3000/=

Estimation of Stream Flows in Ungauged Catchments of the Upper Tana Basin, Kenya

By

John Kimani Mwangi

A thesis submitted in partial fulfilment for the award of the degree of **Master of Science in Hydrology and Water Resources** of Kenyatta University

2007

DECLARATION

I, John Kimani Mwangi, do here-by declare that this thesis is my original work and has not been presented for a degree in any other university.

Signature: Mwangi ------------------------------ Date 23/02/07 --------------------------------

John Kimani Mwangi
Department of Geography,
Kenyatta University

This thesis has been submitted for examination with our approval as university supervisors.

Signature: Ondieki ------------------------------ Date 23/02/07 --------------------------------

Dr. Christopher M. Ondieki
Department of Geography,
Kenyatta University

Signature: -- Date 23/02/2007 ------------------------------------

Dr. John M. Gathenya
Department of Biomechanical and Environmental Engineering,
Jomo Kenyatta University of Agriculture and Technology

DEDICATION

To my mother Wambui Mwangi for all the sacrifice

To my wife Joyce and children Michael, Kevin and Stella
You have provided the shoulders to lean on during the hard times

To all my friends who have stood by me in times of tribulations

ACKNOWLEDGEMENTS

All thanks to the Almighty for the gift of life.

I wish to extend my gratitude and appreciation to my supervisors Dr. Christopher Ondieki of Kenyatta University and Dr. Mwangi Gathenya of Jomo Kenyatta University of Agriculture and Technology for their guidance through suggestions and critiques as well as support which were vital for the successful completion of this thesis.

Many thanks are extended to Dr. Joy Obando of the Department of Geography, Kenyatta University for her words of advice and encouragement especially during the proposal writing. Many thanks are also extended to Mr. Joseph Okola of the Departmental computer laboratory for helping uplift my computer knowledge.

I appreciate the assistance offered by the staff of the data section, Ministry of Water Resources and Irrigation especially Mr. D.M Muikia and the Kenya Meteorological Department especially Mr. P.M. Macharia, who assisted me in getting the necessary data

Many thanks are also extended to Mr. Mboji of Kenya Meteorological Department and Mr.Maina of Kenya Agricultural Research Institute for allowing me access to their respective libraries and for assisting me with the required literature material.
Many thanks also to Mr. S.K. Ng'ang'a of the Regional Centre for Mapping of Resources for Development (RCMRD) and Mr. F.M. Muthusi formerly of the United States Geological Survey (USGS) for their assistance with GIS data.

Many thanks also go to my employer the Jomo Kenyatta University of Agriculture and Technology for granting me the study leave that enabled me to me undertake the course.

Special thanks are also extended to the Chairman and the entire staff of the Department of Civil Engineering, Jomo Kenyatta University of Agriculture and Technology for the encouragement and support provided in various ways during the course of the study.

Thanks to my colleagues Messrs Muli, Ajigoh, Sungu and Maunda for sharing the heat with me. I also wish to thank (posthumously) the late J.M. Wanyonyi a long time friend, colleague, and classmate whose life was cut short by gangsters in the middle of the course. May his soul rest in peace.

I cannot also forget my many friends who supported me through prayers and words of encouragement during my long struggle to pursue further education.

All your prayers were not in vain.

To all I say "Ahsante Sana" and God bless you all abundantly.

ABSTRACT

A regionalisation methodology was applied to catchments of the Upper Tana basin to enable estimation of daily stream flows for any catchment within the region for which physical attributes data and records of rainfall and temperature are available. IHACRES, a lumped conceptual rainfall-runoff model was calibrated at a daily time step to six catchments ranging in size from 49km^2 to 600 km^2 within the upper Tana basin to obtain a set of model parameters that characterise the hydrological behaviour of the catchments. Physical catchment characteristics representing topography, soil and land use were derived from spatial data using GIS. By correlating these two sets of parameters, equations were developed which related the conceptual parameters to catchment characteristics and which enabled the estimation of model parameters from catchment characteristics. The estimated parameters were used to simulate stream flows in two validation catchments within the same region and the goodness of fit evaluated using statistical and graphical methods. Sensitivity tests were carried out at one of the catchments in which flow response to variations in model parameters was assessed in order to analyse calibration errors. The calibration R^2 ranged from 0.57 to 0.85 while the simulation R^2 ranged from 0.55 to 0.77 for all catchments. The Nash-Sutcliffe (1970) efficiency ranged from 0.78 to 0.91 for calibration and from 0.77 to 0.88 for simulation. The stream flows simulated using the estimated parameters agreed well with the observed stream flow series. The R^2 values were 0.21 and 0.67 while the Nash-Sutcliffe efficiency values were 0.21and 0.68 respectively. Although varied, the results indicated that regionalisation has the potential to generate stream flow data in ungauged catchments for purposes of water resources assessment.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ANU	Australian National University
ARPE	Average relative parameter error
CRES	Centre for Resource and Environmental Studies
R^2	Coefficient of determination
DEM	Digital elevation model
DRC	Dynamic Response Characteristics
EROS	Earth Resources Observation System
FAO	Food and Agricultural Organisation
FDC	Flow duration curve
GIS	Geographical Information System
GoK	Government of Kenya
HMLE	Heteroscedastic Maximum Likelihood Estimator
IAHS	International Association of hydrological sciences
IH	Institute of Hydrology
IHACRES	Identification of Hydrographs and Components from Rainfall, Evaporation and Stream flow
JICA	Japan International Co-operation Agency
KMD	Kenya Meteorological Department
PCD	Physical Catchment Descriptors
PUB	Prediction of Ungauged Basins
RCMRD	Regional Centre for Mapping of Resources for Development
SK	Survey of Kenya
SLS	Simple Least Squares

SRTM	Shuttle Radar Topographic Mission
UH	Unit Hydrograph
δ	Time delay factor in the IHACRES model
E_o	Potential Evaporation
USGS	United States Geological Survey
$^{a}R^{2}$	Multiple regression coefficient
f	Temperature modulation factor (°C^{-1})
τ_w	Catchment drying time constant (days)
C	Volume of conceptual catchment wetness storage (mm)
ν_q	Proportional volumetric contribution of quick flow to total stream flow
ν_s	Proportional volumetric contribution of slow flow to total stream flow
S_k	Catchment wetness index
r_k	Rainfall at time step k
t_k	Air temperature at time step k
u_k	Effective rainfall at time step k
x_k	Stream flow at time step k
τ_q	Quick flow response decay time constant (days)
τ_s	Slow flow response decay time constant (days)
S_Grad.	Stream gradient
D_Dens	Drainage density

CHAPTER ONE

INTRODUCTION

1.1 General

Sustainable water resources planning and management requires hydrological data to enable quantification of both water quality and quantity (Oyebande, 2001). River flow measurements are important for water resources planning, conservation, pollution control and for solving many engineering and environmental problems (Mwakalila, 2003, Kokkonen et *al.,* 2003). These include designing bridges and dam structures, flood control, water quality control and stream habitat assessment. In addition they are required on a routine basis to support abstraction licensing and to permit effluent discharge into rivers.

The data can be obtained through measurements at stream gauging stations. However, due to the high costs associated with procurement and maintenance of stream gauging stations, many catchments especially in developing countries are either poorly gauged or not gauged at all and hence such data may not be available when and where it is needed. In addition, a few of the gauging stations operate inconsistently due to maintenance problems leading to gaps in data records. Furthermore, changing land use conditions and water abstractions make past stream flow records of limited use for the assessment of future water resources (Thomas *et al*., 1993).

Inadequacy of data occurs not only in catchments without stream gauging stations but also in gauged catchments, due to inconsistency in the data, inaccuracy of measurements, short duration recorded data or low gauging network density. In addition, gauging stations are sometimes located in remote areas, which become inaccessible during rainy seasons hindering acquisition of stream flow data as well as proper maintenance of equipment. Despite this restriction on availability of stream flow data in both temporal and spatial respects, data needs are inevitable and continue to grow with increasing water demand.

Design of water resources systems is of increasing importance due to the increasing need for efficient water resources management, which requires among other things, efficient control of water resources in a river basin. The need to control water movement arises not only because of the danger it poses particularly under extreme conditions like floods but also to make it available for various activities. To ensure adequate water supply for domestic, agricultural and industrial use, hydraulic structures for storage and distribution of water need to be designed and constructed. This requires reliable and up to date stream flow data. Moreover, availability of stream flow data in real time is required in the management of water resources, reservoir operations and in flood forecasting and control.

In most developing countries such as Kenya, long term data required for the design and planning of water resources projects is not readily available meaning that other ways of generating this data are required. There is need therefore, for the development of a methodology that could help in improving availability of data and which can be applied in both ungauged and poorly gauged catchments. Given that in most tropical-equatorial countries, the duration of rainfall data is usually longer than the duration of stream flow measurements (Solomon, 1967); it becomes impossible to carry out time concurrent analysis of rainfall and stream flow. The need for augmenting discharge records in these circumstances is obvious and leads to the problem of rainfall-stream flow conversion using an appropriate model and a set of model parameters.

A number of modelling approaches have been developed for this purpose and whilst the application of any of these approaches for modelling gauged catchments is straight forward, simulation of flows in ungauged catchments demands independent measurement or estimation of model parameters. Normally one option is to develop models for gauged catchments and link their parameters to physical catchment characteristics, so that the approach can be applied to ungauged catchments in the same geographic region whose physical catchment characteristics can be determined or estimated.

Regionalisation of model parameters is the calibration of models to representative catchments scattered across a region and the assignment of the resultant model parameters to other catchments around them with assumed similar hydrological characteristics (Onyando, 2000). It aims at extending in space the available hydrological data and is undertaken to overcome limitations of lack of hydrological data. An assumption of this approach is that the variability in space of flow characteristics can be explained by basin characteristics (Vogel and Kroll, 1992; Reimers, 1990).

The availability of reliable hydrological data is recognized to be a world-wide problem due to the costs and logistics involved in running extensive gauging networks and because the existing data sets often include missing records. The International Association of Hydrological Sciences (IAHS) has recognized this need and in 2002, adopted the Predictions of Ungauged Basins (PUB) as a decadal research agenda for the period 2003-2012, the principal objectives being to further develop methodologies for predictions in ungauged basins and to reduce uncertainties associated with model predictions (Sivapalan *et al.,* 2003).

Rainfall-runoff models are mathematical expressions of rainfall-runoff processes which can be used to interpolate and extrapolate past or future events. They vary in complexity with some models having conceptual parameters only while others have both conceptual and physical parameters. Majority of these models have been developed for humid temperate regions and need to be adapted to local conditions through calibration and validation before they can be applied locally. The calibration and validation is done using rainfall, stream flow and physiographic data.

The data needed for the determination of physical parameters are normally available from topographical maps, soil maps or remote sensing and are derived using Geographical Information Systems (GIS). Conceptual parameters on the other hand, require rainfall, stream flow and temperature data for their calibration which is achieved through optimization algorithms (Sorooshian and Gupta, 1995). This involves a systematic search for parameter values that yield computed runoff hydrographs that best match observed hydrographs. With model parameters determined, the models can then be used to improve data records in gauged catchments, if rainfall and temperature data for missing stream flow records are available.

In ungauged catchments rainfall–runoff models can be used to generate data after regionalisation. This involves correlating their conceptual parameters to catchment characteristics from several gauged catchments (Shaw, 1996; Nathan and McMahon, 1990; Funke *et al.*, 1999). Describing these hydrological characteristics in terms of physical catchment descriptors (PCDs) allows the estimation of the unit hydrograph (UH) for any catchment within the region. The resulting relationships are then used to derive model parameters for the ungauged catchments in the same geographic and climatic region making it possible to simulate stream flows for the ungauged catchments. Application of this methodology allows stream flow series to be constructed and sensitivity of flow to hydrological and physical catchment characteristics to be determined. Differences in physical factors such as slope, soils, vegetation, drainage pattern and density between catchments have marked differences in their stream flow regimes.

The linking of model parameters to physical catchment characteristics allows the estimation of stream flows, both historical and climate influenced at ungauged sites. When these characteristics include land use indices, it becomes possible to quantify second order effects resulting from changes in land use. The second order effects on stream flows will normally be represented if any of the physical catchment descriptors in the dynamic response characteristics and physical catchment descriptors (DRC-PCD) relationships represents land use such as percentage of forest or arable land and some estimate is available of how land use will change with climate or time.

1.2 Background of the Study

The scarcity and vulnerability of water resources in most parts of Kenya make accurate assessment and planning of the available water resources vital (Nyadawa, 1997)

Relative to area (587,900 km^2) and population 28,686,607 (GoK, 2000), Kenya has limited water resources with most of its perennial rivers concentrated in the central and western parts of the country (JICA, 1992). It is classified as a chronically water scarce country with an annual renewable fresh water supply of only 647 m^3 per capita, well below the global 1000m^3 per capita set as the bench mark for water scarcity. Kenya's water scarcity compares poorly with its neighbours, Uganda and Tanzania with 2,940 m^3 and 2,696 m^3 per capita respectively, (World Bank, 2003).

The rapid population growth and the resultant increase in water demand for domestic, agricultural and industrial use make it necessary to accurately estimate stream flows for reservoir and water supply planning as well as for direct stream abstractions especially during droughts in order to avoid conflicts among various water dependent activities. Given these conditions, improved planning, development and management of the scarce water resources is essential to maximise the available water resources and balance competing uses.

The planning, design, development, management and operation of hydrological systems depend on reliable, accurate and timely stream flow information which can only be obtained from quality hydrological data. The acquisition of this data requires skilled personnel to collect process and interpret it. It is anticipated that in the next decade, the demand for reliable surface water data will be constantly increasing due to expanding development programmes for all water resources projects, so careful monitoring of existing stations and judicious use of gathered data are inevitable (JICA, 1992).

While Kenya's needs for stream flow data are increasing due to rising water demand resulting from population increase, agricultural and industrial development, the financial, human and technical capacities for data collection and monitoring are declining, (table1.1).

Table 1.1 Registered stream gauging stations in Kenya

Name of Drainage Basin	Area of Basin	Registered Stations	Registered Stations (1990)	Registered Stations (2001)	% Drop
Lake Victoria	46,229	229	114	45	80
Rift Valley	130,452	153	50	33	78
Tana River Basin	**126,026**	**205**	**116**	**66**	**67**
Athi River Basin	66,837	223	74	31	86
Uaso Ng'iro Basin	210,226	113	45	29	74
Total	579,770	923	399	204	78

Source: (World Bank, 2003)

Hydrologists and water resource planners therefore need to seek alternative methods of generating this data and because rainfall data is generally in abundance compared to stream flow data, there is a tendency among hydrologists to convert rainfall into stream flow using appropriate rainfall-runoff models.

Since it is cost prohibitive to gauge all existing streams and tributaries in a basin, regionalisation provides a means of improving quantile estimates at gauged river sites and transferring hydrological information from gauged catchments to ungauged catchments. Through regionalisation, expensive redundancies in data collection networks can be avoided besides ensuring sustainable data collection from an optimal number of gauging stations in a given basin (Onyando, 2000)

1. 3 Statement of the Problem

River Tana, which drains the study area is Kenya's largest river, with a catchment area of 126,026 km^2 and comprises of regions of widely differing physical characteristics and development potential. It represents a major contributory sector within the Kenyan economy and contains some of the important resources in the country. These include major hydro-electric power plants (the "seven forks" development scheme), irrigation schemes (Mwea, Bura, and Hola) and large diversion schemes for supplying Nairobi city and its environs with water through inter-basin water transfer schemes. Water within the Tana basin is unevenly distributed seasonally and across locations such that water demands have to be met from areas of water abundance and has also to be stored during the wet seasons for use during the dry seasons.

It is generally accepted that competing water uses are restricting future development in the Tana basin (Harper and Brown, 1998). The use of flood waters in the lower Tana for flood recession agriculture conflicts with regulating the river for hydro-electric power generation and for providing adequate water supplies for flood damage protection of the riverine settlements and infrastructure. This calls for careful economic decisions on water usage which can only be achieved with adequate and reliable stream flow data.

The planning of the final two dams in the cascade which have not been built should take into consideration the environmental impact of stream flow changes and provide flexibility for environmental flows. Resolution of water conflicts within the basin requires comprehensive planning and management which can only be accomplished with adequate hydrological data that cannot be provided by the existing gauging stations in the basin. Although the upper Tana basin is reasonably well covered with both river flow and rainfall gauging stations, in many cases, the length of record and the reliability of the data are insufficient for them to be of much use (Bobotti, 1996). There is need therefore for alternative methods of generating this data to enable proper assessment of the water resources potential of the basin.

1. 4 Research Questions

In response to the research problem there was a need for a study to address the following questions:

1. Can a lumped conceptual rainfall-runoff model be used to extend existing stream flow records in areas with limited historical data?

2. Can GIS be used to derive physical parameters from catchment characteristics for regionalising a conceptual rainfall-runoff model?
3. Can relationships derived between optimized parameters of a lumped conceptual rainfall-runoff model and physical catchment descriptors be used to regionalise rainfall-runoff models for stream flow estimation?
4. Given the declining stream flow records, can regionalised parameters of a lumped conceptual rainfall runoff model be used to estimate stream flow in ungauged catchments of the Upper Tana basin?

A study was carried out to come up with solutions to these questions. The study involved use of a lumped conceptual model IHACRES to estimate flows in the ungauged catchments of the Upper Tana Basin. IHACRES is a catchment- scale rainfall-stream flow model developed jointly by the Institute of Hydrology (IH) and the Centre for Resource and Environmental Studies (CRES), at the Australian National University (ANU), Canberra to assist in the characterization of the dynamic relationship between basin rainfall and stream flow. The model was adopted because of its ability to represent basic information in climate (rainfall, temperature) and stream flow data with a minimum number of parameters. The only field-data required by the model are time series of rainfall and stream flow and a third variable for estimating the effects of evapotranspiration. The variable used in this study was air temperature.

1. 5 Research Objectives

The aim of this research was to regionalise parameters of a lumped conceptual rainfall-runoff model for stream flow estimation in ungauged catchments of the Upper Tana basin.

The specific objectives of the study were to:

1. Calibrate and validate the lumped conceptual rainfall-runoff model IHACRES using selected catchments of the Upper Tana basin.
2. Derive physical parameters from catchment characteristics of the selected catchments for model calibration and regionalisation.
3. Establish regionalised parameters for the Upper Tana basin for stream flow simulations using the rainfall-runoff model.
4. Estimate stream flows in the ungauged catchments of the Upper Tana basin using regionalised parameters of the conceptual model.

1.6 Justification and Significance

The Tana basin was selected for this study because of its economic importance to the country as river Tana and its tributaries comprise the country's major surface water resource. It is a major source of water for various uses within and outside the basin. Since the basin has a high potential for economic development in terms of irrigated agriculture and hydropower generation activities, these activities can be planned, implemented and managed well if the available water resources potential of the basin is properly assessed.

To meet the ever increasing water demand and to alleviate persistent water shortages especially during droughts, there is need for appropriate water storage structures such as dams for storing excess water for release during periods of shortages. These structures are required even in areas without river gauging stations yet their design depends on the availability of adequate stream flow data. This study attempted to estimate stream flows in such ungauged catchments for purposes of water resources assessment.

Previous studies in the Tana basin have focused on general aspects of the basin resource management (Bobotti, 1996), flooding (Mutua, 1993) sedimentation and erosion (Otieno and Maingi, 1993), social-economic aspects (Mutisya and Mutiso, 1998) and hydrologic drought (Agwata, 2005). A review of past studies on rainfall-runoff modelling in Kenya has been done by Nyadawa (1997). Further in Onyando (1994), Onyando and Sharma (1995), TAMS (1980), Thomas *et al.,* (1993) and Onyando (2002) studies have been carried out on direct runoff simulations in gauged catchment where event-based rainfall and runoff data were available for model calibration. Simulations in ungauged catchments have also been carried out in the past using the rational formula and intensity-duration frequency curves but these have simulated only peak runoff rates for design purposes. From the literature reviewed, no studies have been carried out on continuous daily stream flow simulations in ungauged catchments using regionalised parameters of a lumped conceptual model. The results of this study will provide a valuable insight into the overall hydrologic behaviour of the Upper Tana basin and in addition provide a useful alternative method of water resources assessment in catchments with inadequate gauging stations or no gauging stations at all within the basin.

1.7 Scope and Limitations

Since most rivers within the basin are fairly regulated in terms of artificial water abstractions, the abstractions were assumed to be insignificant on the natural stream flow records since the abstractions are not continuous, are regulated through issuance of permits and take place at specific times and in specified quantities. No naturalization of historic flow records was carried out. The study was limited to those rivers whose gauging stations had at least 20 years of recorded stream flow data within the upper catchment area of the Tana basin that covers an area of approximately 7,000 km^2. The middle and lower catchments of the Tana basin did not form part of the study area due to lack of sufficient stream flow data for model calibration as most of the rivers are not sufficiently gauged.

