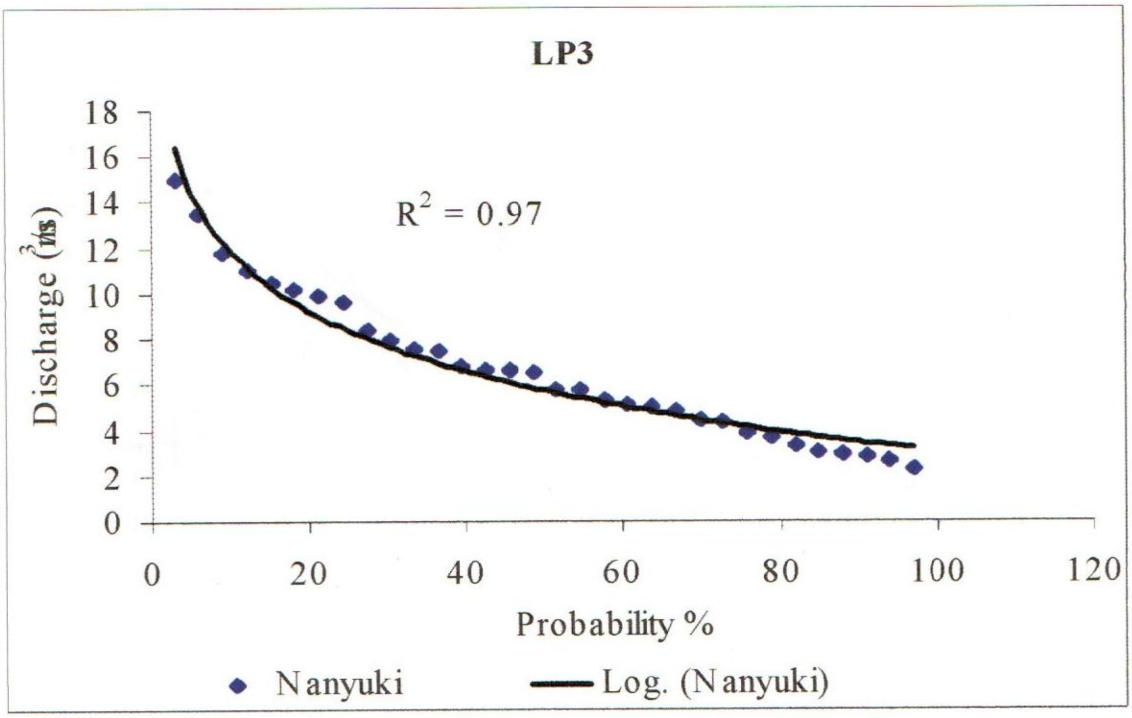

Figure C2: Nanyuki flows in Kenya, 1970-2001.