1. 8 Description of the Study Area

1.8.1 Location

The Upper Tana basin, covering an area of 6678 Km^2 includes the greater part of central Kenya, part of the middle regions of eastern Kenya and small sections of Rift Valley province. It forms part of the Tana basin and extends from the river drainage systems of 4A to 4C. It lies between longitudes 36.58°E and 37.54°E and between latitudes 0.16°S and 1.20°S. Altitude varies from 4000m above mean sea level near the slopes of Mt Kenya to about 1000m above mean sea level around Kiambere area. It encompasses a diverse area bounded to the north by Mt Kenya and the Nyambene ranges, to the west by the Aberdare ranges and to the south by River Tana (figure 1.1). The relief of the area is sharp with numerous perennial streams descending in deeply incised valleys separated by long narrow ranges. The lithology of the area comprises mainly of igneous rocks but with a significant proportion of metamorphic rocks. Small patches of sedimentary rocks are also found in the basin mainly in Kiambu and Nyandarua districts.

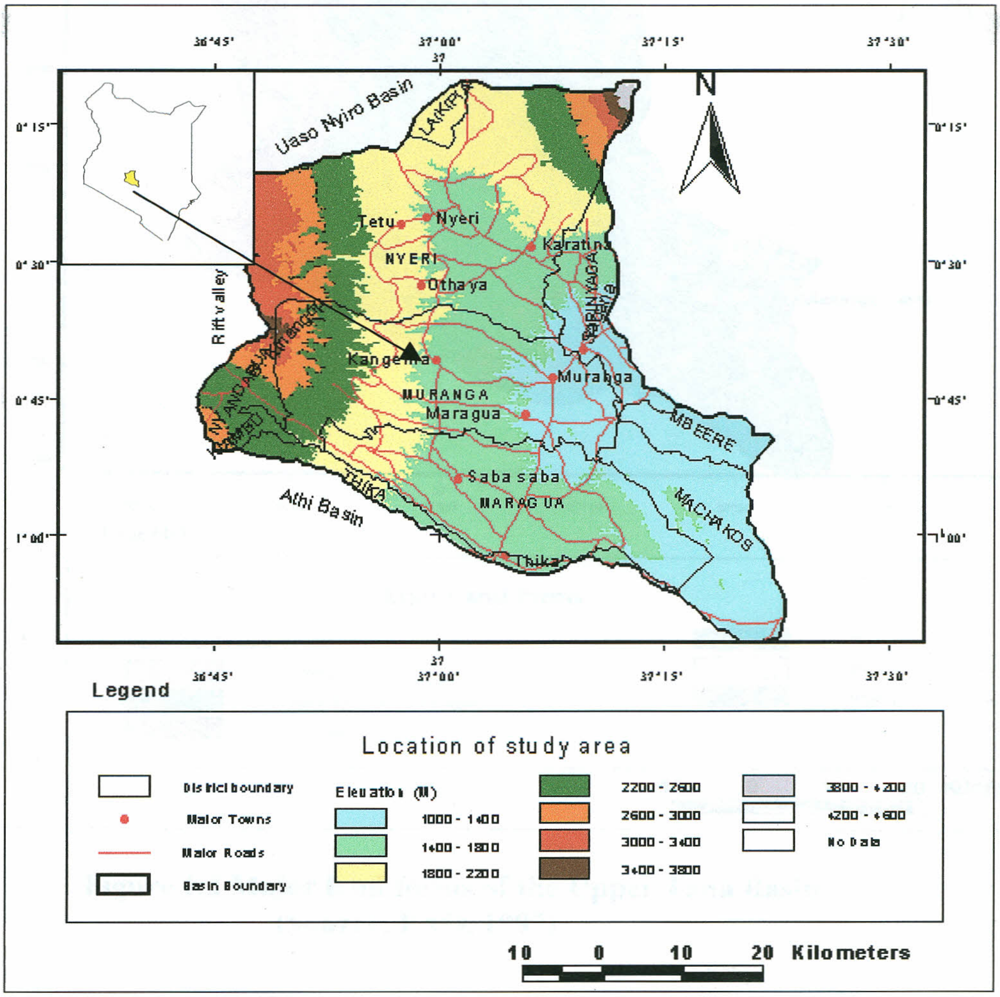


Figure 1.1 Location map of the Upper Tana basin

Landforms in the area comprise of mountains in the North-East (Mt Kenya) and the western part (Aberdare ranges), hills and mountain foot ridges in the central part, plains in the south-eastern part and small patches of foot slopes in the southern and south-western parts as shown in figure 1.2.

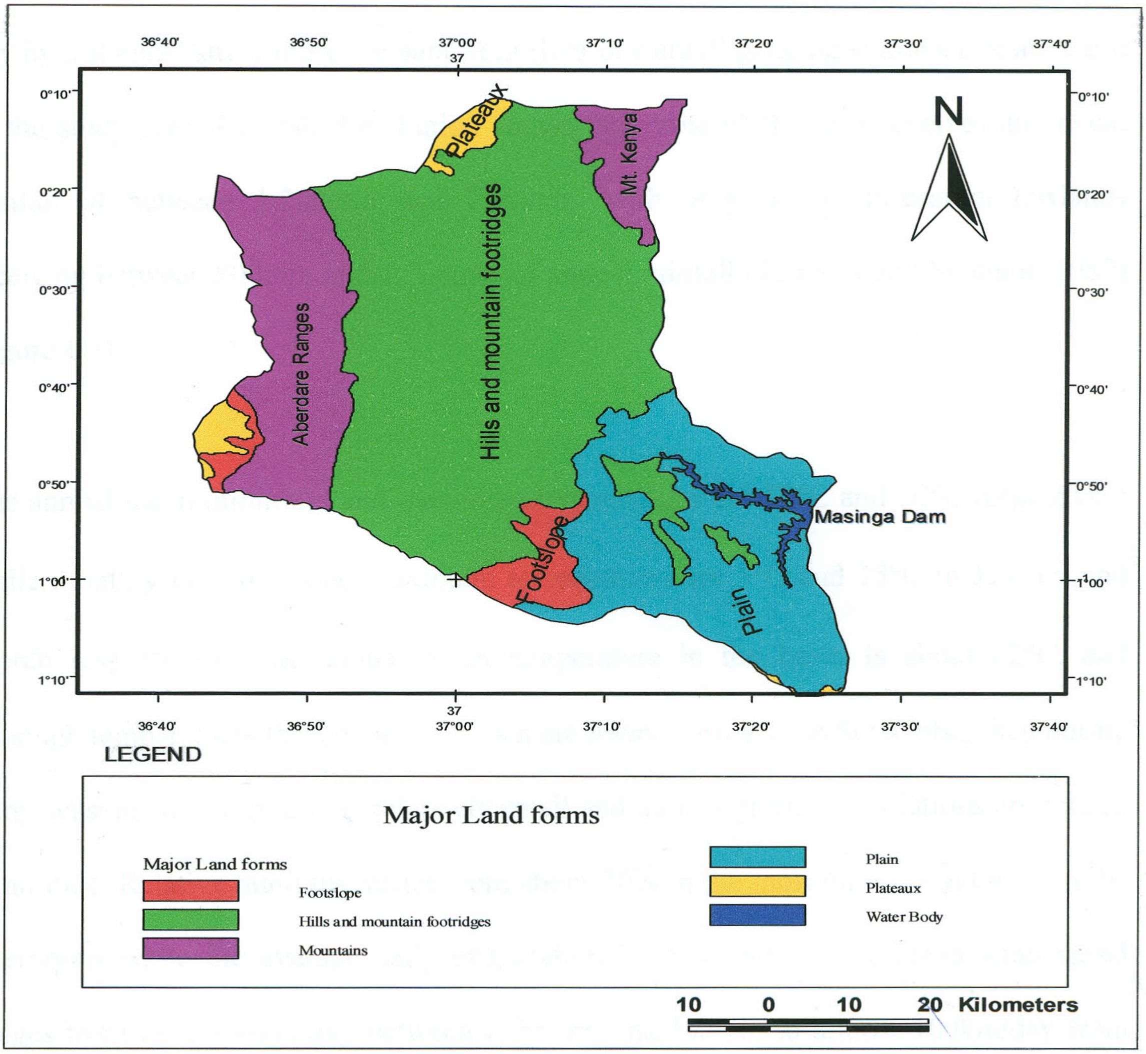


Figure 1.2 Major land forms of the Upper Tana Basin
(Source: FAO, 1997)

1.8.2 Climate

Like most parts of Kenya, the area experiences a bimodal rainfall pattern (two rainfall seasons) with the long rains starting from March to May and the short rains starting from October to December. Rainfall decreases with increase in temperature and decrease in altitude from the mountain summits to the south east lowlands, with areas lying within 1400-1800m above mean sea level receiving high amounts of rainfall and those below

1100m above mean sea level receiving low unevenly distributed rainfall. Mean annual rainfall varies greatly even within short distances and due to this pronounced variations any hydrological study must use data from river or rainfall gauging stations that are close to the study area. Rainfall distribution shows that most of the area receives an annual rainfall of between 1,200mm and 2,400mm with only the south-eastern lowlands receiving between 800mm and 1,200mm of annual rainfall (Jaetzold and Schmidt, 1983) (figure 1.3)

The annual mean minimum and maximum temperatures are 10°C and 20°C respectively while monthly minimum and maximum temperatures are 5°C and 15°C in January and March respectively. The annual mean temperature in the basin is about 22°C, and although temperatures throughout the basin are lower during June-September than during rainy seasons, this variation is relatively small and more significant variations are related to altitude. Relative humidity varies from about 70% in the morning to about 45% in the afternoons while the average daily evaporation is about 6mm. The mean wind speed ranges from over 140km/day between February and March, to almost 100km/day from May to July while daily sunshine hours range from about 4 hours in July to about 9 hours in February (Jaetzold and Schmidt, 1983). In the Upper Tana basin potential evaporation is about 1500mm. The area has fertile clay-loam soil of volcanic origin that is well suited for coffee and tea growing. It is densely populated and experiences intensive irrigable agricultural practices that considerably reduce the dry season's low flow in rivers.

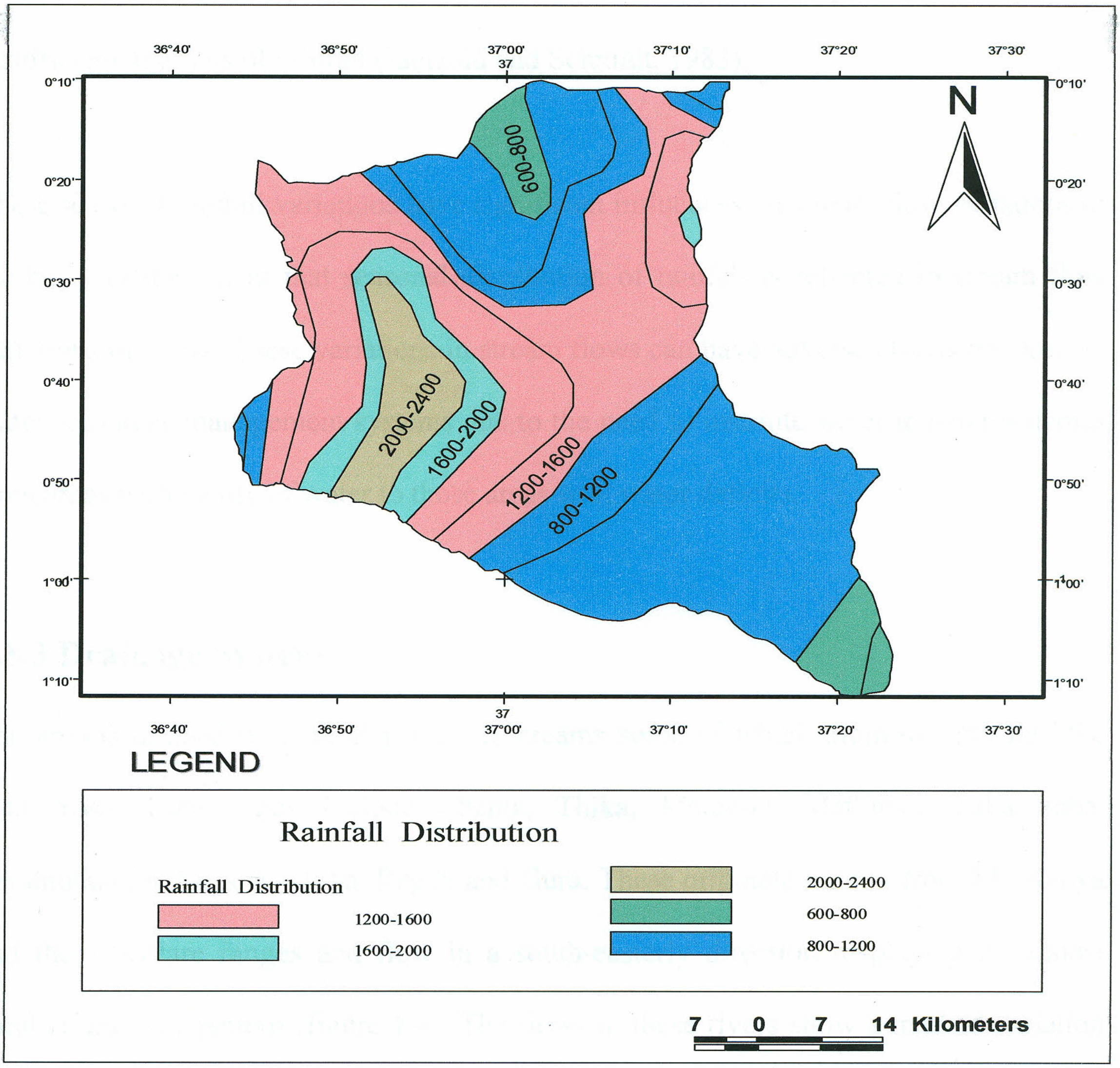


Figure 1.3 Rainfall distribution within the Upper Tana basin
(Source: http://www.ilri.cgiar.org/gis)

Climate influences stream flows in most parts of the basin with Mt Kenya to the north-east and the Aberdare ranges to the west in particular influencing rainfall patterns through orographic lifting of moist air masses in the northern and western parts of the basin. Topography and relief affect both seasonal and annual rainfall distributions within the basin to the extent that the upper parts to the North and West receive high amounts of rainfall while the lower parts to the extreme East and South-West receive unreliable and insufficient amounts of rainfall (Jaetzold and Schmidt, 1983).

These seasonal rainfall variations have significant influences on stream flow variations in the basin to the extent that seasonal distribution of rainfall is reflected in stream flow variations in rivers. These variations in stream flows can have adverse effects on existing water resources management systems due to the need to institute water transfer schemes from areas with plenty of water to those areas with water deficits.

1.8.3 Drainage System

The area is drained by several perennial streams some of which drain directly into the main river Tana. They include Chania, Thika, Maragua, Mathioya, Saba Saba, Rwamuthambi, Sagana, Thiba, Ragati and Gura. These originate mainly from Mt. Kenya and the Aberdare ranges and flow in a south-easterly direction displaying an almost parallel drainage pattern (figure 1.4). The flows in these rivers show a marked variation in seasonal and annual volumes that pose a major challenge to proper management of water resources in the basin. The high flows generally occur from March to June and from October to December while low flows occur during the months of January and February and between July and September (Bobotti, 1996).

1.8.4 Surface Geology

The surface geology of the area comprises mainly of the Precambrian, metamorphic and volcanic rock formations with nearly half of the area covered by tertiary and quartnary volcanic rocks such as basalt, phonolites and trachytes. However, some sedimentary rocks occur within the main river valleys while glacial deposits form the main rock type below the peaks of Mt. Kenya. The volcanic rock formations originate mainly from Mt Kenya and the Aberdare ranges. Areas below 1000m above mean sea level contain rocks consisting mainly of gneisses, schists and calcareous rocks that originate from sediments that have undergone geological changes as a result of shearing, folding and faulting.

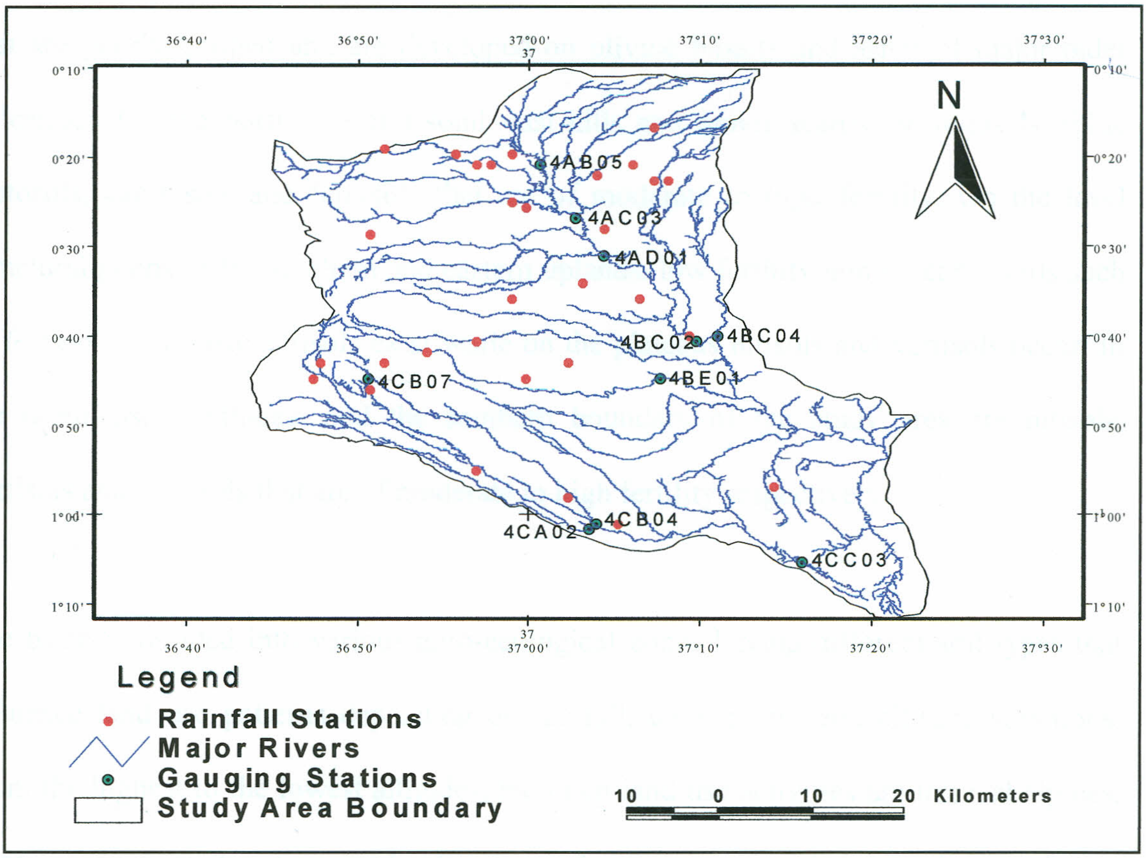


Figure 1.4 Hydrological network of the Upper Tana basin

The youngest rocks are quartnary and tertiary volcanic rocks that originate from Mt. Kenya and volcanoes in the Aberdare ranges. Examples of these include basalts (Hughes, 1984). Most of the river flood plains consist of recent alluvial sediments although some parts overlie volcanic rocks, the Mozambique belts and metamorphic rocks (Baker, 1967).

1.8.5 Soils and Agro-Ecological Zones

Most soils in the basin are volcanic in nature and have high infiltration rates besides being highly permeable and resistant to erosion. The higher parts of the basin such as mountains and escarpments comprise of histosols, humic andosols and lithic leptosols that are poorly drained and are developed on olivine basalts and ashes of major older volcanoes. On the north-east and south-east hills and minor scarps are regosols, lithic leptosols, cambisols and ferrasols that are of moderate to high fertility. On the level structural plains of the southern and eastern uplands, low fertility non-volcanic soils such as ferrasols and ferric acrisols exist while on the plateaus nitisols and vertisols occur. In the north-east, south-east and the southern boundary of the study area are nitisols, vertisols and fluvisols that are of moderate to high fertility respectively.

The basin is divided into various agro-ecological zones having different soil types that influence land use patterns depending on rainfall, temperature and altitude variations. From the highest to the lowest altitudes, the main land use activities are tropical alpines, forest, sheep-dairy, tea-dairy, and coffee-tea and coffee growing. Other land use activities include National parks, forests, pyrethrum-wheat, wheat-barley, wheat-maize-pyrethrum and wheat-maize-barley growing (Jaetzold and Schmidt, 1983) figure 1.5

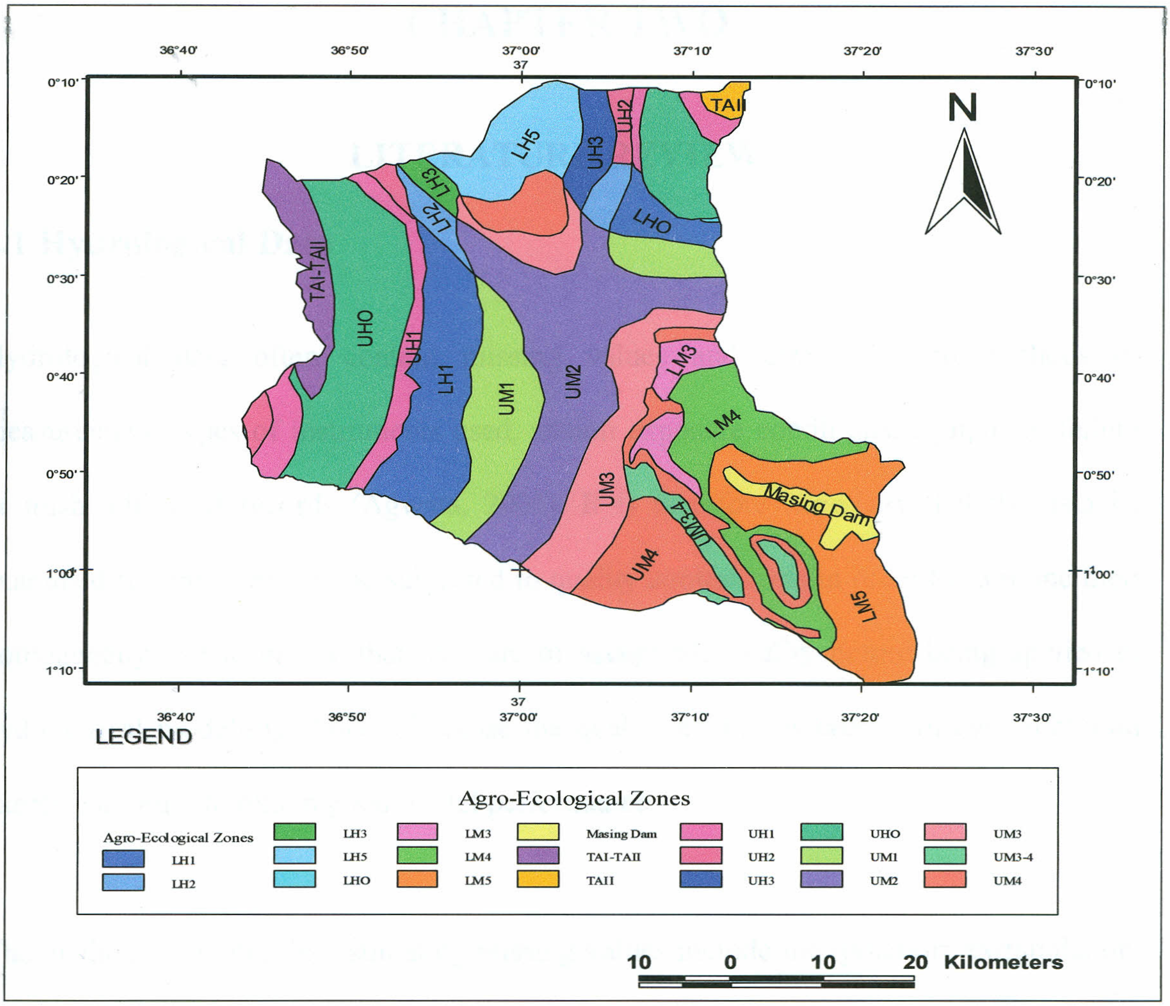


Figure 1.5 Agro-Ecological Zones of the Upper Tana basin
(Source: Jaetzold and Schmidt, 1983)

Various land use activities occur in the various Agro-Ecological zones and are related to the different soil types in the basin. Most of these activities are dependent on the availability of water in the basin for their development and sustenance and in view of the decreasing quantities of water in most rivers of the basin there is need to plan these activities according to the available water resources of the basin and hence the need for accurate and reliable assessment of the water resources potential of the basin.

CHAPTER TWO

LITERATURE REVIEW

2.1 Hydrological Data

Hydrological data often contain missing values and errors due to methods of measurement, types of instruments used, station exposure conditions, equipment failure or mishandling of records (Agwata, 2005). It is therefore necessary that the data be examined for any gaps and be subjected to quality control tests in order to examine their homogeneity and to ensure that they are of acceptable quality before being applied in hydrological modelling. This is because the quality of data is crucial in the calibration exercise in order to obtain good model performance.

The methods available for estimating missing values include interpolation, extrapolation, grid square mapping, regression and correlation analysis. These methods have various levels of accuracy depending on the density of observations or station network (Shaw, 1988). Correlation analysis based on the correlation matrix of gauging stations is preferred when the correlation structure is statistically significant, (Chow *et al.,* 1988). It is particularly useful when neighbouring gauging stations in a basin or those from the same stream are highly correlated meaning that the neighbouring gauges have similar basin characteristics such that the flows originate from similar rainstorms within given catchments.

Hydrological data may contain errors that render the records heterogeneous at various gauging stations. This may be due to changes in the location of the gauging stations, exposure of the stations due to developmental works and land use demands, urbanization or techniques used for taking observations. The quality and consistency of data need therefore to be checked before it is analysed. Methods available for quality control include runs tests, regression methods, and mass curve analysis techniques (WMO, 1994)

2.2 Conceptualisation of Rainfall-Runoff Relationships

One of the first 'event' models relating storm runoff to rainfall was the rational formula (Mulvaney, 1851). Later, the development of the unit hydrograph concept (Sherman, 1932) and its subsequent evolution to the instantaneous unit hydrograph provided the basis for the storm response models of Nash (1957) and Dooge (1959). From these beginnings current methods of estimating catchment response to individual storm events have developed and grown in complexity. However, models of this type do not estimate continuous flows but rather flood peaks for engineering design purposes only, modelling of the former being of more recent origin (Blackie and Eeles, 1985).

One of the first lumped conceptual hydrologic models to emerge for simulating continuous flows was the Stanford watershed model (Linsley and Crawford, 1960) which estimated daily flows from inputs of daily rainfall using infiltration, unit hydrograph and recession functions. This was further developed by the inclusion of soil moisture budgeting, evapotranspiration estimates and flow routing techniques to give hourly estimates (Linsley and Crawford, 1966). However, increased computing power has advanced the development of conceptual hydrologic models (Kokkonen, 2002) which have found wide applications in practical problems such as operational flood forecasting.

2.3 Estimation of Stream Flows

The availability of measured stream flow data is restricted in both temporal and spatial respects and for this reason extension of existing data is usually required in solving many practical problems. This is achieved through the use of conceptual rainfall- runoff models. Provided good stream flow record covering a few years is available for model calibration and validation, the models may be used to extend the stream flow record to the full duration of the rainfall and other meteorological records of the catchment.

When no historical stream flow data is available to support model calibration and validation, information from other catchments become very valuable (Kokkonen, 2002). To estimate stream flows from ungauged catchments for existing or future conditions, model parameters may be extrapolated from gauged catchments within the same geographic and climatic region. This transfer of information from one or more catchments to another is known as regionalisation (Bloschl and Sivapalan, 1995). The ability to predict stream flow in an ungauged catchment (one without adequate quantity and quality of data) remains an important goal of hydrology and its success indicates an understanding of the principal drivers of catchment hydrologic response

2.4 Regionalisation and Uncertainty

Regionalisation is an attempt to relate flow characteristics at gauging stations to physical and climatic characteristics of their drainage basins (Riggs, 1990). It is the transfer of information in the form of characteristics describing the hydrological data or models from one catchment to another (Bloschl and Sivapalan, 1995). These characteristics should be those that are easily obtainable from existing data sources and which can also be obtained for the ungauged catchments. Assuming the identified relationships hold good for the ungauged catchments, stream flow simulations can be carried out for these catchments by estimating the model parameters from physical catchment descriptors and then using the models to simulate the required stream flows.

Regionalisation of conceptual rainfall runoff models is a popular approach to estimating stream flows in ungauged catchments (Post and Jakeman, 1996, Sefton and Howarth, 1998 and Kokkonen *et al.,* 2003). The technique involves calibrating models on gauged catchments and determining relationships between these model parameter values and physical catchment attributes such as topography, geology, soils and vegetation cover. These relationships are then be used to estimate model parameters for the ungauged catchments. With the parameters determined, the models can be used to simulate flows.

However, to effectively transfer information on hydrological behaviour from gauged to ungauged catchments, these catchments should form a relatively homogeneous group (Nathan and McMahon, 1990; Pilgrim, 1983; Post and Jakeman, 1996). According to Bates (1994) successful regionalisation of rainfall-runoff models depends on accurate estimation of model parameters for the gauged catchments, selection of PCDs with significant influences on catchment response to rainfall, proper identification of homogenous regions and the degree of correlation between model parameters and physical catchment descriptors.

Regionalisation aims at deriving transfer functions between model parameters and catchment characteristics and using these functions on ungauged catchments (Onyando, 2000). It stems from the fact that model parameters are region specific and should be derived for every region. However, application of universal models without regionalisation has been a common practice among hydrologists, one such model being the SCS curve number model for the estimation of rainfall excess (Chow *et al.,* 1988). The limitation of this application however, is that the simulated flows are likely to be overestimated or underestimated since the model parameters used do not reflect local conditions.

Depending on data availability, regression based procedure is the most widely used approach for relating catchment characteristics to model parameters and has been presented in the past by many authors. One of this dates back to 1960 when Nash linked multiple topographic factors to Nash model parameters (In Shaw, 1996). A regression technique through clustering of catchments into homogeneous sub-groups was used by Nathan and McMahon (1990, 1992) and Burn and Boorman (1992). Such clustering however, requires many catchments, which limits the applicability of the method. Sefton and Howarth (1998), Funke *et al.*, (1999), Mwakalila, (2003) have in separate studies presented step-wise regionalisation methodology applicable in situations where the number of gauged catchments is limited. Most often in regionalisation studies predictive focus has been on estimation of event characteristics. In particular estimation of flood indices such as mean annual flood (Mimikou and Gordios, 1989) or flood frequency (Burn and Zrinji, 1994) for ungauged catchments has received a lot of attention. Other event and index based studies have looked at low flows (Nathan and McMahon, 1990, 1992) and mean annual runoff (Reimers, 1990).

Another group of studies have estimated unit hydrograph characteristics (Burn and Boorman, 1992) with a view to reconstructing flow records. In many such studies, the catchments are grouped in an attempt to improve the fit of the observed data to the regression model. Of more recent focus has been the estimation of water balance model parameters with the aim of simulating continuous flows. Vandewiele and Elias (1995) simulated monthly flows for basins considered ungauged while Post and Jakeman (1996, 1999), Sefton and Howarth (1998) and Mwakalila (2003) predicted daily stream flows by developing relationships between the parameters of a daily time step rainfall-runoff model and physical catchment descriptors. In such studies, the models used should be parsimonious in order to capture efficiently the hydrological behaviour of the catchment. Servat and Dazetter (1993) found that it was easier to relate model parameters to catchment attributes for parsimonious models than for over-parameterized models. The consequences of over-parameterization (with respect to the information required to calibrate models) common in distributed models are well documented in literature (Pilgrim, 1983; Jakeman and Hornberger, 1993). Over-parameterisation arises when parameters of a model are so many such that the 'best fit' of model outputs to measured outputs can be generated by a number of different sets of parameter values raising doubts as to the reality of the parameters.

The main drawback to regionalisation type approach to ungauged catchments predictions is the need for data from many catchments belonging to the same region (Kokkonen, 2002). Schaake *et al.,* (1997) demonstrated that a large number of catchments were necessary in order to obtain meaningful relationships between model parameters and catchment characteristics. Another drawback is the limited number of gauged catchments usually available. Using large regions increases the number of available gauged catchments but at the same time increases the variation of climate and physiography between these catchments leading to the introduction of more variables in the regression analysis (Seibert, 1999). Furthermore, some scatter may be introduced as the relationships may change between regions (Mimikou, 1984).

Runoff depends on the aggregation of many climatic, geological and physiographic conditions in a catchment and these may vary in different respects between catchments making it difficult to find relationships between model parameters and a few catchment characteristics. Model parameters are sometimes also poorly defined meaning that almost equally good simulations can be obtained at different locations in parameter space (Jakeman and Hornberger, 1993). This may cause a scatter in the relationships preventing the parameters from being detected easily. A pre-requisite for any regionalisation approach to be successful is that differences in hydrological response should be identified and quantified with reasonable confidence (Kokkonen *et al.,* 2003)

The regionalisation process is usually accompanied by loss of accuracy due to optimization error, adjustment of boundary conditions and use of transfer functions. To determine the overall regionalisation efficiency, simulations should be carried out using the regionalised parameters and the efficiency determined which should be compared with that of optimization. Regional differences in climate and catchment characteristics normally cause variations in hydrologic processes (Pilgrim, 1983), therefore the validity of the transfer functions developed in the regionalisation process can only be ascertained if they are tested in separate catchments but in the same climatic region as those used in their development (Onyando, 2000).

While regionalisation of catchment model parameter values is a straightforward approach to the problem of stream flow prediction in ungauged catchments, the uncertainty in model predictions depends on the accuracy and precision of the catchment attributes, relationship between catchment attributes and model parameters, and on the choice and appropriateness of the model structure. The uncertainty in the relationship between catchment attributes and model parameters depends on the uncertainty in the catchment attributes and the calibrated model parameter values for the gauged catchments used to derive the relationships.

Another source of uncertainty in predicted response characteristics for ungauged catchments is the criteria used in fitting the model parameters to gauged catchments, which may be insensitive to the aspect of the rainfall-runoff behaviour determining the response characteristics of the catchments. Another contributing factor may be how well the model structure represents the key hydrological processes since a model structure good for a restricted purpose may oversimplify or omit processes which have an important influence on another response characteristic. Finally, gauged catchments may exhibit similar behaviour in a given response characteristic while differing markedly in others which affect model parameter estimates. This tends to limit the ability to identify relationships between model parameter values and catchment attributes. Since the purpose of regionalisation is to estimate flow characteristics at ungauged catchments rather than to estimate model parameters, the performance of regionalisation is usually assessed by comparing the estimated and observed response characteristics for gauged test catchments.

2.5 Hydrological Modelling

Hydrological processes are usually interconnected and operate in a continuous cycle, the hydrologic cycle. Each process receives an input and produces an output which can either be measured or estimated. Those, which cannot be measured, are estimated using hydrologic models. These models are essentially a set of mathematical equations based on theoretical principles and they link the inputs to the outputs (Onyando, 2000). They are used to study complex problems and to synthesize different kinds of information necessary for planning and decision making in matters related to water resources planning, development and conservation (Woolhiser and Brakensiek, 1982)

In stream flow prediction hydrological models provide an alternative method of extending stream flow data in both time and space for planning purposes and a disaster management tool capable of warning people in advance of incoming floods. Models are also useful in the design of flood events, where calibrated hydrological models may be used to derive rainfall events that correspond to design floods of specific return periods required for the design of hydraulic structures. The purpose of hydrological modelling may be summarized as to:

1. Determine the volume of water available in a catchment and to make efficient and cost-effective quantitative estimates of water- related variables at ungauged locations.
2. Determine the distribution in time and space of both water quantity and quality components and to optimize the project design of water resources.
3. Determine peak discharges along with computation of water levels to assist in flood forecasting and design of hydraulic structures.
4. Study the responses of the various hydrological systems through anthropogenic-induced changes e.g. upstream deforestation, urbanization, wetland drainage, reservoir construction, ground water contamination, global climatic change etc.

The various types of data required for hydrological modelling include rainfall, stream flow, meteorological (e.g. evapotranspiration), topographic and remote sensing data.

2.6 Types of Hydrological Models

Hydrological models vary greatly in nature and are applied across a wide range of scales. Vertessy *et al.*, (1993) broadly classified models into statistical, black box, lumped parameter and physically- based models. Although each of these modelling approaches has a role to play in hydrological predictions they suffer from certain inadequacies (Wheater *et al,* 1993) with some approaches being better suited than others to predict a variety of hydrologic impacts in a given situation. Each modelling approach has limitations which relate to their power (potential to predict specific effects of specific changes), utility (transferability to other sites and different conditions), accuracy (uncertainty implicit in model predictions) and ease of use (complexity and input data requirements) (Muthusi, 2004). The four types of models are described in the following sections:

2.6.1 Physically Based Models

Physically based models represent individual run-off producing mechanisms and cater for spatial variations in catchment characteristics such as soil properties and vegetation cover. However, according to Beven (1989), these models require significant amounts of input data, which is difficult to obtain, making them complex and too costly to use. Refsgaard *et al.,* (1992) notes that although physically based models rank highly on the 'power' and 'utility' scales, they rate poorly on the 'ease of use' and 'accuracy' unless the input data is of very high quality. However, physically based distributed models make use of parameters which have a physical interpretation and allow representation of spatial variability in the parameter values (Abbott *et al,* 1986). Because they relate model parameters directly to physically observable land surface characteristics, spatially distributed hydrological models have important applications in the interpretation and prediction of the effects of land use changes and climatic variability on hydrological response.

Physically based models make certain assumptions on the operations of hydrological systems. The physics on which the options are based is the small-scale physics of homogeneous systems. In real applications, it becomes a necessary to lump the small scale physics to the model grid scale, for example the 250m by 250m used in the SHE model applied to the Wye catchment (Bathurst, 1986). There is no theoretical frame work for carrying out this lumping of sub-grids processes for the spatially heterogeneous grid squares. It is assumed that the same small-scale physical equations can be applied at the model grid scale with the same parameters and in doing so a conceptual leap is made.

2.6.2 Statistical Models

Statistical models are usually based on regression relationships derived from 'paired catchment' and 'catchment treatment' experiments. They have been applied widely in catchment run-off predictions particularly in land use change contexts because they are relatively simple to construct (Hornbeck, 1973; Stoneman and Schofield, 1989). A major difficulty with these models, however, is that they need to be based on long term rainfall-runoff records. Furthermore, the statistical associations on which these models are based are not necessarily transferable to sites where catchment characteristics differ or planned treatments differ (Bosch and Hewlett, 1982).

2.6.3 Black- Box Models

Black box models do not explicitly consider the governing physical laws of the processes involved, but only relate the input to output through some transfer function (Leavesley, 1994). They employ time series analysis techniques such as moving average or filtering methods. Jakeman and Hornberger (1993) described various configurations of the black-box models in which they reported very good agreement between observed and predicted daily stream flows using the time series analysis transfer function method. However, the models appeared incapable of predicting the hydrologic impact of any catchment disturbance unless the model was first calibrated against stream flow data collected after such a disturbance.

2.6.4 Lumped Parameter Models

Lumped parameter models usually have a mechanistic basis but rarely discriminate between the many intervening processes which occur between rainfall hitting the ground and run-off arriving at the stream (Weeks and Hebbert, 1980). They represent the effective response of an entire catchment, without attempting to characterize the spatial variability of the response explicitly and assume that the system inputs and dynamics are uniform in space. A critical shortcoming of lumped parameter models is their inability to represent the spatial variability of hydrological processes and catchment parameters (Moore *et al.,* 1991). However, they simulate catchment response reasonably well although some accuracy may be lost as the scale of lumping increases (Kirkby, 1999). Among the lumped parameter models used for transforming rainfall excess into direct runoff is the unit hydrograph.

The complexities of the environment and data collection constraints have favoured widespread use of lumped conceptual models. This is because most models especially distributed ones, are over- parameterised with respect to the data required to calibrate them (Jakeman and Hornberger, 1993) and as a result model parameters estimated from them tend to have large uncertainties associated with them. Thus any relationships derived between these parameters and catchment attributes will inherently contain large uncertainties. Over- parameterization can be avoided by including only those processes that can be identified from observed data and by not making ***a* priori** decisions about the model structure (Post and Jakeman, 1996). Currently, precipitation, total stream flow and temperature time series on a catchment scale are the major data available for model identification and thus any modelling approach that wishes to avoid the problems of over-parameterization of model parameters must restrict itself to using these data.

Lumped parameter models require less input data, pose little computational burden and hence their use is widespread. As with statistical and black box models, lumped parameter models cannot be applied with confidence to conditions not reflected in the calibration data set. This is because the lumping of processes is often oversimplified and ignores many complex process feedbacks. Moreover, they do not account for important spatial interactions that occur in catchments. An important factor in the application of lumped parameter models is the stability of the catchment system, stable spatial distribution of precipitation, vegetation and soil characteristics (Blackie and Eeles, 1985).

Refsgaard and Knudsen (1996) compared the ability of a lumped conceptual model, a semi-distributed model and a physically based model to simulate flows at ungauged catchments in Zimbabwe and concluded that, when the objective is to simulate time series of stream flow only, a lumped conceptual model gives the best results. The IHACRES which is a lumped conceptual model has been chosen for this study because of the relatively few parameters required for its calibration and its successful application in previous regionalisation studies. It seems capable of representing the hydrological processes at work in a variety of catchments and has been applied successfully in the United Kingdom (Sefton and Howarth, 1998), in Australia (Post and Jakeman, 1996 and 1999) and in North America (Kokkonen *et al.*, 2003).

2.7 Selection of Modelling Method

When selecting suitable techniques for assessment on a catchment scale basis, consideration should be given not only to the problem at hand but also to the limitations of data availability, degree of accuracy required and the available computing power. Simplicity should be balanced against quality and reliability of results (Muthusi, 2004). According to Woolhiser and Brakensiek (1982) the appropriateness of a model depends largely on the problem to be solved and will change as the problem changes. A technique with intensive data requirements may produce detailed results, but may be impractical to apply for a whole catchment. A technique which requires less data may be practical for catchment scale application, but may not produce results with the desired accuracy because the analysis may be over simplified (Acreman, 1994). When the aim is to simulate stream flows, simple models are adequate if historical data is available for model calibration. Simple models require few input parameters and have fewer problems arising from parameter uncertainties than complex models, making them suitable for regionalisation studies (Perrin *et al*, 2001)

2.8 Model Simulation

According to Beven (1989), the two main aims of model simulation are to explore the implications of making certain assumptions about the nature of the real world system and to predict the behaviour of the real world system under a set of naturally occurring circumstances. However, no matter how complex simulation models are, they remain extreme simplifications of reality. According to Clarke (1994) model simulation is an iterative process which starts by advancing a candidate model which, from graphical observation of the available data or otherwise, appears likely to represent their principal characteristics. The next step is the determination of the respects in which this model fails in its intended purpose. This involves plotting the fitted values given by the model and comparing them with the observed data set. The observed discrepancies between the fitted and observed values give an indication of how the model should be modified. The modified model output is again compared with the observed data and the process repeated severally until the two sets of data agree to within acceptable ranges.