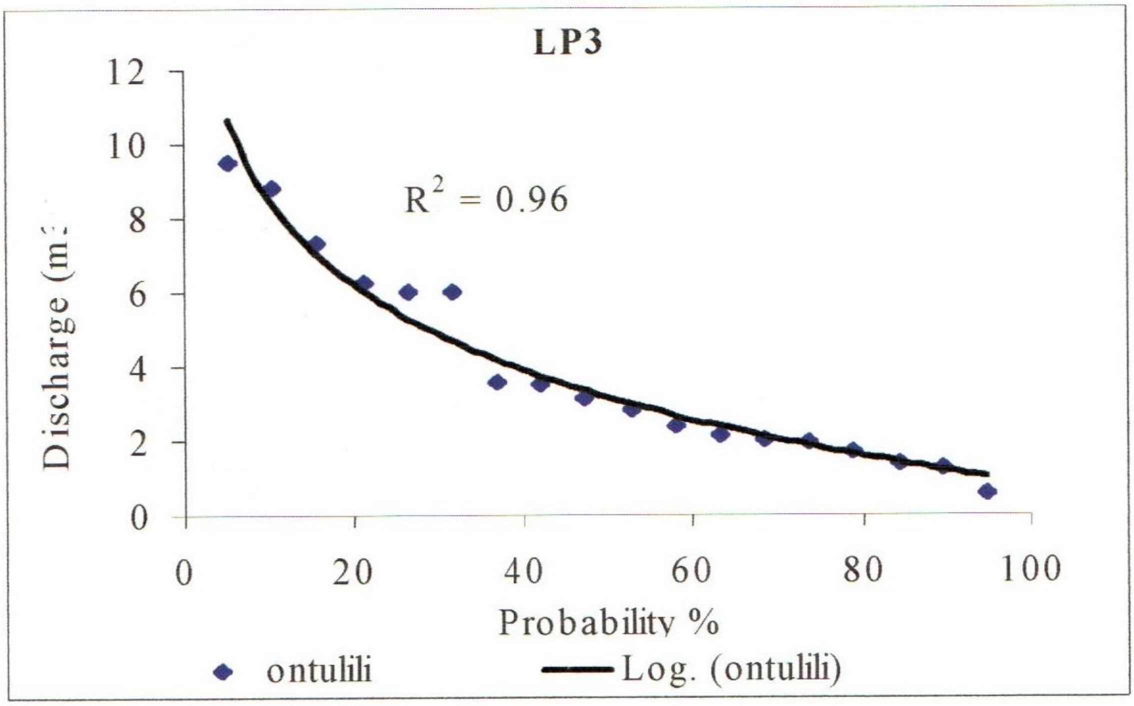

Figure C3: Ontulili flows in Kenya, 1970-1987.

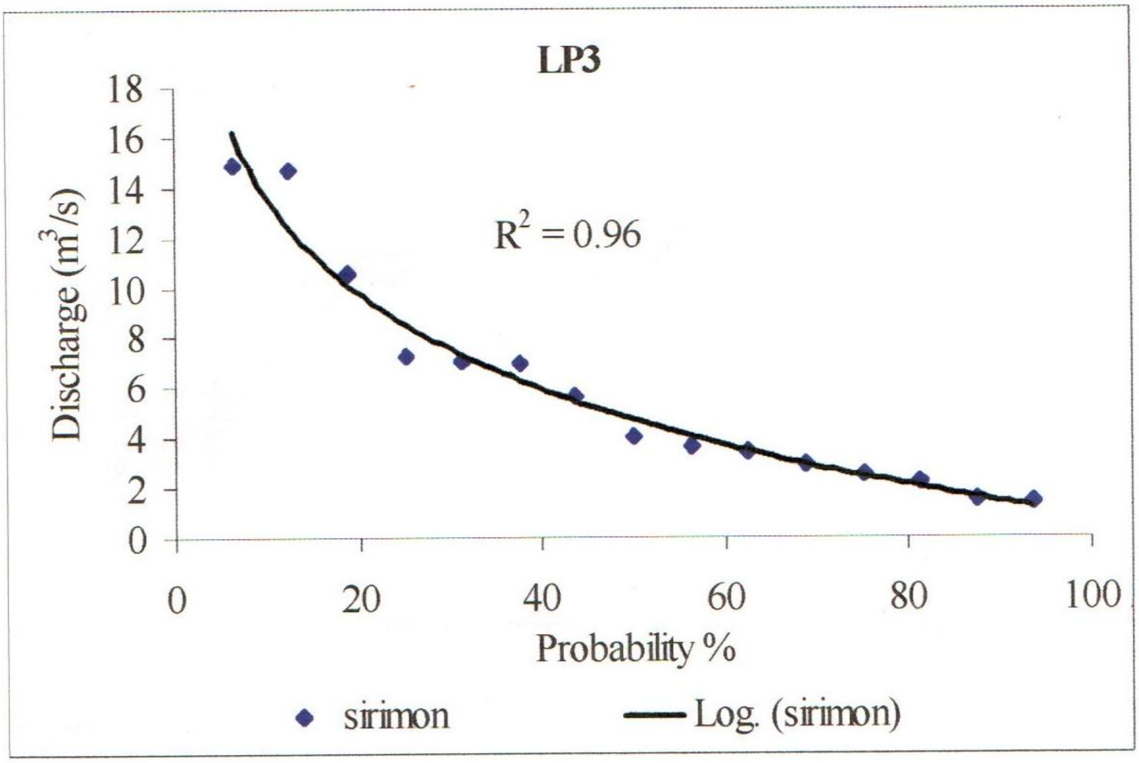

Figure C4: Sirimon flows in Kenya, 1970-1987.

Appendix D: (Figures of Maximum Annual Daily Rainfall Trends)