2.9 Model Calibration

The values for the model parameters are selected once a suitable model for a given watershed has been defined. The process by which the parameters are selected is called model calibration (Sorooshian *et al.,* 1993). In recent times, a pragmatic approach to the identification of model parameter values has been adopted where it is assumed that some parameters especially those to which the model predictions are insensitive can be estimated. Other parameters are assumed to vary only on the basis of soil and vegetation type. This reduces the number of parameters actually supplied to the model to manageable numbers (Abbott *et al.,* 1986).All the parameters governing the various components in a model have some physical meaning and accepted ranges for their values. However, the exact values of the parameters cannot be fixed in advance hence the need for calibration, which is the process of estimating these parameters. The main purpose of model calibration (parameter estimation) is to obtain a parameter set for a particular catchment which gives the best possible fit between the simulated and the observed hydrographs. The importance of model calibration in regionalisation has been reported by Schaake *et al.,* (1997) who found out that the catchments in which regionalisation methods performed best corresponded to those catchments in which the best calibrations were obtained.

The process of calibration involves three steps namely:

1. Identification of the major model parameters (those which carry more weight in the model). This involves the determination of the physical meaning of the parameters, the significance of the parameters in the model and the range of values for the parameters.
2. A systematic way of changing the parameter values. This process can be done either automatically, manually or semi-automatically depending on the number of parameters under consideration.
3. Assessment of the model performance

The model parameters changed during calibration can be classified into two: physical and process parameters. The physical parameters represent physically measurable properties of the watershed such as surface area of the streams and surface slopes among others. The process parameters represent watershed properties that are not directly measurable for example the average or effective depth of surface soil moisture storage. Prior knowledge about the watershed properties and behaviour is made use of in specifying the initial estimates of the model parameters. The parameters are then typically fixed at these measured values and not adjusted further unless determined to be in error. For the process parameters, estimates of the range (minimum and maximum) of the possible values of these parameters are determined based on judgment and understanding of the hydrology of the watershed (Sorooshian *et al.,* 1993). To reduce the uncertainty of the parameter estimates the initial parameter values are selected somewhere in the ranges previously specified. The parameter values are then adjusted to more closely match the model behaviour to that of the watershed. The process of adjustment can be done either manually or using computer based automatic methods.

2.9.1 Manual Calibration

To calibrate a model, some aspects of watershed behaviour to which the model is to be matched are selected; typically, stream flow hydrographs at one or more locations on a river. The model parameters are then adjusted to get the simulated stream flow hydrograph to match the observed hydrograph for some historical data period (Sorooshian *et al.,* 1993). In manual calibration, trial and error process is used in parameter adjustments. After each parameter adjustment, the simulated and observed hydrographs are visually compared to see if the match has improved. Developments in computer graphics have made the process easier by enabling the effects of parameter adjustments to be rapidly observed and compared to the parameter trials, (Brazil, 1988). The main disadvantage of manual calibration is the absence of generally accepted objective measures of comparison. This makes it difficult to know when the process should be terminated, that is whether the 'best' possible fit has been obtained. Different operators may therefore obtain different parameter values for the same watershed (Beven *et al.,* 1995).

2.9.2 Automatic Calibration

Various factors have motivated the development of computer based methods for the automatic calibration of watershed models. Some of these factors are the need to speed up the process of calibration, the fact that there are few model calibration experts available for each watershed model and the need to assign some measure of objectivity and confidence to the resultant model predictions. Automatic calibration methods cannot however, entirely replace the manual methods since although they provide quick solutions; they still require user expertise and are typically used in conjunction with manual calibration procedures.

The main advantage of the automatic calibration procedure over the manual calibration is that it is relatively rapid, reduces subjectivity of the modeler and maintains proportionality between different parameters, which are established a priori, (Eckhardt and Arnold, 2001). According to Sorooshian *et al.,* (1993), a typical automatic parameter estimation procedure consists of four major elements: objective function, optimization algorithm, termination criteria and calibration data. In addition, process verification and sensitivity analyses are necessary to establish confidence in the results. A discussion of the various elements is given below.

2.10 Objective Function(s)

An objective function is an equation for computing a numerical measure of the difference between observed and simulated outputs. The purpose of automatic model calibration is therefore to find those values of the model parameters that optimize (minimise or maximise, as appropriate) the numerical value of the objective function. The choice of an objective function depends primarily on the intended application of the model and is therefore user-dependent (Diskin and Simon, 1977).

2.11 Optimization Algorithms

An optimisation algorithm is a logical procedure used to search the response surface constrained to the allowable ranges on the parameter values that optimize the numerical value of the objective function. The procedure is typically implemented on a digital computer to enable very rapid search using either the local, global or random search method. It is generally agreed that the quality of model parameters depends partly on the efficiency and robustness of the optimisation algorithm (Duan *et al.,* 1992). Considerable efforts have therefore been made over the last two decades to develop and implement complex and more efficient calibration techniques to cope with many parameters and highly non-linear model structures (Perrin *et al.,* 2001).

2.12 Calibration Data

Sorooshian and Gupta (1983) noted that proper choice of calibration data could reduce the difficulties encountered during calibration of hydrological models. Use of long period data marginally improves the parameter estimates. Generally, from statistical point of view, the data set used should be at least 20 times the number of parameters to be estimated. Qualitatively, the data set used should be informative. The quality of data is evaluated on the basis of the information (about the parameters) contained in the data and the noise (errors) in the data. The information content should be as large as possible while the noise should be as small as possible.

2.13 Model Validation

Model validation is the process of demonstrating that a given site specific model is capable of making sufficiently accurate predictions. It involves testing the ability of a model to simulate hydrologic response of a catchment for conditions different from those used during calibration period (Refsgaard and Knudsen, 1996). This implies the application of the calibrated model in simulation without changing the parameter values that were set during the calibration period. The model is said to be validated if its accuracy and predictive capability during validation period lie within acceptable limits.

CHAPTER THREE

METHODOLOGY

3.1 Data used in the Study

In the course of carrying out this research, different organisations were contacted to provide data and information on the Upper Tana basin for use in the study. Stream flow data was obtained from the Ministry of Water Resources and Irrigation for 30 gauging stations in the Upper Tana basin covering the period 1970-1990. After careful evaluation of the recorded data, 8 gauging stations with continuous flow records extending over the period 1970-1975 were selected for further analysis (table 3.1). The criteria used in selecting the stations to be used for further analysis was based on the quality of data, length of available records, spatial distribution of the stations within the study area and minimum computed percentage of missing data at a threshold value of 1%).

Table 3.1 Stream gauging stations used in the study

Station ID	River Name	Longitude (Deg)	Latitude (Deg)	Start Year	End Year	% of Data Missing
**4AC03	Sagana	37.04°	-0.45°	1970	1975	0
**4AD01	Gura	37.08°	-0.52°	1970	1975	0
*4BC02	Tana	37.21°	-0.67°	1970	1975	0.04
*4BC04	Rwamuthambi	37.24°	-0.67°	1970	1975	0
*4BE01	Maragwa	37.16°	-0.75°	1970	1975	0
*4CA02	Chania	37.06°	-1.03°	1970	1975	0
*4CB04	Thika	37.07°	-1.02°	1970	1975	0.08
*4CB07	Thika	36.75°	-0.75°	1970	1975	0

* Calibration Catchments ** Validation catchments

The Kenya Meteorological Department (KMD) provided data on rainfall for 130 stations in and around the study area covering the period 1970-1990. After careful evaluation of the recorded data 26, rain gauge stations with continuous records covering the period 1970-1975 corresponding to the selected stream gauging stations were adopted for further analysis based on maximum computed percentage of missing data of 10% of station record as recommended by the World Meteorological Organisation (WMO,1994). Since no monthly temperature data covering the area and period of study was available from the Kenya Meteorological Department, mean monthly temperatures were extracted from Kalders (1988) for use in calibrating the model. To identify which monthly temperature records were applicable to particular catchments, the temperature stations (appendix 1.2.) were converted to Arc View shape files using their position co-ordinates and then overlaid on the study catchments. Where more than one station was applicable in a catchment, the mean value was adopted as the monthly temperature for the catchment.

Soils data was extracted from the digital Soil map of the world (FAO, 1997) while the land use and land cover data for Africa was obtained from the USGS Earth Resources Observation System (EROS) data centre at URL http://edcaac.usgs. gov/glcc/glcc.html. The major soil and land cover types of the Upper Tana basin are shown in figures 3.1 and 3.2 respectively.

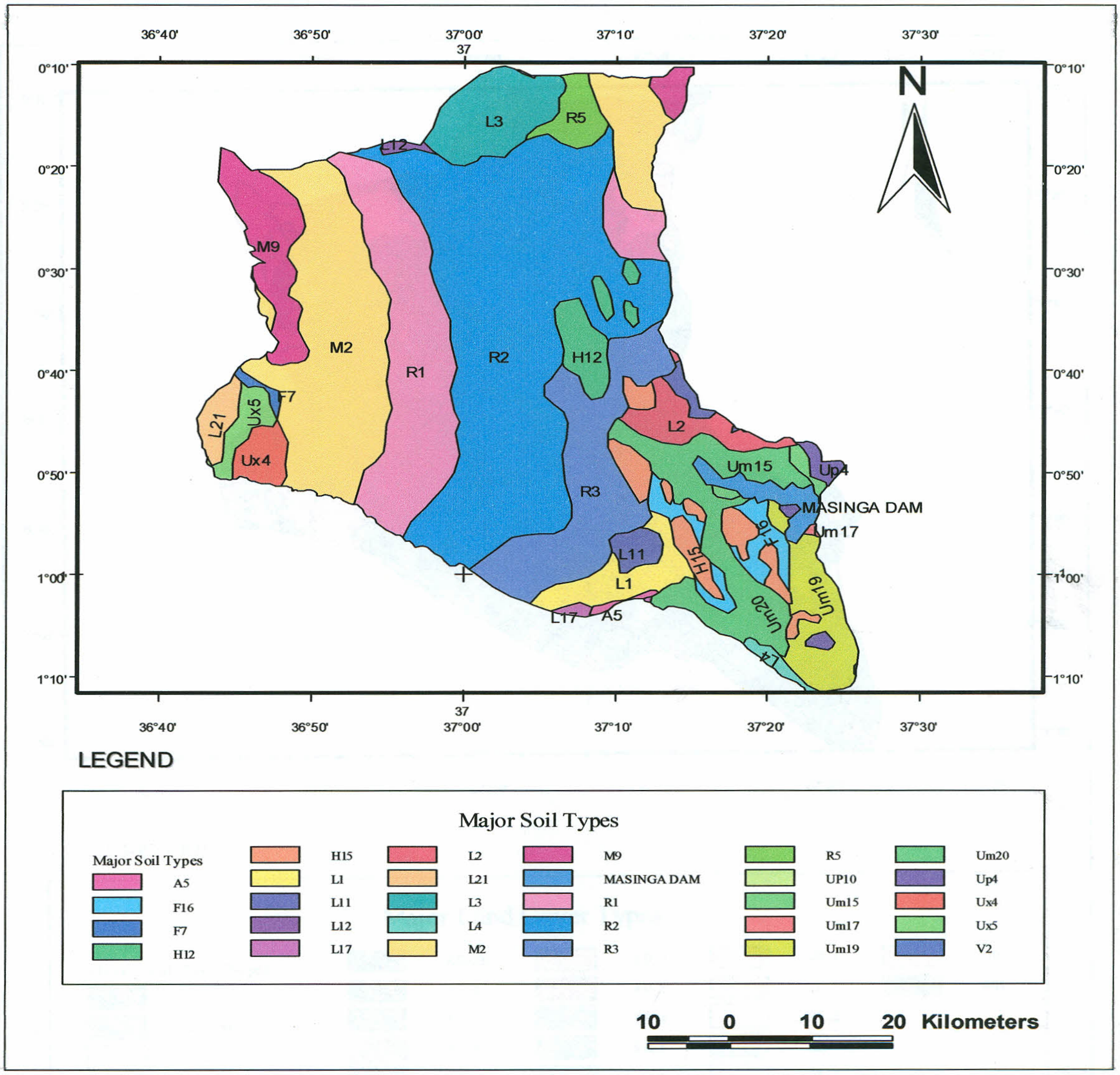


Figure 3.1 Major soil types of the Upper Tana basin

(Source, FAO, 1997)

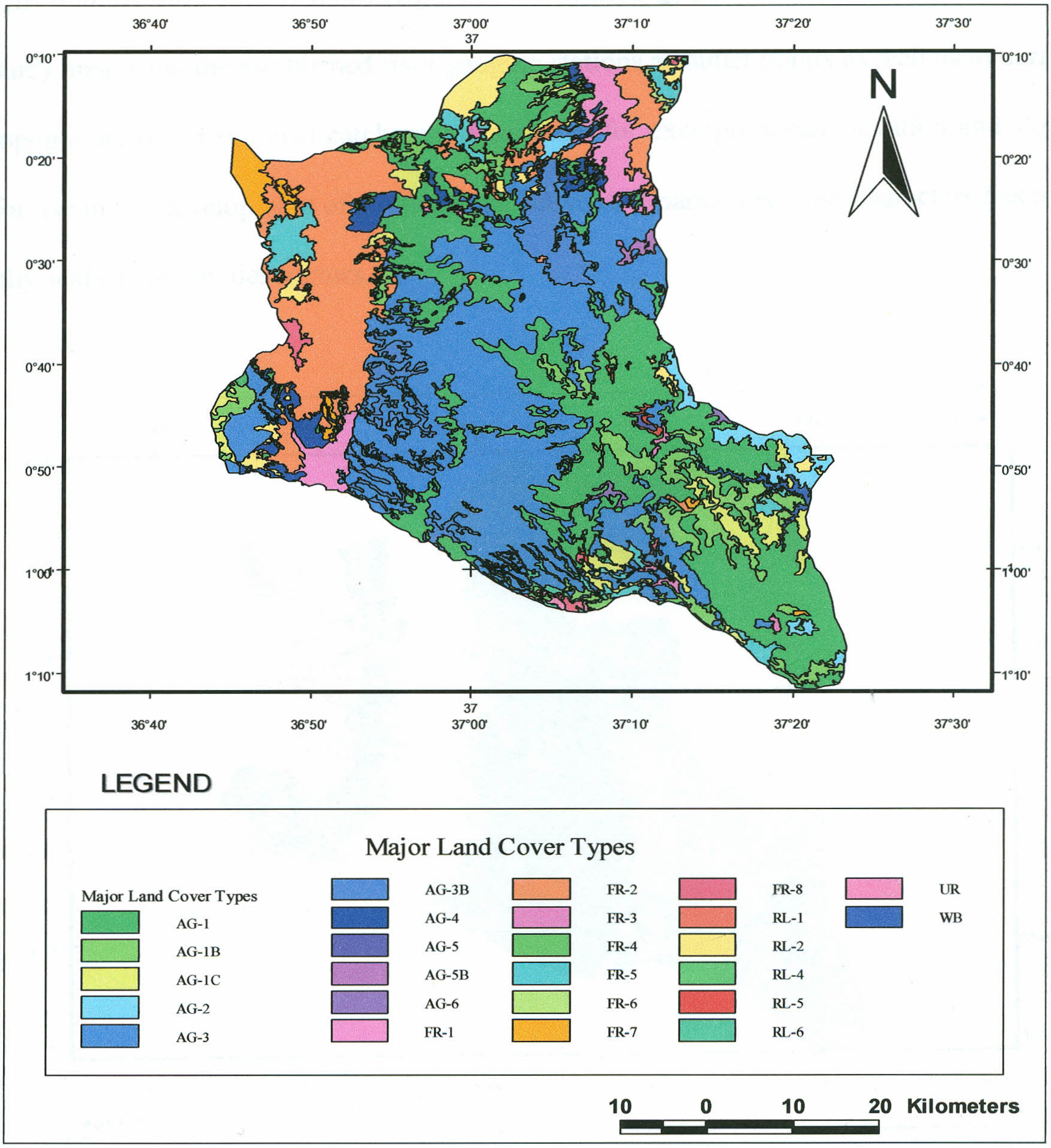


Figure 3.2 Major land cover types of the Upper Tana basin (Source: http: //edcaac.usgs. gov/glcc/glcc)

3.1.1 Upper Tana Basin Digital Elevation Model (DEM)

A Shuttle Radar Topographic Mission (SRTM) digital elevation model (DEM) at a horizontal spatial resolution of 90m was obtained from the United States Geological Survey's Earth Resources Observation System data centre at URL *http//srtm.usgs.gov*.

This 90m by 90m SRTM DEM (figure 3.3) was used to delineate catchments within the study area using the established river gauging stations as outlet points as well as to derive topography based physical catchment descriptors for example mean elevation and slope for use in the development of relationships between dynamic response characteristics and physical catchment descriptors.

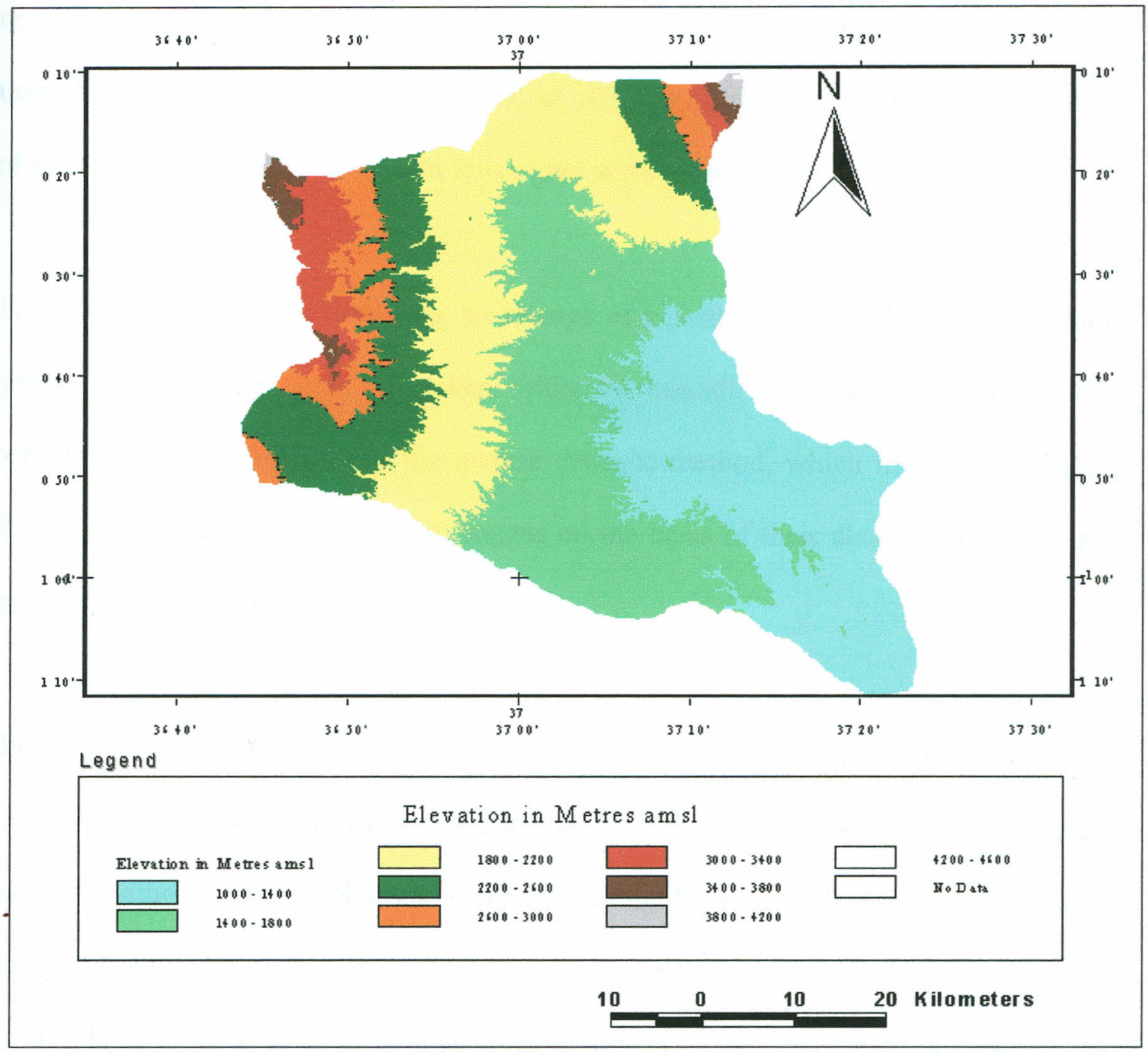


Figure 3.3 Digital Elevation Model (DEM) of the Upper Tana basin

(Source:http/srtm.usgs.gov)

3.1.2 Preparation of Raw Data

The raw rainfall and stream flow data obtained from the various government departments were checked to identify and eliminate any out-lier data points and to fill in missing values. This is because the quality of data has an effect on model calibration and validation and may introduce uncertainties to model predictions. There were gaps in some daily stream flows that were associated with seasons when the rivers were flooded or when station gauge's data was lost. Those stations that had a lot of missing data (more than 1% for gauging stations and more than 10% for rainfall stations) were dropped out of further analysis while those with fewer gaps were filled before further analysis.

To fill in missing stream flow data, an average of a particular day's flow over the entire period of record (20) years was taken, while for rainfall data, the missing value for a particular day was filled using the inverse distance method, which involves computing the weights of the surrounding rain gauges on the basis of their distances from the rain gauge with the missing rainfall data.

Based on rainfall data from 30 rain gauge stations, areal rainfall for each individual catchment was computed from the surrounding rainfall gauges using the Thiessen polygon method with the polygons constructed using ArcView GIS (figure 3.4). Thiessen polygon is a method that attempts to make allowance for irregularities in gauge locations by weighting the record of each gauge in proportion to the area which is closer to that gauge than to any other gauge. The average depth of rainfall is usually given by:

$$P = \frac{A_1P_1 + A_2P_2 + A_3P_3 .... + A_nP_n}{A_1 + A_2 + A_3 .... + A_n} \tag{3.1}$$

Where: P is the average depth of rainfall

P_1, P_2, P_n are the rainfall values recorded at the gauging stations

A_1, A_2 and A_n are polygonal areas around the rainfall stations

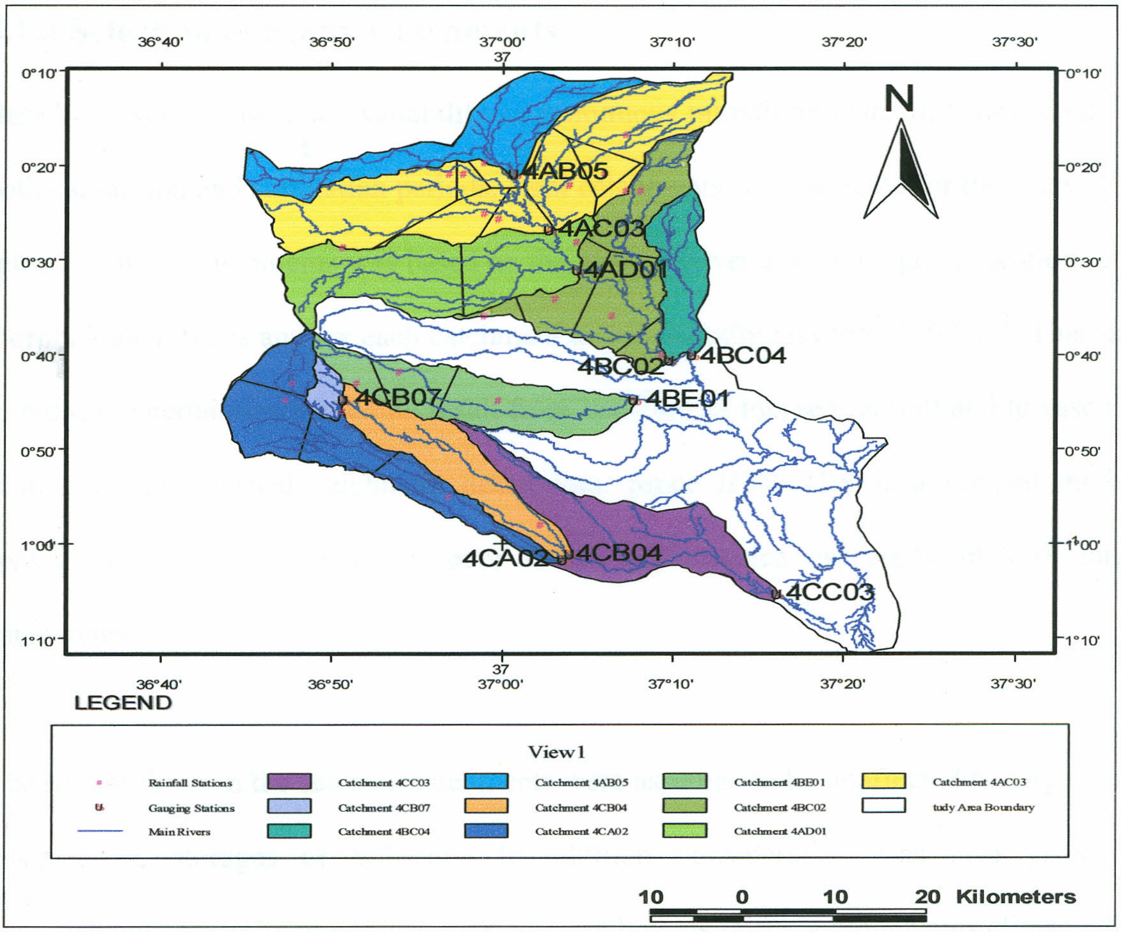


Figure 3.4 Thiessen polygons used in the computation areal rainfall

Using river coverages and the 90m by 90m SRTM DEM, mean elevation, slope, area and drainage density for the individual catchments were computed using the relevant extensions in ArcView GIS. These were required for the determination of physical catchment descriptors required in the derivation of DRC-PCD relationships. The soil and land cover coverages for the individual catchments were extracted from digitized Kenyan maps of the same and the percentage of individual catchment areas occupied by the respective soil and land cover types were computed using ArcView GIS. These were required in the computation of soil and land cover indices.

3.1.3 Selection of Study Catchments

Based on average size and availability of continuous rainfall and stream flow record for both calibration and simulation periods, eight catchments were selected for the study. The selection of the catchments was based on the need to cover a wide range of locations and morphological types and for each catchment to be generally less than 1000km^2. This was to reduce uncertainties that could result from handling of lumped rainfall and to ease the abstraction of physical catchment descriptors. Since IHACRES is a lumped model, spatial variability can result in poor model performance especially in very large catchments.

The stream flows in the selected catchments were assumed to be unaffected by large scale abstractions, storages or effluents. In addition consideration was also given to geographical proximity in order to make the catchments fairly homogeneous climatically, hydrologically and physiographically. In this way, the problems of model identification was minimised and relationships between catchment hydrological responses and physical characteristics made easy to detect, understand and interpret (Kokkonen, 2002). Attention was also given to catchment characteristics such that the selected catchments were representative of the statistical population of catchments in the study area both geographically and in parameter space. Six of these catchments were used for calibrating the model while the remaining two were used for model validation.

Although the study catchments were of similar scale and were geographically close to each other, they exhibited differences in land cover, topography, soil types and drainage density network structure. The study aimed at relating these differences to the variations in the hydrological responses of these catchments.

3.1.4 Structure of IHACRES model

The model applied is known as Identification of Hydrographs and Components from Rainfall, Evaporation and Stream flow (IHACRES), and the version is the PC-IHACRES v1.03 (Littlewood *et al.*, 1997, Littlewood *et al.*, 2002). The model uses an approach which attempts to capture identifiable catchment-scale dynamic response characteristics (DRCs) from data in its parameters. It comprises of two modules in series:

1. A non-linear loss module that links rainfall and air temperature ($\mathbf{r_k}$ and $\mathbf{t_k}$) to effective rainfall ($\mathbf{u_k}$) parameters $\mathbf{c}$, $\boldsymbol{\tau_w}$ and $\mathbf{f}$. It uses temperature and rainfall data to estimate the relative catchment moisture index which determines the proportion of rainfall that becomes effective rainfall.
2. A linear unit hydrograph (UH) module that links effective rainfall $\mathbf{u_k}$ to stream flow $\mathbf{x_k}$ parameters $\boldsymbol{\tau^q}$, $\boldsymbol{\tau^s}$ and $\mathbf{v^s}$. It routes the effective rainfall through any configuration of stores in parallel and/or in series, which is identified from the time series of rainfall and stream flow data but is typically either one store only, representing ephemeral streams or two

stores in parallel allowing both slow and quick flows to be represented (Croke *et al.*, 2005) as shown in figure 3.5.

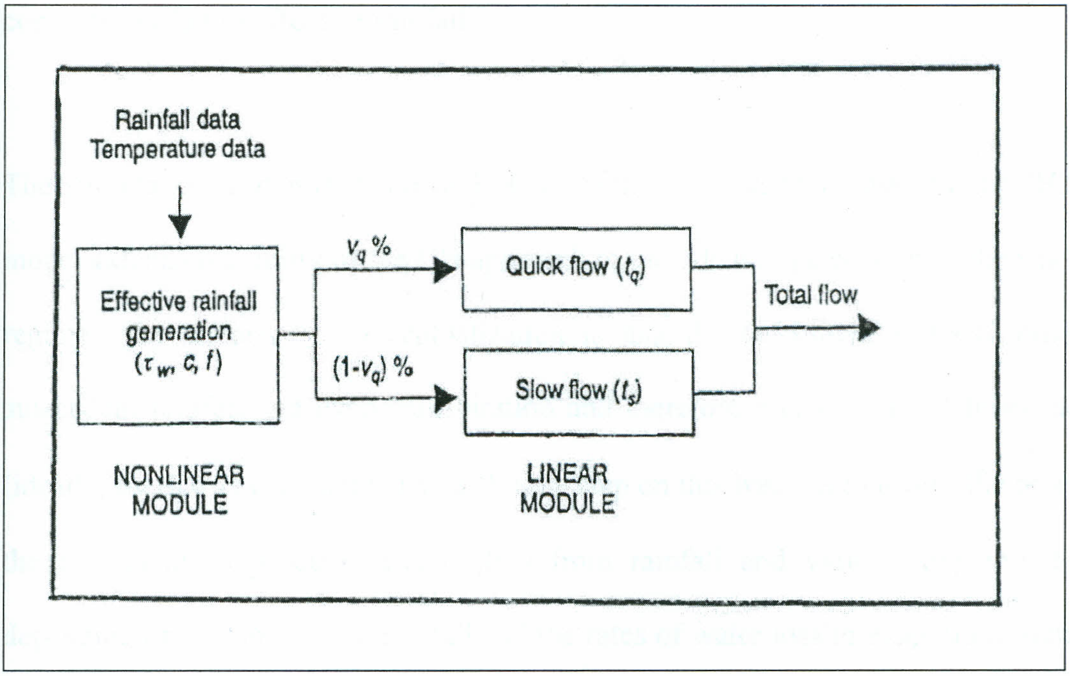


Figure 3.5 Schematic representation of the IHACRES model

Source: Kokkonen *et al.*, (2003)

The IHACRES is a lumped conceptual model, which attempts to simulate rainfall-runoff response of catchments as total stream flow. It uses temperature and rainfall data to estimate stream flow, with parameters calibrated prior to simulation by comparison with observed stream flow data (Jakeman and Hornberger, 1993). All of its response characteristics can be identified from the raw data and the model parameters are not systematically related to the climatic sequence in the calibration period. It assumes that there is a linear relationship between effective rainfall and stream flow. Effective rainfall $\mathbf{u}_k$ for time step k is that part of rainfall which is not lost to evapotranspiration and therefore eventually leaves the catchment as stream flow, $\mathbf{x}_k$. This assumption allows the application of the UH theory which conceptualizes the catchment as a configuration of linear stores acting in series and/or in parallel. The non-linearity normally observed between rainfall and stream flow is therefore accommodated in the module, which converts rainfall to effective rainfall.

The simultaneous identification of UH for both high and low flows by the IHACRES model extends the utility of the UH approach to include a large portion of the whole flow regime. The underlying conceptualization is that the catchment wetness varies with antecedent rainfall and evapotranspiration and therefore a catchment wetness index, S_k (ideally, $\mathbf{0 < S_k < 1}$) is computed at each time step on this basis. It indicates the potential of the catchment to produce stream flow from rainfall and varies from zero to unity depending on the antecedent rainfall and the rates of water loss to evapotranspiration and stream flow. A value of zero indicates a relatively dry catchment with rainfall falling at this time producing no effective rainfall while a value of unity indicates that the catchment is relatively saturated and all rainfall falling at this time will become effective rainfall. The index $\mathbf{S_k}$ is given by:

$$S_k = \frac{r_k}{c} + \left[1 - \frac{1}{\tau_w(t_k)}\right] S_{k-1} \qquad (3.2)$$

Where c is a parameter which determines the impact that a unit input of rainfall has on the catchment storage, r_k is rainfall and $\tau_w(t_k)$ is the time constant (days) of catchment losses at daily mean temperature t_k (°C) according to:

$$\tau_w(t_k) = \tau_w \exp(20f - t_k f) \qquad (3.3)$$

Where τ_w is the time constant (days) of catchment losses at 20°C. The constant 20°C is a reference temperature chosen depending on conditions in the study area and may therefore vary while f is a factor describing the effect of a unit change in temperature on the loss rate. Effective rainfall u_k is then calculated according to:

$$u_k = \frac{1}{2}\left(s_k + s_{k-1}\right)r_k \qquad (3.4)$$

Once the effective rainfall has been calculated, the total unit hydrograph is determined by parameterising and discretizing the following linear convolution as a series of linear reservoirs:

$$y(t) = \int_0^t h(t-s)u(s)ds \qquad (3.5)$$

Where y(t) is the observed stream flow, u(s) is the effective rainfall and h(t) is the unit hydrograph. The parameters describing the unit hydrograph are determined by a simple refined version of the instrumental variable technique described in Jakeman *et al.,* (1990).

Although any configuration of reservoirs in series or in parallel is possible for the model, the most commonly identified structure is the two linear reservoirs acting in parallel and having different response times to rainfall and hence referred to as quick and slow flow components of stream flow. The model has six parameters, three relating to the non-linear component (τ_w, f, and c) and three relating to the linear component (τ_q, τ_s and v_s).

The percentage of rainfall which becomes effective rainfall in any time step varies linearly (between 0 % and 100 %) as the catchment wetness index, $\mathbf{S_k}$ varies between 0 and unity). The advantage of the spatially 'lumped' approach of the model is that it requires only six parameters, three in the non-linear module to model rainfall to rainfall excess and three in the linear module of routing rainfall excess to stream flow.

A model with a few well-defined parameters gives better statistical relationships between DRCs and the selected PCDs. Parameter $\boldsymbol{\tau_w}$ is the rate at which the catchment wetness declines in the absence of rainfall. Hence a large $\boldsymbol{\tau_w}$ gives more weight to the effect of antecedent rainfall on catchment wetness than a smaller one. The value of parameter **c** is set such that the volume of rainfall excess is equal to the total stream flow volume over the estimation period. It is the increase in catchment wetness index per unit rainfall in the absence of any decrease due to evapotranspiration. The parameter **f** is a temperature modulation factor which determines how $\boldsymbol{\tau_w (t_k)}$ changes with temperature for a constant $\mathbf{s_k}$. The parameters τ_q and τ_s are the recession time constants for quick and slow flow components respectively. The volume $\boldsymbol{\nu_q}$ is the proportional volumetric contribution of quick flow to total stream flow while $\boldsymbol{\nu s\ (1-\nu_q)}$ is the proportional volumetric contribution of slow flow to total stream flow. The quantities $\mathbf{c}$, $\mathbf{f}$, $\boldsymbol{\tau_w}$, $\boldsymbol{\tau_q}$, $\boldsymbol{\tau s}$, $\mathbf{v_s}$ are considered as the DRCs of the catchment as shown in figure 3.6.

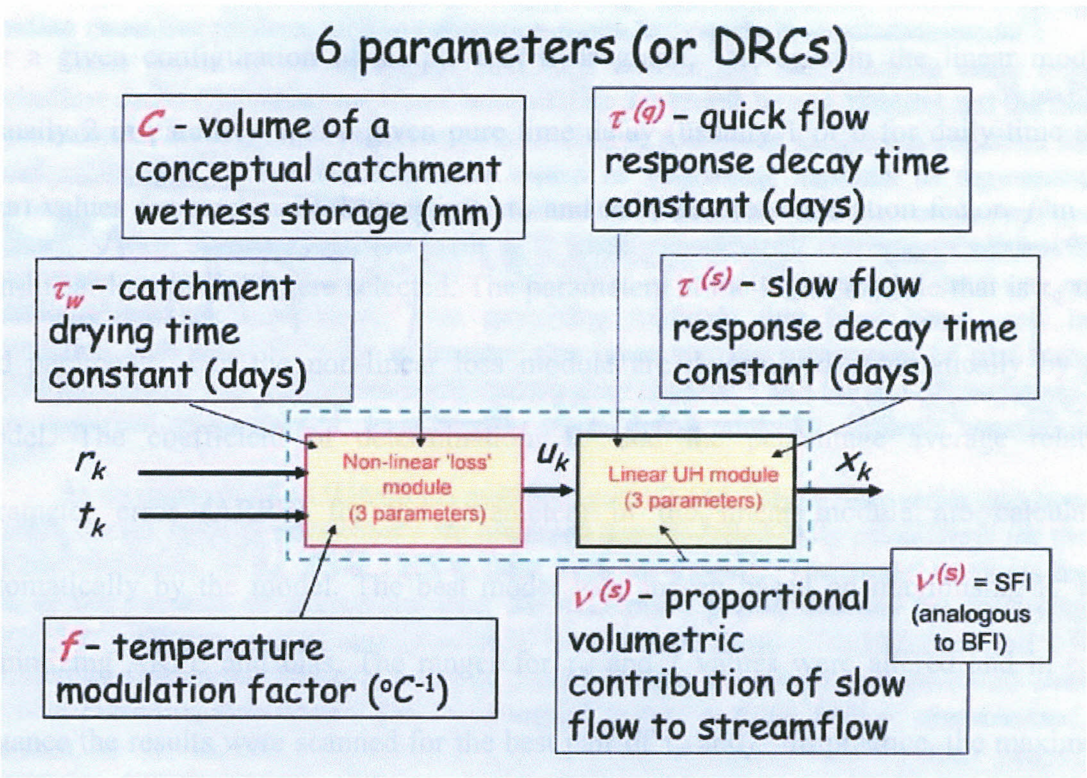


Figure 3.6 IHACRES model structure and DRCs

(Source: Littlewood *et al.*, 2002)

Generically, IHACRES is a "data based mechanistic" type model (Young, 2002) and, as such, is able to make efficient use of existing data sets. For calibration, the model requires time series of rainfall, stream flow and temperature data. Its parametric efficiency makes it easy to relate its DRCs to PCDs making it useful in regionalisation studies. Despite its structural simplicity, the IHACRES model has been applied successfully to a wide range of catchment types and for regionalisation studies, (Post and Jakeman 1996, 1999; Kokkonen *et al.*, 2003: Post and Croke, 2002). In practice, selection of the 'best' IHACRES model is based on performance in both calibration and simulation (Littlewood *et al.*, 2002).

3.2.1 Calibration and Validation of IHACRES Model

For a given configuration of simple unit hydrograph, storages in the linear module (usually 2 in parallel) and a given pure time delay (usually 1 or 0 for daily time step data) values for catchment drying time, τ_w and temperature modulation factor, f in the non-linear loss module were selected. The parameters in the linear module that is τ_q τ_s v_s and parameter, **c** in the non-linear loss module are determined automatically by the model. The coefficient of determination, R^2 and the percentage average relative parameter error (ARPE) for the parameters in the linear module are calculated automatically by the model. The best model was chosen based on maximising R^2 and minimizing ARPE and bias. The ranges for τ_w and f values were altered and in each instance the results were scanned for the best pair of τ_w and f. In practice, the maximum R^2 and minimum ARPE define ranges of the catchment drying time constant ($\tau_{w)}$ and the temperature modulation factor f and a subjective trade off was made between a high R^2 value and a low ARPE in selecting an optimum pair. This procedure gave an invariant model parameter set in the sense that any experienced model user would choose the same set of optimised parameters.

Once an optimal pair had been identified for a given configuration of simple unit hydrographs (UHs) and a given value of pure time delay **(δ)**, different configurations of UHs storages in the linear module, and different values of δ were applied to confirm that the optimum combination of these factors had been selected. These model calibrations were then assessed in terms of their ability to reproduce the observed stream flows by using them to simulate stream flows from an independent period of record.

The study catchments were divided into two sets, one set for model calibration (six) and another set for model validation (two) (table 3.1). For each individual catchment, historical daily rainfall, stream flow and monthly temperature data were used to calibrate and validate the IHACRES. To account for climatic variability, the model calibration was performed on daily rainfall- stream flow time series during the common period of observation, 1970-1975 for all the six catchments. The 1970s was selected for modelling because of the reasonably good rainfall and stream flow data that was obtained for the decade compared to other decades.

Because of the wide range of study catchments in terms of size and response times, a daily time step was adopted for calibrating the model. This was because of the need to capture the quick flow dynamics of small catchments while at the same time avoiding over-parameterisation of the model. At the limits of catchment size however, some bias may be introduced, for example in τ_q at small catchments where a daily time step may be too coarse to characterize the quickest response (Post and Jakeman, 1996)

Firstly, the period of record for each catchment was divided into three non-overlapping two year calibration periods. Selection of a two year calibration period (Jakeman, *et al.*, 1993) allowed each model to be exposed to some inter-annual variability while at the same time ensuring that the hydrological response of the catchment does not change dramatically within the calibration period. It also helped to balance problems of variance and bias. Shorter calibration periods generally yield model parameters with very high variance whilst longer calibration periods are likely to encompass changes in the system, such as vegetation cover, or changes in measurement variability such as shift in the stage-discharge rating curve.