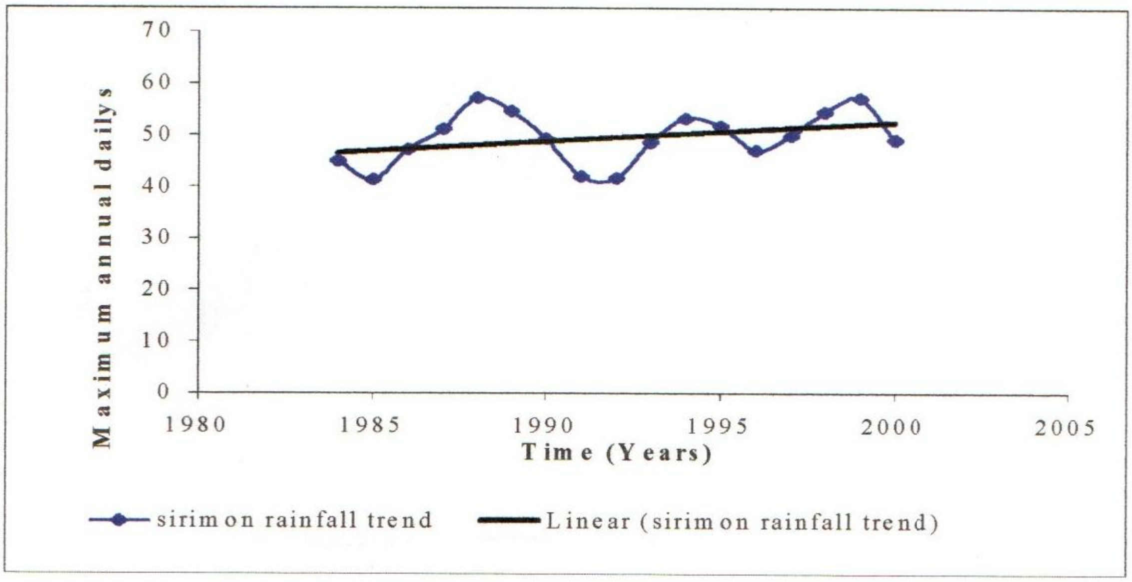

Figure D1: Sirimon maximum annual daily rainfall trend.

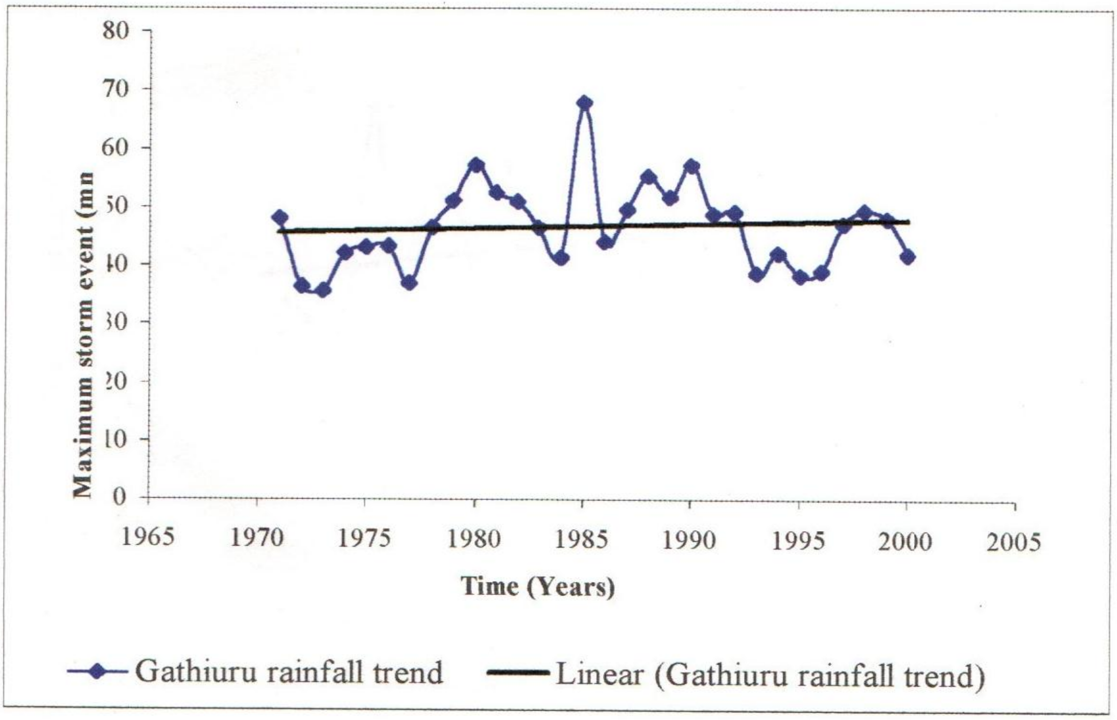

Figure D2: Gathiuru maximum annual daily rainfall trend.