However, although a two year calibration period may be sufficient in humid catchments (as in the present study), in arid and semi-arid catchments longer calibration periods may be required as flow events in these catchments are usually less frequent (Croke *et al.,* 2005). For the six catchments, models were calibrated for three periods of two years each (1970-71,1972-73 and 1974-75) and validated on the other two periods (4 years) as well as on the entire period of study (1970-75). Since the IHACRES model assumes an initial catchment wetness index (S_k) of zero, each calibration period was made to start and end on a low flow and for that reason January was selected as the starting month for all calibrations since in Kenya, January is generally considered to be a dry month.

The parameters used in evaluating the performance of the models were: R^2 (a measure of the goodness of fit), bias (the difference between mean observed and modelled flow indicating where there was systematic overestimation or underestimation of flow) and the average relative parameter error, ARPE which combines the efficiency of parameterization and goodness of fit (Sefton and Howarth, 1998). The Nash Sutcliffe (1970) efficiency E was used to assess the deviation of observed stream flow from modelled stream flow. After multiple period calibrations, the third sub-period 1974-1975 was found to give the best calibration results for all catchments and when simulated over the entire period of study (1970-1975) also gave good results. In simulation the unit hydrograph derived for each model was linearly convoluted with the effective rainfall determined by inserting the raw rainfall and temperature records in equations 3.2–3.4 for the entire period of study and the resultant modelled stream flow compared with the corresponding observed stream flow.

The model parameter set obtained from calibration using the 1974-1975 data were therefore considered as the best model and used to calibrate all the catchments. The best model was that which provided a good simulation of stream flow over the entire period and also fitted the observed stream flow closely for the calibration period. The parameters corresponding to this model were therefore those, which on average, gave the best approximation of stream flow for each catchment. These parameters were considered as the dynamic response characteristics of these catchments and were then related with the physical attributes of these catchments.

For the purpose of demonstrating that the calibrated model was capable of making accurate simulations, a model validity test was performed. This involved keeping the model parameter values set during calibration constant and running the model in simulation mode. The resulting simulated stream flow was then compared with the observed (recorded) stream flow and the goodness of fit evaluated using both visual and numerical methods. The calibrated model for each catchment for the period 1974-75 was used to simulate stream flows over the entire study period (1970-75) the first four years (1970-73) of which served as validation since they were not used in the initial calibration.

3.2.2 Derivation of Physical Catchment Descriptors

Since IHACRES is a conceptual model no field measurements of the physical catchment attributes were carried out and therefore only those catchment properties that could be evaluated from existing data sources were considered. The information available on the physical catchment attributes was therefore limited; a situation that is common in applied research. However, this introduced some degree of subjectivity into the quantification of some of the catchment characteristics.

Catchment attributes used in the study were those that were considered to characterize well the factors that drive the hydrologic response of the catchments and were easily derivable from existing data sources. A plethora of different descriptors have been used in literature in search of connections between the hydrological response of catchments and some observable indices of its physical properties. Sefton and Howarth (1998) and Post and Jakeman (1996), contain a list of physical catchment descriptors applicable in studies of this nature, some of which are given in table 3.2

Physical catchment indices representing physical attributes of the catchment were therefore derived for the eight catchments using the DEM, topographical, soil and land cover maps to represent topography, soil and land use. In total, 14 PCDs (5 topography based, 5 land cover based and 4 soil type based) were selected on the basis of probability of success, low correlation with other indices and wide availability. The inclusion of indices that change with time (land cover) allowed hydrological response to be affected by changing land use, an implicit assumption being that the derived relationships between DRCs and PCDs were constant and that change in land use did not alter the hydrological responses of these catchments or the processes governing them. The methods used to quantify the PCDs, are described in the following sections.

Table 3.2 List of commonly used physical catchment descriptors

Soil related descriptors	Land cover related	Topography related
Ground water (%)	Moorland Grass (%)	Elevation
Peaty soils (%)	Heath (%)	Mean slope
Semi-permeable soils (%)	Deciduous woodland (%)	Area
Mineral soils (%)	Coniferous woodland (%)	Channel slope
Shallow ground water (%)	Upland & lowland bog (%)	Stream frequency
Deep ground water (%)	Arable crop land (%)	Aspect
No ground water (%)	Urban development (%)	Elongation
	All woodland (%)	Drainage density
		Stream gradient
		Main stream length

Source: Sefton and Howarth (1998); Post and Jakeman (1996)

3.2.2.1 Topographical Indices

1. The drainage density index (*D_Dens*) was obtained by dividing the total stream length within the catchment by the catchment area. It is a measure of how dissected a basin is and is an important landscape attribute since it affects the transformation of rainfall into runoff (Mazvimavi, 2002).
2. The mean catchment elevation (*Elev.*) and mean catchment slope (*Slope*) were computed directly for each catchment using the 90m by 90m SRTM DEM and the relevant extensions in Arc View GIS. Slope is an important characteristic of a catchment as it gives an indication of the kinetic energy available for water to move towards the basin outlet, and has been found to be related to total runoff and to base flow (Vogel and Kroll, 1992).
3. The catchment area (*Area*) defined as the area draining to the catchment outlet was determined directly for each catchment using ArcView GIS.
4. The stream gradient (*S_Gradient.*) was determined by dividing the total length of the main stream with the elevation difference between its highest point and the outlet point.

3.2.2.2 Geology and Soil Indices

Geology affects stream flow in at least two ways, the first being direct in the sense that ground water is stored in rocks, especially if they are fractured and this contributes to stream flow. The second effect is on soil formation as different rocks tend to produce different types and depths of soil under the influence of weathering and plant action and hence differences in recharge, ground water and stream flow. The derivation of quantitative geological indices that express the geological effects on runoff processes at the basin level is a major challenge in hydrology especially in developing countries where hydrogeological data is rarely available. Due to lack of quantifiable geological data for the study area, geology was omitted from consideration as a physical catchment descriptor.

Because of the great variability of soil profiles in most catchments of the study area (especially between ridge tops and valley bottoms) it was not possible to attribute a particular predominant soil type to any catchment. Percentage coverages of various soil types in individual catchments were therefore extracted from a digital and terrain data base of East Africa map (FAO, 1997) at a scale of 1:1000, 000 by overlaying them with the individual catchments areas using ArcView GIS. The percentage areal coverages for the various soil types were divided by the total catchment area to obtain a dimension less soil index for each soil type in all the study catchments.

3.2.2.3 Land Cover/ Land Use Indices

The land cover and land use data along with soils data determine the amount of excess rainfall, the amount of recharge to the ground-water system, and the amount of water held in storage in the soils. It has been shown in several studies to affect catchment flow characteristics (Bosch and Hewlett, 1982) and was therefore selected for this study. The percentage coverage of land cover types was derived from vegetation and land use data available on digitized Africover land cover maps at a scale of 1,000,000. The derived land cover areas for individual catchments were used to compute a dimensionless index by dividing them with the total catchment area. For analytical purposes, each land cover index was expressed as a fraction of the total catchment area.

3.2.2.4 Climatic Indices

Since climate requires many parameters for its adequate description some of which are difficult to measure due to unavailability of the relevant equipments, climate was not considered in the present study. In addition it was assumed that the rainfall runoff model IHACRES is capable of filtering out, to a sufficient extent, any effects arising from different climatic descriptors Kokkonen *et al.,* (2003).

3.2.3 Derivation of a Regional Model Parameter Set

A regional model parameter set was derived to test the functional relationship between physical catchment descriptors and model parameter values that would allow simulation of daily stream flow in ungauged catchments of the upper Tana basin and also to enable discussion on the physical basis of the IHACRES model. It involved the following sequential steps:

3.2.3.1 Development of DRC-PCD Relationships

Generally, more than one catchment attribute will influence any one particular aspect of the hydrological response of a catchment. However, for exploratory purposes and to identify which catchment attributes were important, relationships were determined between individual catchment attributes and the catchment response characteristics. In addition interactions between catchment attributes and between model response characteristics which needed to be understood were also determined. This was done to avoid drawing misleading relationships.

For an effective model, the developed relationship should contain independent variables, be statistically significant, physically sensible, yield good estimates of calibrated DRCs and more importantly good estimates of observed flows. The relationships were developed based on DRC and PCD data from six catchments, the remaining two being for statistical model validation. Acceptance or rejection of variables was based on both statistical and hydrological significance, such that statistics were not used to justify hydrologically improbable processes and conversely were not ignored in the quest for an expected relationship. Statistical significance was evaluated for each independent variable in the regression and for the relationship as a whole. The physical processes by which each variable affected the model parameters were identified as well as the relative magnitudes of the regression coefficients. The strategy adopted for developing DRC-PCD relationships involved a progression of techniques, from inspection and correlation analysis, through stepwise regression and multiple regression while taking into consideration all criteria for a good relationship. The derived relationships were expected to have some meaning in the context of individual parameter conceptualizations and not to be mere mathematical abstractions. Preliminary analysis involved investigations into correlation within DRCs and PCDs and between DRCs and PCDs data sets.

3.2.3.2 Stepwise Regression

This is one of the most commonly used procedures for selecting the 'best' regression equation and involves building the equation one variable at a time by adding at each time step the variable that explains the largest amount of the unexplained variance. After each step, all the variables in the equation are examined for significance and discarded if they don't explain a significant amount of the variation

The first variable added is normally the one with the highest simple correlation with the dependent variable and the second variable added is the one explaining the largest variation in the dependent variable that remains unexplained by the first variable added. At this point the first variable is tested for significance and discarded or retained depending on the results of the test. The third variable added is the one that explains the largest portion of the variance not explained by the variables already in the regression equation. The variables in the equation are again tested for significance and the procedure continued until all variables not in the regression equation are found to be insignificant and all variables in the equation are found to be significant.

The quality of the resulting regression model was judged by its ability to predict the dependent variables for observations on the independent variables not used in estimating the regression coefficients. To make adequate comparisons of this nature the 'split record test' procedure was adopted where the data was divided into two parts; one part for developing the regression model and the other part for testing the model. Catchments 4AD01 (Gura) and 4AC03 (Sagana) were selected for validation purposes.

3.2.3.3 Multiple Regression

Multiple regression is commonly used when one dependent variable and several independent variables are available and it is desired to find a linear model for predicting unobserved values for the dependent variable. A model has therefore to be developed that does not have to contain all of the independent variables. The development of the right model is normally complicated by the fact that in most cases the independent variables are not statistically independent of each other but are correlated. The first step in regression analysis therefore is to determine the correlation matrix of the independent variables.

When two variables in a regression equation are linearly related, they both attempt to explain the same thing in the model such that when both are in the model, part of the variation in the dependent variable that either would explain when used alone may be split between them in such a way that neither will be significant. Retaining variables in a regression equation that are highly correlated therefore makes the interpretation of the regression coefficients difficult. This is because sometimes the sign of the regression coefficient may be the opposite of what is expected if the corresponding variable is highly correlated with another independent variable in the same regression equation.

It is important therefore when selecting a multiple regression model to perform several regressions on a given data set data using different combinations of the independent variables and then choosing the regression model that best fits the data as indicated by the largest value of R^2. The variables retained in a regression equation should make a significant contribution to the regression as well as have a physical significance. With the advent of digital computers it is now possible to perform many regressions on large data sets using a variety of techniques such as stepwise regression.

A stepwise regression was therefore applied in this study to identify those PCDs with the strongest statistical relationships with the DRCs. Although such relationships were purely empirical, they were useful in considering PCDs that could have been overlooked or obscured during correlation analysis.

For each DRC, the variables suggested by correlation analysis, and stepwise regression were combined in a multiple regression model enabling each DRC to be expressed in terms of PCDs as follows:

$$y(k) = \beta_o + \beta_1 x_1(k) + \beta_2 x_2(k) + \text{-----------------------------------} + \beta_p x_p(k) + \varepsilon(k) \qquad (3.6)$$

With k = 1, --------, N Where y (k) represents DRC; $x_1(k)$, $x_2(k)$--------$x_p(k)$ denotes the PCDs in the k^{th} model simulation; N represents the total number of catchments; $\beta_o, \beta_1, \beta_2$-----------$\beta_p$ denotes the ordinary regression coefficients obtained by minimizing the regression residual ε and p is the number of PCDs used in the regression model.

3.2.3.4 Validation of DRC –PCD Relationships

A model developed for the purpose of estimating stream flows in ungauged catchments and simulating observed conditions within a strict level of significance is of little value if it cannot be applied to other catchments within the region where the region is delineated both geographically and as a parameter space defined by the limits of the set of PCDs of the calibration catchments (Mwakalila, 2003). The usefulness of the developed estimation equations was therefore tested by re-calculating the model parameter values (DRCs) of all the calibration catchments as well as another two catchments not used in their derivation Gura at 4AD01; area 604km^2 and Sagana at 4AC03; area, 1035km^2. A statistical comparison was then made between the calibrated parameter values and those obtained through estimation using the derived DRC-PCD relationships (table 4.12)

The successful development of the relationships between DRCs and PCDs makes it possible to regionalise IHACRES model by providing a regional parameter set where model parameter values vary with catchment characteristics that can be measured or determined. The characteristics so determined can then be used with the developed relationships to estimate model parameters which, together with daily time series of rainfall and temperature, can be used to simulate daily stream flows in any ungauged catchment of the Upper Tana basin using the IHACRES model. In addition, relating model parameters to catchment characteristics makes it possible to analyse the physical basis of the IHACRES model. Both stepwise and multiple regressions were carried out using the SYSTAT software and the graphs plotted using Microsoft Excel Spread sheet programme.

3.2.4 Estimation of Daily Stream Flows

Since the six dynamic response characteristics (DRCs) fully define the IHACRES model, the estimated DRCs were used together with the daily rainfall and monthly temperature data time series to simulate daily stream flow of Gura (4AD01) and Sagana

(4AC03) respectively for the period 1970-1975 assuming they were ungauged for stream flow. However, since these catchments are in fact gauged, the estimated daily stream flows were then compared with the observed (recorded) daily stream flows for the same period (1970-1975) in order to assess the accuracy of the estimations.

The assessment of the accuracy of these estimations was done by comparing the values of the coefficients of determination, R^2 of estimated and observed stream flows as well as visual inspection of time series plots of observed and simulated stream flows using the estimated model parameters.

The coefficient of determination R^2 is defined by:

$$R^2 = 1 - \frac{\sum (y_k - x_k)^2}{\sum (y_k - y)^2} \tag{3.7}$$

Where: y_k is the observed stream flow at time step k

y is the mean observed stream flow

x_k is the simulated stream flow using estimated DRCs

As the model fit improves the co-efficient of determination approaches unity.

Further validation of the DRC-PCD relationships was by estimating DRCs for the six calibration catchments and using these DRCs together with daily rainfall and monthly temperature time series for the period 1970-75 to simulate daily stream flow. The resulting simulated stream flows were then compared with those obtained using calibrated DRCs as well as with the observed stream flows.

In their application to estimate stream flows in ungauged catchments it was assumed that the established relationships do not change spatially within the region. However, it was acknowledged that the relationships implicitly accounted for features of the physical nature of the catchments and the driving variables that were relatively uniform within the region. This implied that the relationships were invalid outside the regions where these factors differed significantly.

3.2.5 Model Sensitivity Analysis

Sensitivity tests in which model parameters are varied over their calibration range are used to find out the effect of change of the various model parameters on the hydrologic response of catchments. Such tests serve as a further validation of the relationships in that effects are explainable in physical terms using PCDs linked to each varied DRC. In the process the less sensitive parameters whose mean values can be adopted are identified. It is a necessary step in extrapolation and regionalisation of models. During sensitivity analysis the parameter under investigation is varied between its minimum and maximum boundaries while keeping the other parameters constant at their optimal (calibrated) values and calculating the efficiency. A graphical plot of percentage change in efficiency against percentage change in parameter value is then used to reveal the differences between the sensitivities of different model parameters.

The IHACRES model package automatically determines the values of the parameters τ_q,τ_s, v_s and c. The other two parameters τ_w and f are usually determined by the modeler through a trial and error procedure, whereby all possible values of these parameters are tried and optimum values selected by evaluation of a range of model performance statistics. These parameters were therefore varied during sensitivity analysis and the resulting change in simulated stream flow for each parameter change noted. The two parameters, that is f and τ_w were varied by increasing/decreasing their values at intervals of 50% while keeping the other parameter constant and the new parameter values used to simulate stream flow. A comparison was then made of the simulated flows before and after the variations in order to assess the impact of individual parameter change on stream flow simulation. The aim was to determine which of the parameters was more sensitive in terms of its effect on stream flow simulation.

3.2.6 Assessment of Model Performance

Comparison of predicted and observed hydrographs is a necessary step in assessing the performance of any model. An objective function is one of the common means of evaluating model performance. It is an equation used to compute the numerical measure of the difference between simulated model output and the observed (measured) catchment output.

3.2.6.1 Least Squares Method

Drawing from statistical regression and model fitting theory, the most commonly used objective function has been some form of the weighted least squares (WLS) function:

$$f(\theta) = \sum_{t=1}^{n} Wt * \left[q_t^{obs} - q_t(\theta) \right]^2 \tag{3.8}$$

Where q^{obs} = Observed (measured) stream flow value at time (t)

$q_t(\theta)$ = Simulated model stream flow value at time (t)

θ = Vector model parameters

Wt = weight at time t

n = Number of data points to be matched

The weights Wt indicate the importance to be given to fitting a particular hydrograph value. Setting all the weights equal to zero reduces the WLS to the simple least squares (SLS) function. The minimum value of f (θ) that can be attained is zero, if the model is able to perfectly reproduce the observed stream flow hydrograph.

3.2.6.2 Maximum Likelihood Method

The most common form of maximum likelihood criteria has been the Heteroscedastic Maximum Likelihood Estimator (HMLE) which accounts for the non-stationary variance in the stream flow measurement errors. It is a maximum likelihood minimum variance estimator, which in its simplest form can be computed as:

$$\text{HMLE} = \frac{\frac{1}{n} R_d}{\exp\left[2(\lambda - 1) a_d\right]} \tag{3.9}$$

Where: n = number of data points

λ = Unknown transformation parameter, which stabilizes variance

$$R_d = \sum_{t=1}^{n} Wt \in_t^2 \tag{3.10}$$

$\in_t$ = Model residual at time t

$= q_t^{obs} - q_t(\theta)$ (observed and simulated flows respectively)

Wt= weight assigned to time t

$= f_t^{2(\lambda-1)}$

f_t = expected flow ($q_{t,\,true}$) at time t

3.2.6.3 Correlation Coefficient (r)

Correlation analysis is a statistical tool used to assess the degree of dependence between variables. Correlation coefficient (r) summarises in one number the direction and magnitude of the correlation. It is another numerical means of assessing the performance of a model. The value of the coefficient, r, varies from zero to unity; with the highest value unity indicating the best performance i.e. $|r| \leq 1$ and

$$r_d = 1 - \frac{S^2}{\sigma_{y^2}} \qquad (3.11)$$

Where: r_d = coefficient of determination

S= standard error of simulated flows

$$S = \sqrt{\frac{\sum (y_f - y_o)^2}{n-1}}$$

σ_y = Standard deviation of observed flows

y_f = Measured value

y_o = Mean value

The extent of correlation between variables can usually be qualitatively ascertained by plotting scatter diagrams.

3.2.6.4 Nash-Sutcliffe Coefficient of Efficiency (E)

This is the most commonly used objective function for assessing the goodness of fit after model calibration or simulation. It measures how well the simulated and observed flows correspond and its value ranges from zero to unity with the highest value of E indicating a perfect fit. It is given by:

$$E = 1 - \frac{\sum (q_i - \hat{q}_i)^2}{\sum (q_i - \bar{q})^2} \qquad (3.12)$$

Where: $\bar{q}$ is the mean of the Observed stream flow

q_i is the Observed stream flow

$\hat{q}_i$ is the simulated stream flow.

The coefficient can be used to compare:

1. Model performance between catchments
2. Model performance between different periods of study
3. Flow predictions between different models

3.2.6.5 Ratio of Total Simulated Flow to Total Observed Flow

.This is given by:

$$R = \frac{\sum_{i}^{n} Q_{i(sim)}}{\sum_{i}^{n} Q_{i(obs)}} \qquad (3.13)$$

Where $Q_{i(obs)}$ and $Q_{i(sim)}$ are respectively the observed and simulated stream flows. For a good model fit the value of R should have positive value and be as close to unity as possible. Visual inspection of the joint plots of observed, calibrated and estimated stream flows helped to subjectively judge the ability of the model to simulate inter-annual and seasonal variability as well as extreme conditions.

CHAPTER FOUR

RESULTS AND DISCUSSION

4.1 Delineation of Study Catchments

The study area was delineated into 8 sub-catchments using the 90m by 90m SRTM DEM and the established gauging stations as outlet points (figure 4.1). Using the available daily stream flow, rainfall and monthly temperature data for six of these catchments, the IHACRES model was calibrated in order to obtain the dynamic response characteristics (DRCs) of these catchments. Using topographic, land use and soil indices, correlation analysis, stepwise regression and multiple regressions were performed in order to establish a relationship between these dynamic response characteristics and physical catchment descriptors of these catchments. These relationships were then validated and an attempt made to predict daily stream flows in gauged test catchments which were assumed ungauged for the purpose.

4.2 Derivation of Dynamic Response Characteristics

The models calibrated over the period (1974-75) were found to characterise well the catchments of the Upper Tana basin in that they performed well over the period immediately before the calibration (1970-73). This is shown by the good simulation results obtained when the models were used to simulate stream flows over the entire period of record (1970-75).

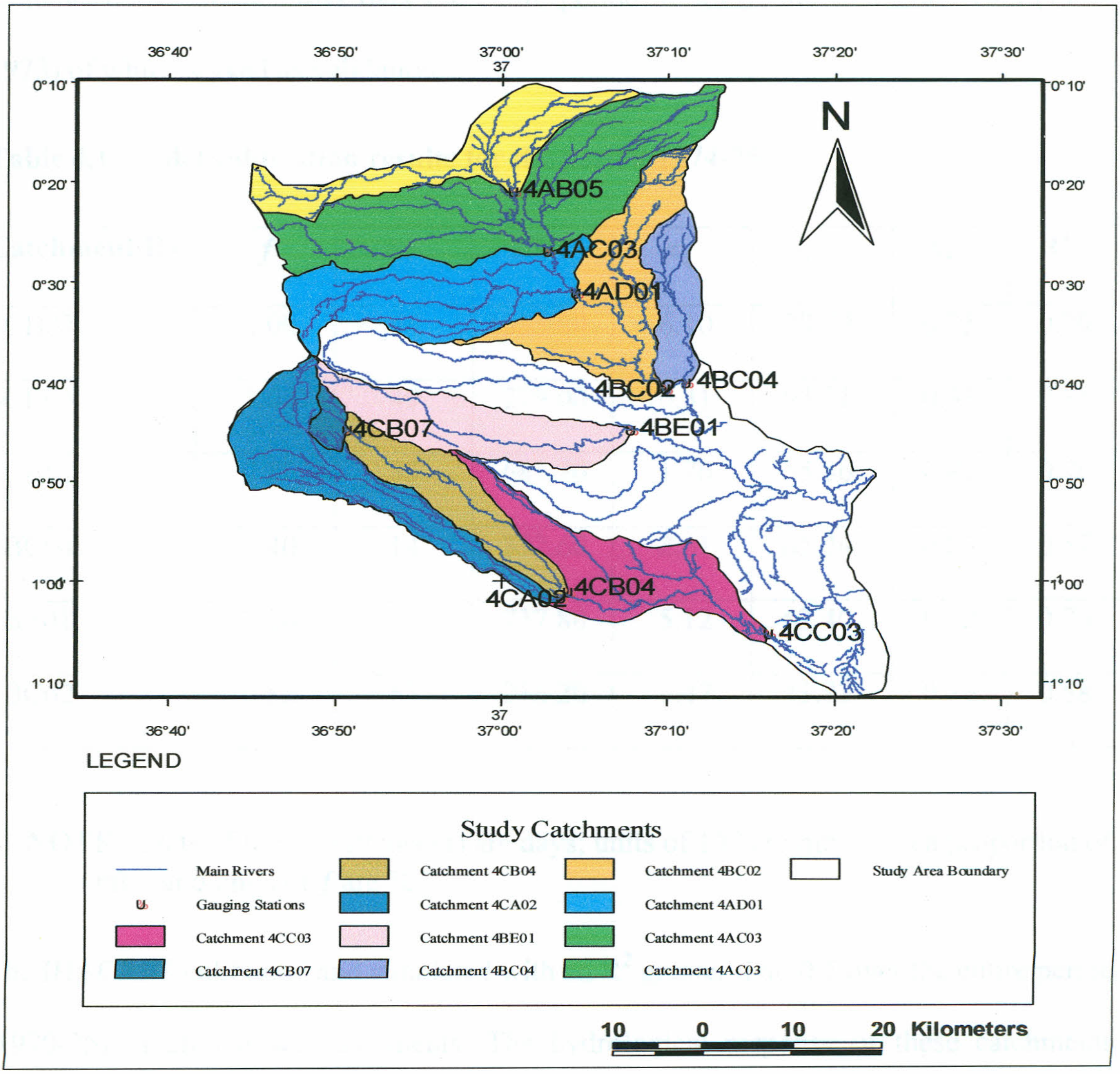


Figure 4.1 Delineated study catchments

The mean calibration R^2 obtained was 0.72 while the mean simulation R^2 was 0.64. Therefore model parameter set calibrated for the period 1974-1975, was considered as the dynamic response characteristics (DRCs) of these catchments. All the models were therefore calibrated over the two year period (1974-1975) and the optimum combination of loss characteristics chosen using the objective guidelines of maximising R^2 and minimizing ARPE and BIAS. The calibrated model for each catchment was then used to simulate daily stream flows over the entire period of study, the first four years (1970-1973) of which served as validation.

Table 4.1 Model calibration results for the period 1974-75

Catchment ID	f	τw	**1/C**	τ_q	τ_s	v_s	R^2
4CB07	4.00	64	1535.00	6.10	78.79	0.35	0.79
4CB04	1.00	49	279.00	10.31	93.51	0.33	0.85
4CA02	0.80	33	681.20	1.76	55.36	0.82	0.76
4BC04	0.40	14	332.60	2.61	65.24	0.85	0.57
4BE01	2.20	59	737.80	5.12	62.40	0.62	0.76
4BC02	0.80	63	236.20	1.47	43.83	0.76	0.58

NOTE: Units of time constants (τ) are days; units of 1/C are mm; v_s is a proportion of unity; and units of f are $°C^{-1}$

The IHACRES calibrated and simulated with an R^2 greater than 0.5 over the entire period (1970-75) in all the six catchments. The hydrological response of these catchments appear to be well represented, and adopting an acceptance threshold of $R^2>0.5$ no catchment was excluded from further analysis. Deviations of simulated flows from observed flows were assumed to occur principally due to the recorded rainfall not being representative of basin rainfall. The calibration and simulation results obtained are shown in tables 4.1 and 4.2. These results indicate that the model has been adequately calibrated. Figures 4.2 and 4.3 show sample calibration and simulation results for Maragwa catchment at station 4BE01 for the periods January 1974 to December 1975 and January

1970 to December 1975 respectively. The R^2 values obtained were 0.76 and 0.77 respectively.

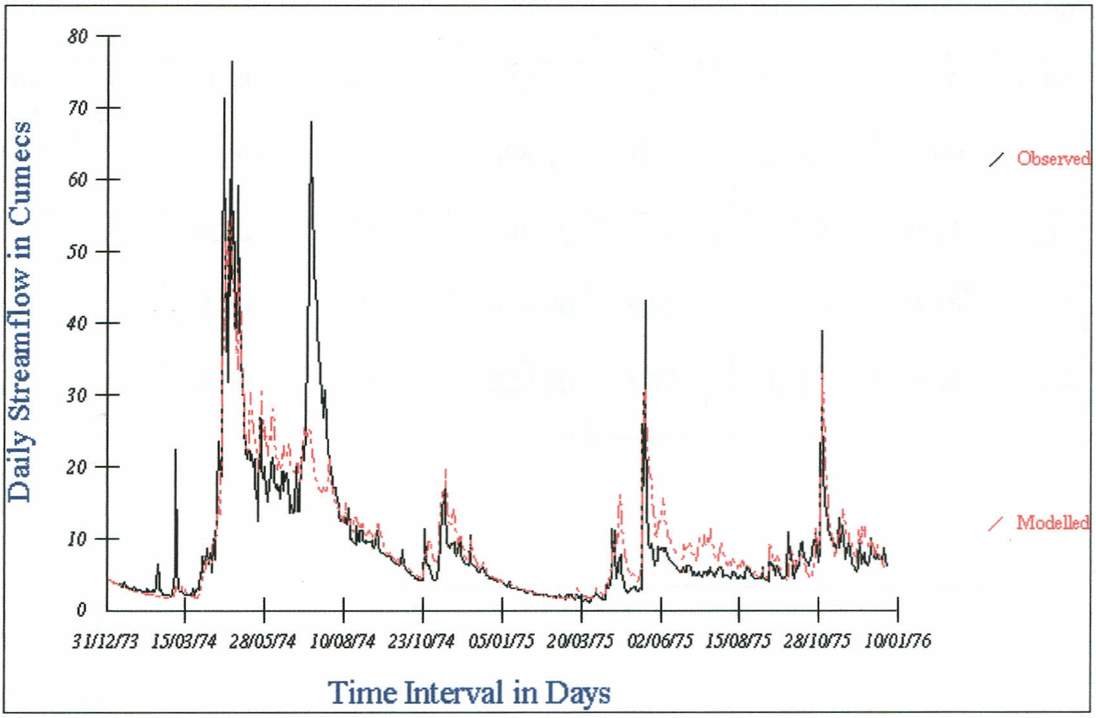


Figure 4.2 Observed and calibrated stream flow for catchment 4BE01 (1974-75)

The results indicate that the model has been well validated and the six IHACRES model parameters may be adopted as the dynamic response characteristics (DRCs) of the respective catchments and used for further analysis.

Table 4.2 Model simulation results for the period 1970-75

Catchment ID	f	τw	1/C	τ_q	τ_s	v_s	R^2
4CB07	4.00	64	1836.8	0.17	8.77	1.02	0.66
4CB04	1.00	49	332.1	2.35	14.98	1.10	0.67
4CA02	0.80	33	696.1	0.32	47.80	0.99	0.55
4BC04	0.40	14	346.90	2.71	57.06	0.83	0.55
4BE01	2.20	59	673.70	8.51	18.14	0.35	0.77
4BC02	0.80	63	229.80	5.87	121.2 3	0.58	0.64

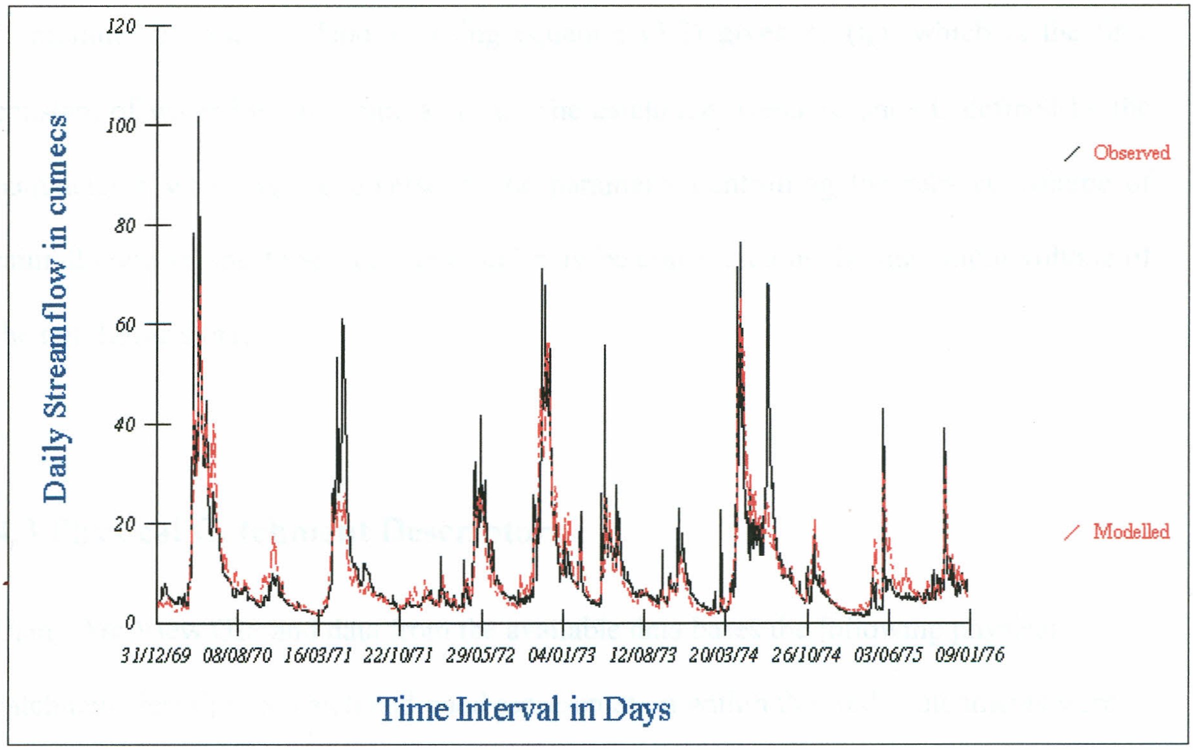


Figure 4.3 Observed and simulated stream flow for catchment 4BE01 (1970-75)

4.2.1 Catchment Dynamic Response Characteristics

Parameters τ_q and τ_s control the rate of stream flow recession from a catchment where τ_q is the time constant governing the rate of recession of water from the quick store while τ_s is the time constant governing the rate of recession of water from the slow flow store. The relative amounts of quick flow and slow flow are defined by v_s, which is the relative volume of water passing through the slow flow as opposed to the quick store ($v_q = 1 - v_s$).

Parameter f controls the effect of a unit increase in temperature on the rate of catchment water loss and thus can be considered as characterizing the difference in evapotranspiration between seasons. The rate of catchment water loss is also controlled by parameter τ_w, which is the time constant of water loss at the reference temperature. Combining parameters f and τ_w using equation (3.3) gives τ_w (t_k), which is the time constant of water loss at temperature t_k. The catchment water balance is defined by the parameter c which is the inverse of the parameter controlling the relative volume of rainfall contributing to stream flow, and may be considered as the maximum volume of the non-linear store.

4.3 Physical Catchment Descriptors

Using Arc View GIS and data from the available data bases the following physical catchment descriptors which had good representation within the study catchments were extracted and used in the derivation of DRC-PCD relationships.

Table 4.3 Key physical catchment descriptors used in the study

PCD & Category	UNIT	4CB07	4CB04	4CA02	4BC02	4BC04	4BE01
Area (T)	Km^2	49	308	493	555	210	419
D_Dens (T)	km/km^2	279	227	205	229	151	219
S_Gradient (T)	ratio	0.03	0.09	0.03	0.04	0.03	0.04
Elevation (T)	m	2678	1901	2320	1729	1555	1919
Slope (T)	%	10.14	6.86	6.75	7.36	4.12	9.77
AG1 (L_C)	ratio	0.00	0.04	0.08	0.16	0.30	0.12
AG3 (L_C)	ratio	0.00	0.72	0.28	0.56	0.53	0.56
AG3B (L_C)	ratio	0.00	0.08	0.05	0.09	0.00	0.09
FR2 (L_C)	ratio	0.00	0.00	0.12	0.04	0.00	0.15
RL2 (L_C)	ratio	0.00	0.00	0.03	0.00	0.00	0.18
R1 (S)	ratio	0.00	0.32	0.13	0.05	0.30	0.27
R2 (S)	ratio	0.00	0.28	0.10	0.58	0.36	0.39
R3 (S)	ratio	0.00	0.12	0.05	0.05	0.24	0.06
M2 (S)	ratio	1.00	0.28	0.28	0.14	0.06	0.30

Key: (T) Topography based PCDs, (L_C) Land cover based PCDs, (S) soil based PCDS

4.4 DRC-PCD Relationships

Preliminary analysis involved investigations into correlation within DRCs and PCDs and between the DRCs and PCDs (tables 4.4, 4.7 and 4.8).

Table 4.4 Correlation matrix for calibrated DRCs

	f ($°C^{-1}$)	τ_w (days)	**1/C** (mm)	τ_q (days)	τ_s (days)	ν_s (ratio)
f	1					
τ_w	0.63	1				
1/C	**0.93****	0.39	1			
τ_q	0.33	0.29	0.13	1		
τ_s	0.33	0.03	0.23	**0.93****	1	
ν_s	-0.65	-0.58	-0.45	**-0.90***	**-0.83***	1

*Correlation significant at 0.05 level ** Correlation significant at 0.01 level.

Table 4.5 Summary of univariate statistics for IHACRES model parameters

DRC	N	Mean	Standard error	Standard Deviation	Minimum	Maximum	Skewness
f	6	1.53	0.55	1.35	0.40	4.00	1.56
τ_w	6	47.00	8.12	19.89	14.00	64.00	-1.07
I/C	6	633.63	199.99	489.88	236.20	1535.00	1.53
τ_q	6	4.56	1.38	3.37	1.47	10.31	1.05
τ_s	6	66.52	7.16	17.53	43.83	93.51	0.46
$\mathbf{v_s}$	6	0.62	0.09	0.23	0.33	0.85	-0.54

The results in table 4.5 show that most DRCs are normally distributed although τ_w and v_s are both negatively skewed but the degree of skew ness appears to be insignificant. The distribution of time constant τ_q is close to normal while that of τ_s is less than one. It is only parameters f and **1/C** that are positively skewed but the magnitude of the skew ness is not significant. To ease the interpretation of dependencies, no attempt was made to normalize these distributions which are important in the relationships between DRCs and PCDs. Within the DRCs the correlation was generally low with f and τ_q having correlations significant at the 0.01% level with only **1/C** and $\tau_{s.}$ In addition, both τ_q and τ_s had negative correlations that were significant at the 0.05%level with $\mathbf{v_{s.}}$ This low correlation among IHACRES model parameters was an advantage in that strong correlation among IHACRES model parameters increases the uncertainty of the calibrated values of individual parameters.

Table 4.6 Summary of univariate statistics for physical catchment descriptors

PCD	**N**	**Mean**	**Standard error**	**Standard deviation**	**Minimum**	**Maximum**	**Skewness**
Area	6	339	77.13	188.93	49	555	-0.55
D_ Density	6	219	16.92	41.45	151	279	-0.38
S_Gradient	6	0.043	0.01	0.21	0.026	0.086	2.03
Elevation	6	2017	168	412	1555	2678	0.82
Slope	6	7.50	0.90	2.21	4.12	10.14	-0.25
AG1	6	0.11	0.04	0.10	0.00	0.30	1.04
AG3	6	0.44	0.10	0.26	0.00	0.72	-1.12
AG3B	6	0.05	0.17	0.04	0.00	0.09	-0.54
FR2	6	0.05	0.03	0.07	0.00	0.15	0.85
RL2	6	0.04	0.03	0.07	0.00	0.18	2.31
R1	6	0.18	0.06	0.14	0.00	0.32	-0.29
R_2	6	0.29	0.09	0.21	0.00	0.58	-0.07
R_3	6	0.09	0.04	0.08	0.00	0.24	1.44
M_2	6	0.34	0.14	0.34	0.06	1.00	1.99

The results in table 4.6 show that most variables are normally distributed although some like M_2, RL_2 and S_Gradient are slightly skewed positively reflecting the relatively small coverages of these physical catchment attributes within the study catchments compared to the other attributes.

The PCDs that correlated well with the model parameters (DRCs) are considered to be major drivers of catchment hydrological response. Those PCDs that did not have significant correlations with the DRCs, meant that their hydrologic effects if any, were masked by the more powerful drivers of catchment hydrological response. In this respect parameter f was strongly correlated with stream gradient (significant at the 0.05% level) and soil type M_2 (significant at the 0.01%level) which is deep well draining Andosols. Parameter τ_w was highly correlated with both drainage density and stream gradient (both significant at the 0.05% level), while parameter 1/C was highly correlated with soil type M_2 (significant at the 0.01% level).

Within the Upper Tana basin the major drivers of catchment hydrologic response were therefore identified as stream gradient, drainage density and soil type M_2 which is a predominant soil type by virtue of its widespread distribution in all the study catchments.

Table 4.7 Correlation matrix for physical catchment descriptors

	Area	D_dens	S_Grad	Elevat.	Slope	AG1	AG3	AG3B	FR2	RL2	R1	R2	R3	M2
Area	1													
D_dens	-0.22	1												
S_Grad	0.10	0.00	1											
Elev.	-0.39	0.74	-0.29	1										
Slope	-0.13	**0.86***	-0.12	0.65	1									
AG1	0.17	**-0.87***	-0.26	-0.81	-0.72	1								
AG3	0.48	-0.50	0.70	**-0.86***	-0.44	0.42	1							
AG3B	0.79	0.12	0.50	-0.33	0.22	-0.17	0.64	1						
FR2	0.62	-0.07	-0.16	0.07	0.33	-0.06	0.04	0.50	1					
RL2	0.28	-0.02	0.00	-0.06	0.48	-0.01	0.18	0.45	**0.82***	1				
R1	0.03	-0.65	0.64	-0.63	-0.47	0.40	0.75	0.19	0.06	0.30	1			
R2	0.59	-0.38	0.17	**-0.87***	-0.29	0.61	0.77	0.59	0.08	0.18	0.28	1		
R3	-0.14	**-0.87***	0.22	-0.75	**-0.85***	0.79	0.55	-0.27	-0.35	-0.19	0.74	0.33	1	
M2	-0.66	**0.82***	-0.26	0.89	0.73	-0.73	-0.81	-0.43	-0.20	-0.08	-0.59	-0.74	-0.65	1

* Correlation significant at the **0.05** level.

Table 4.8 Correlation matrix for calibrated DRCs and PCDs

	f	τ_w	1/C	τ_q	τ_s	ν_s
Area	-0.60	0.08	-0.61	-0.45	-0.68	0.54
D_Dens	0.80	**0.88***	0.66	0.40	0.27	-0.74
Mean Elev.	-0.28	0.06	-0.48	0.78	0.64	-0.51
Mean Slop.	0.77	0.40	0.89	0.18	0.25	-0.48
S_Gradient	**0.86***	**0.87***	0.72	0.29	0.11	-0.57
AG_1	-0.64	-0.66	-0.58	-0.56	-0.47	0.77
AG_3	-0.70	-0.17	-0.89	0.21	0.03	0.14
AG3B	-0.30	0.47	-0.50	0.13	-0.22	-0.02
FR_2	-0.04	0.11	0.03	-0.33	-0.49	0.37
RL_2	0.20	0.24	0.11	0.01	-0.17	0.07
R_1	-0.50	-0.52	-0.56	0.38	0.38	0.05
R_2	-0.56	0.06	-0.77	-0.27	-0.48	0.41
R_3	-0.65	-0.82	-0.61	-0.03	0.13	0.37
M_2	**0.94****	0.52	**0.94****	0.36	0.42	-0.68

*Correlation significant at the **0.05** level

Correlation significant at the **0.01 level.