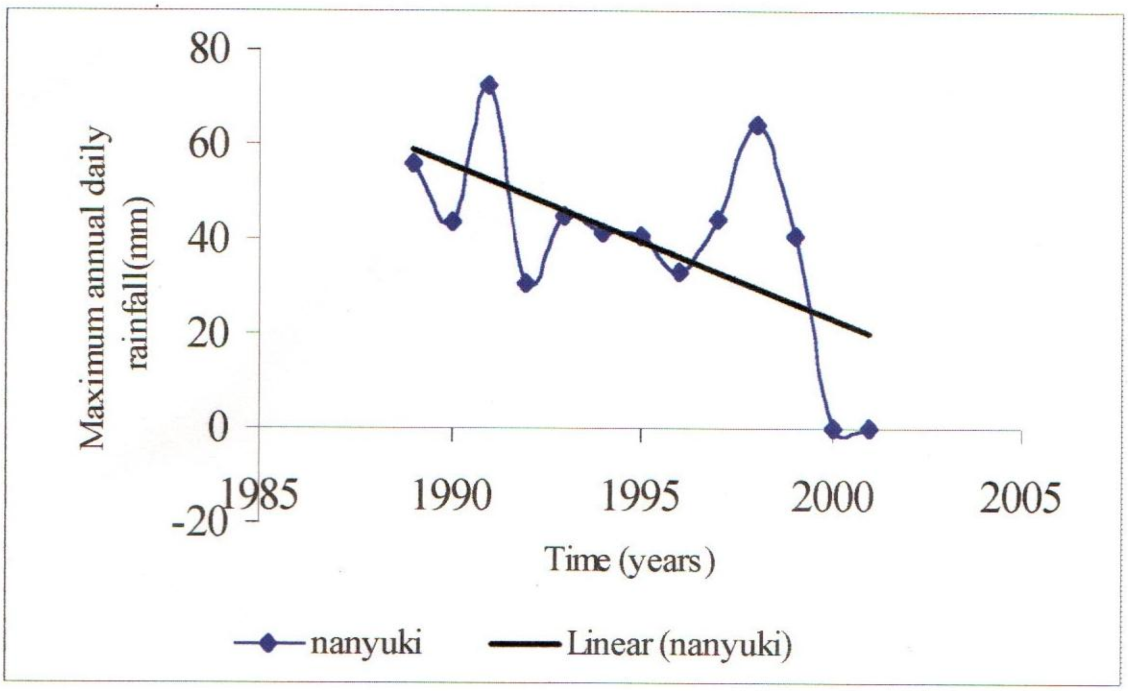

Figure D3: Nanyuki maximum annual daily rainfall trend.

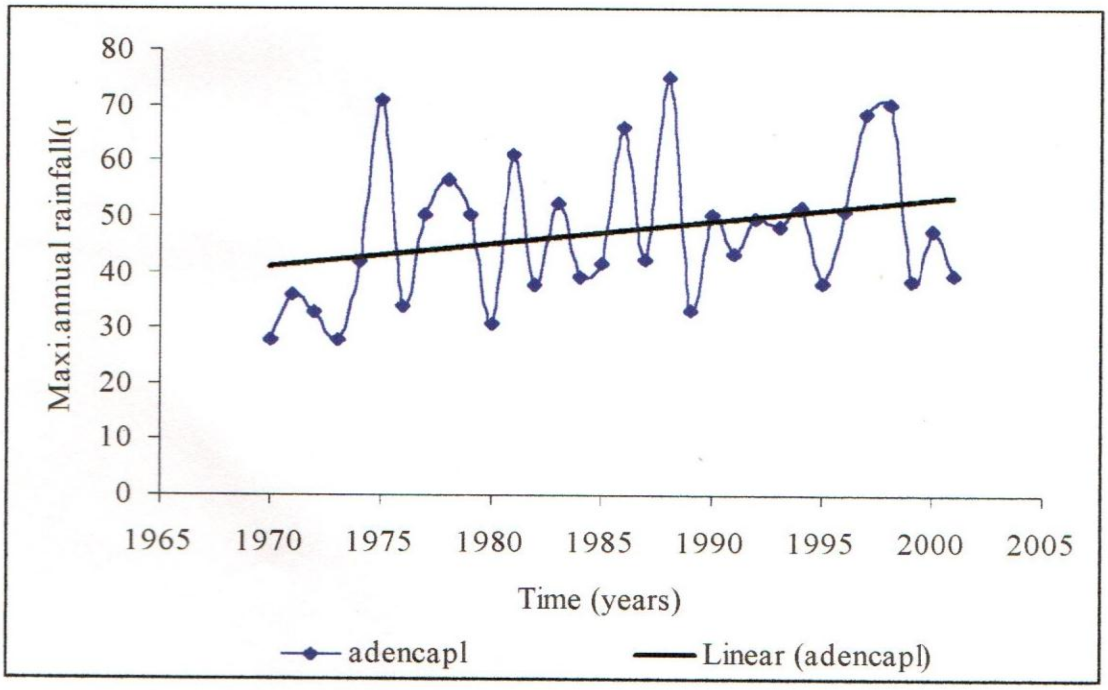

Figure D4: Adencaple maximum annual daily rainfall trend.