4.5 Stepwise and Multiple Regressions

A stepwise regression was carried out to identify those PCDs that had the strongest statistical relationship with the DRCs. Using a forward approach the best regression equation was built one variable at a time by adding at each time step the variable that explained the largest amount of unexplained variance. For each DRC the variables suggested by correlation analysis, and stepwise regression were combined in a multiple regression equation where each DRC was expressed in terms of PCDs. These multiple regression equations enabled the estimation of DRCs from PCDs.

Table 4.9 DRC-PCD relationships for the estimation of DRCs from PCDs

DRCs	**Optimized multiple regression equations of DRCs in terms of PCDs**	**$^aR^2$**
f	$0.013+5.233RL_2+3.895M_2$	0.99
τ_w	$-76.702+0.505Dens+40.946R_2+50.884RL_2$	1.00
1/C	$162.052+1373.537M_2$	0.88
τ_q	$-8.181+178.720Gradient+9.810M_2+14.201AG_1$	0.99
τ_s	$65.647-0.056Area+596.878S\_Gradient-20.363R_2$	0.99
v_s	$0.426+1.680AG_1$	0.59

$^aR^2$ is a multiple regression coefficient.

The resulting multiple regression equations were then used to re-calculate the model parameters for the calibration catchments and to estimate model parameters for the validation catchments (tables 4.10 and 4.11).

Table 4.10 Calibrated and estimated DRCs for the calibration catchments

Catchment		f	τ_w	1/C	τ_q	τ_s	v_s
4CB07	Calibrated	4.00	64.00	1535.00	6.10	78.79	0.35
	Estimated	3.91	64.12	1535.60	6.20	78.18	0.43
4CB04	Calibrated	1.00	49.00	279.00	10.31	93.51	0.33
	Estimated	1.10	49.53	546.64	10.45	93.85	0.49
4CA02	Calibrated	0.80	33.00	681.20	1.76	55.36	0.82
	Estimated	1.26	32.60	546.64	1.74	56.18	0.56
4BC02	Calibrated	0.80	63.00	236.20	1.47	43.83	0.76
	Estimated	0.56	62.77	354.35	1.84	44.07	0.69
4BC04	Calibrated	0.40	14.00	332.60	2.61	65.24	0.85
	Estimated	0.25	14.19	244.46	2.66	66.56	0.93
4BE01	Calibrated	2.20	59.00	737.80	5.12	62.40	0.62
	Estimated	2.12	59.13	574.11	4.47	60.98	0.63

Table 4.11 Calibrated and estimated DRCs for the validation catchments

Catchment		f	τ_w	1/C	τ_q	τ_s	v_s
4AD01	Calibrated	1.00	71.00	868.30	6.89	50.83	0.53
	Estimated	1.25	35.43	560.38	4.80	47.76	0.78
4AC03	Calibrated	1.00	147.00	968.90	0.63	17.99	0.83
	Estimated	0.88	102.00	450.49	7.87	34.95	0.81

It can be seen from table 4.11 that the estimated values of the DRCs for the validation catchments did not vary greatly from the calibrated values and hence the resultant stream flow predictions for these catchments although varying were fairly accurate. Generally, the stream flows estimated using the estimated parameter sets fitted well with the observed stream flow values for most catchments, although the fits were not as good as those obtained using the calibrated parameters.. The results were plotted along side the calibrated results for the two parameters f and τ_w for comparison (figures 4.4 and 4.5).

4.6 Relationships between DRCs and PCDs

All physical catchment descriptors and dynamic response characteristics were examined for cross-correlations. Relationships between catchment attributes can exist when two or more attributes essentially measure the same catchment characteristic or when some catchment attributes interact with each other in ways that are not very obvious. Interactions between specific catchment attributes and model parameter values are presented and discussed in the following sections:

4.6.1 Non- Linear Loss Model Parameters

The most influential model parameter governing the rate of catchment water loss is f which is the rate of change of catchment water loss per unit change in temperature. It is the parameter that determines differences in evapotranspiration (ET) between seasons. The parameter is strongly correlated with catchment stream gradient and soil type M_2 (table 4.8). This is probably due to the fact that all the catchments face the same general direction and therefore those that slope more steeply receive less radiation and hence have lower values of evapotranspiration. Also since these catchments are fairly small (maximum area 555km^2) catchment gradient may have an indirect link with relief and hence temperature and may be considered an adequate measure of evapotranspiration. However, for larger catchments, other catchment attributes may be required to characterize the differences in ET values between seasons. The parameter f is also affected by the vegetation present in the catchment and this can be seen from the high f value in catchments 4CB07 (100% forest) as compared to other catchments.

The other catchment response characteristic related to the rate of catchment water loss is parameter τ_w the time constant of water loss at 20°C. This parameter was found to be strongly influenced by drainage density and catchment gradient (table 4.8). This relationship is due to the fact some of the water lost by the catchment occurs as stream flow and as the drainage density increases, τ_w decreases as the water is lost from the catchment as stream flow much more quickly. Impacts of evapotranspiration on the rate of water loss occurs through parameter f and this can be seen in the strong relationship between this parameter and catchment gradient (table 4.8)

From equation (3.2) parameter c determines to what degree a unit input of rainfall contributes to catchment wetness and therefore stream flow through equation (3.4). This implies therefore that the larger the value of c, the lower the stream flow and vice versa. The value of c is dependent to a large extent on how quickly the catchment loses water as governed by parameter τ_w. This relationship occurs because the amount by which the catchment storage, S_k, must be increased so that rainfall can contribute to stream flow, equation (3.4) is regulated not only by the ratio of rainfall to stream flow, but also by the initial value of S_k itself. This implies that c will need to be larger in a catchment which dries out quickly than in a catchment which dries out slowly. It follows therefore that parameter c is related to the same catchment attributes as parameter τ_w that is catchment gradient and drainage density. In addition, this hydrologic characteristic, like parameter f is driven largely by vegetation because of the effect of vegetation on the rate of evapotranspiration. Vegetation cover, especially, with cultivation, enhances storage capacity of catchments through infiltration resulting in substantial contribution to base flow. Catchment area has a strong relationship with parameter c as large catchments tend to have more storage capacities than small catchments (Post and Jakeman, 1996).

4.6.2 Linear Unit Hydrograph Model Parameters

At the onset, it was expected that only two components of flow recession could be identified from rainfall and stream flow data alone, (Jakeman and Hornberger, 1993). This was the case in this study and the two components were represented by the time constants of quick and slow flow, τ_q and τ_s, as well as the relative volumetric throughput v_s. The parameter τ_q was highly correlated with τ_s (0.93 at 0.01 level) and with v_s (0.90 at 0.05 level). Parameter τ_s was found to be inversely related to the volumetric throughput v_s (-0.83 at 0.05 level).

The value of time constant τ_q was found to be related to drainage density and to elevation (table 4.8). This is rather unexpected as in catchments with greater drainage densities, water will normally find its way to the stream much quicker and thus be expelled from the catchment more quickly and hence a low τ_q value. Instead, quick flow tends to recede very slowly from most of the catchments except 4CA02 and 4BC02 where it tends to recede quickly as shown by the low τ_q values for these two catchments (table 4.12). This is despite the fact that all the catchments have high drainage densities except 4BC04 which has a comparatively lower drainage density. This means that the reason for the low recession of quick flow in these catchments has to do with a combination of catchment attributes other than drainage density. It could also be due to the impact of some catchment feature or attribute that was not considered. Firstly, all the catchments are reasonably large (smallest 49km^2) and since there is a relationship between τ_q and catchment area, the time taken for in-stream travel (governed primarily by area) seems more important than how long the water takes to reach the stream (governed partially by drainage density). In addition, all the catchments are long (defined by the maximum distance of water flow to the outlet point in the catchment) and hence the stream flow takes long to reach the outlet point. Finally, parameter τ_q is inversely related to most land cover attributes especially AG$_1$ (rain fed herbaceous crop) which is predominant in all catchments except 4CB07 which is entirely covered by forest. The effect of catchment vegetation on the movement of water is to delay its movement and increase the time it requires to reach the outlet point. This increases the quick time recession constant even in the presence of dense stream network.

The time constant governing slow flow (τ_s) shows a strong positive correlation with τ_q implying that as the quick flow constant increases the water gets sufficient time to move to the ground and become sub-surface flow which moves at almost the same pace as the quick flow component. It was expected that (τ_s) will show an inverse relationship with catchment slope (Post and Jakeman, 1996) since steeply sloping catchments of similar soil types and depth will normally have a shorter τ_s as subsurface water will drain from them much more quickly. However, this was not the case and there was a very weak positive relationship. This is probably because all the study catchments generally have very steep gradients. The presence of vegetation cover delays surface runoff, enhances the infiltration process and has the overall effect of increasing base flow.

The relationship between the volume of slow flow to total flow (v_s) and physical catchment descriptors indicated that the volumetric throughput of slow flow of these catchments does not vary greatly (v_s ranged from 0.62 to 0.85) for all catchments except 4CB07 (0.35) and 4CB04 (0.33). The close similarities between the v_s values is probably due to the close similarities of these catchments in terms of soils and vegetation cover. The predominant land cover type in Catchment 4CB04 is AG_3 (rain fed shrub crop) while the entire catchment 4CB07 is covered by forest. In terms of soil coverages, catchment 4CB07 is entirely covered by soil type M_2 while catchment 4CB04 has more widely distributed soil types (R_1, R_2, R_3 and M_2). It appears from the results of correlation analysis (table 4.8) that v_s is dependent on the type and depth of soil as well as on vegetation cover present.

Table 4.12 Linear UH parameters for the calibration catchments

Catchment Identity	**Area (km^2)**	**D_Dens (Km/km^2)**	**Elevation (m)**	**Mean slope (%)**	**τ_q (days)**	**τ_s (days)**	**v_s (ratio)**
4CB07	49	278.86	2677.78	10.14	6.1	78.79	0.35
4CB04	308	227.26	1900.49	6.86	10.31	93.51	0.33
4CA02	493	205.31	2320.09	6.75	1.76	55.36	0.82
4BC02	555	229.15	1728.87	7.36	1.47	43.83	0.76
4BC04	210	150.79	1554.84	4.12	2.61	65.24	0.85
4BE01	419	219.22	1919.34	9.77	5.12	62.40	0.62

4.7 Validation of DRC-PCD Relationships

The validation of DRC-PCD relationships was successful in that the estimated DRCs compared well with the calibrated values (table 4.11). The degree of deviation from the calibrated values in most of the relationships was relatively small and this is despite the limitations of the modelling and calibration approaches adopted, and the limited information that was available to evaluate some of the PCDs. Figures 4.4 and 4.5 show the plots of the comparison between calibrated and estimated parameter values for the temperature modulation factor (*f*) and the catchment drying time constant (τ_w). Plots of the other model parameters are shown in appendix 2.1. The re-calculated dynamic response characteristics for the six calibration catchments closely match the calibrated values as can be seen statistically (table 4.10) and from the graphical plots of the same (figures 4.4, 4.5 and appendices 2.1-2.4)

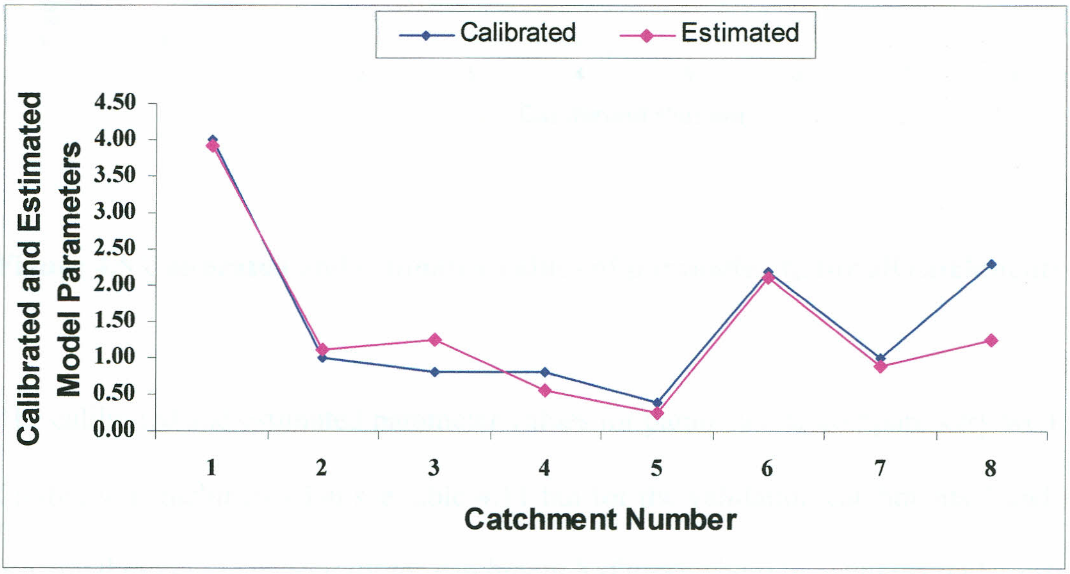


Figure 4.4 Calibrated and estimated values of parameter *f* for all catchments

The calibrated and estimated values for parameter *f* compare very well for all the calibration catchments 1-6, table 4.10 as well as one of the validation catchments 7, table 4.11 However, for validation catchment 8 there is a small underestimation of this parameter due probably to the low percentage value of variable RL_2 which is a major component in the regression equation used to estimate this parameter. The effect of land cover, a seasonal component in the estimation of parameter *f* is therefore significant. This is because this parameter is affected by vegetation status in the catchment.

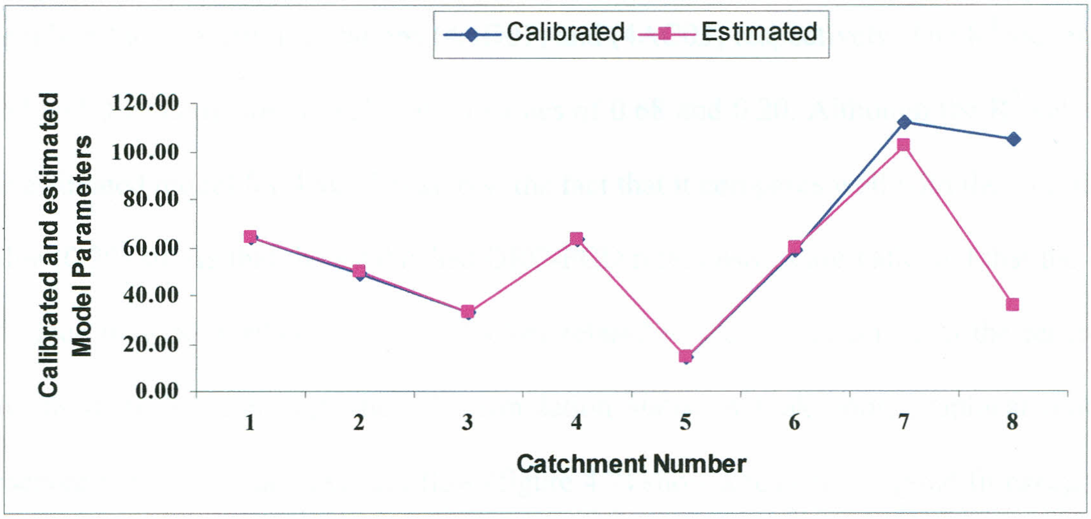


Figure 4.5 Calibrated and estimated values of parameter τ_w for all catchments

The calibrated and estimated parameter values for parameter τ_w compare very well for the Calibration catchments 1-6 see table 4.11 but for the validation catchments 7 and 8 there is a small deviation especially for catchment 8 (Gura) where the value is under estimated. This may probably be due to the high drainage density in this catchment and the low percentage of land cover type RL_2. These two are the major variables in the regression equation used to estimate this parameter. The existence of a relationship between drainage density and parameter τ_w was expected and is due to the fact that some of the water lost by the catchment occurs as stream flow and as drainage density decreases, τ_w increases as the water is being lost from the catchment much more slowly.

4.8 Estimation of Daily Stream Flows

Time series plots of observed and estimated stream flows for the validation catchments in figures 4.6 and 4.7 show that the estimated models compare well with the calibrated models at the two gauging stations (4AD01) and (4AC03) respectively. The R^2 values are 0.67 and 0.19 compared to calibration values of 0.68 and 0.20. Although the R^2 value of the estimated model for 4AC03 was low, the fact that it compares well with the calibrated value (0.20) shows that the established DRC-PCD relationships are valid and that the low R^2 value may be attributed to other factors related to the catchment or to the recorded data used. In addition, although the simulation statistics look poor, graphical plot of observed versus estimated stream flow (figure 4.7) show a reasonably good fit except for the peak flows which are consistently underestimated by very big margins. However, the determination of peak flows is usually accompanied by very low accuracy and in addition; the observed high peaks are not reflected in the corresponding rainfall records.

In an attempt to explain the hydrologic response of this catchment, correlation analysis was carried out between the recorded stream flow records and those of the neighbouring gauging station (4AD01), with which it was found to correlate well (correlation coefficient 0.65). A simple regression equation was then developed and used to fill in stream flow values at the peaks (only those above $50m^3/s$). When the model was calibrated with the new peak flow values there was little change in the R^2 value and even when these new values were applied in simulation mode the impact was negligible. The implication is that the poor simulation results may be as result of the nature of rainfall data causing model identification problems.

Considering that this catchment is extremely large compared to other catchments ($1035km^2$), and that it had the largest number of rainfall stations used to generate Thiessen polygons for the computation of areal rainfall, errors in rainfall records as well as those arising from lumping of rainfall may be contributing to this type of response. In addition, the recorded stream flow data may not be a true reflection of the gauge readings at station 4AC03. However, these tests served as a further validation of the developed relationships, in that effects were explainable in simple physical terms, using the PCDs linked to each DRC.

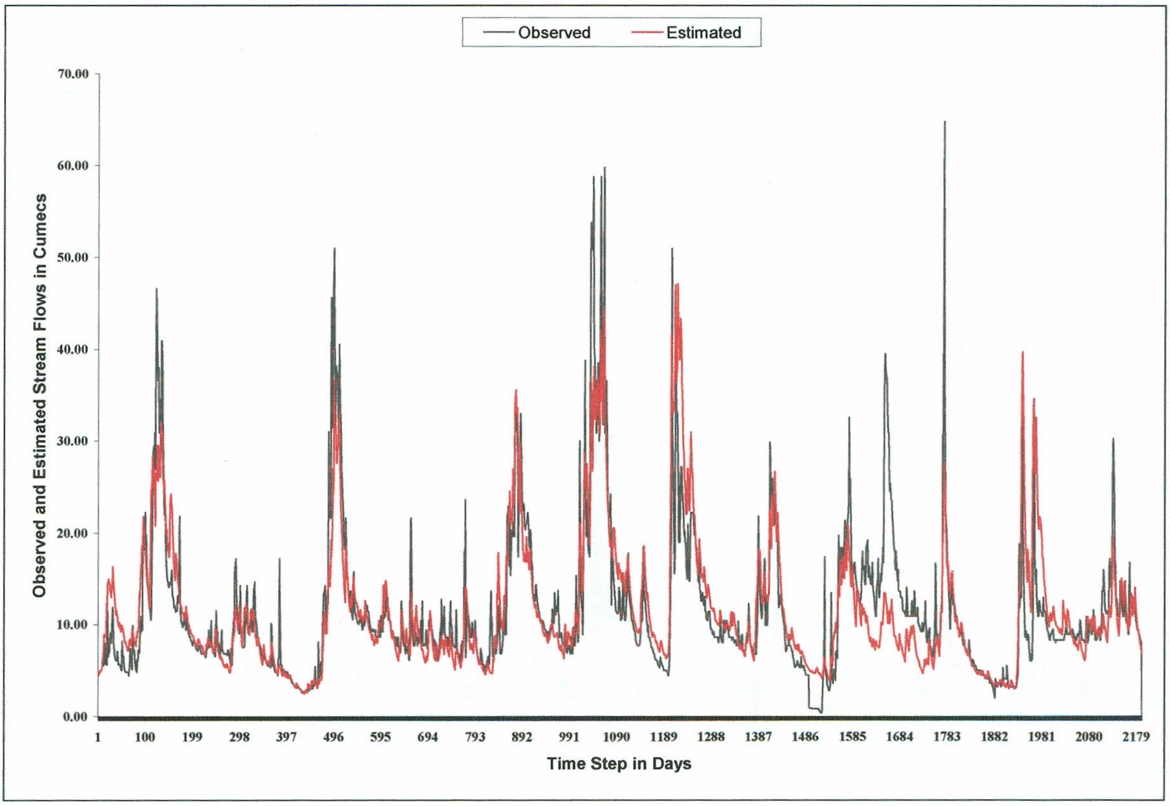


Figure 4.6 Observed and estimated stream flows for Gura river at Station 4AD01 for the period 1970-1975

The results of the predictions in catchment 4AD01 show that the predicted values compare well with the observed values and that there is an even distribution of model under and over estimates. However, some peak flows were underestimated by small margins (figure 4.6)

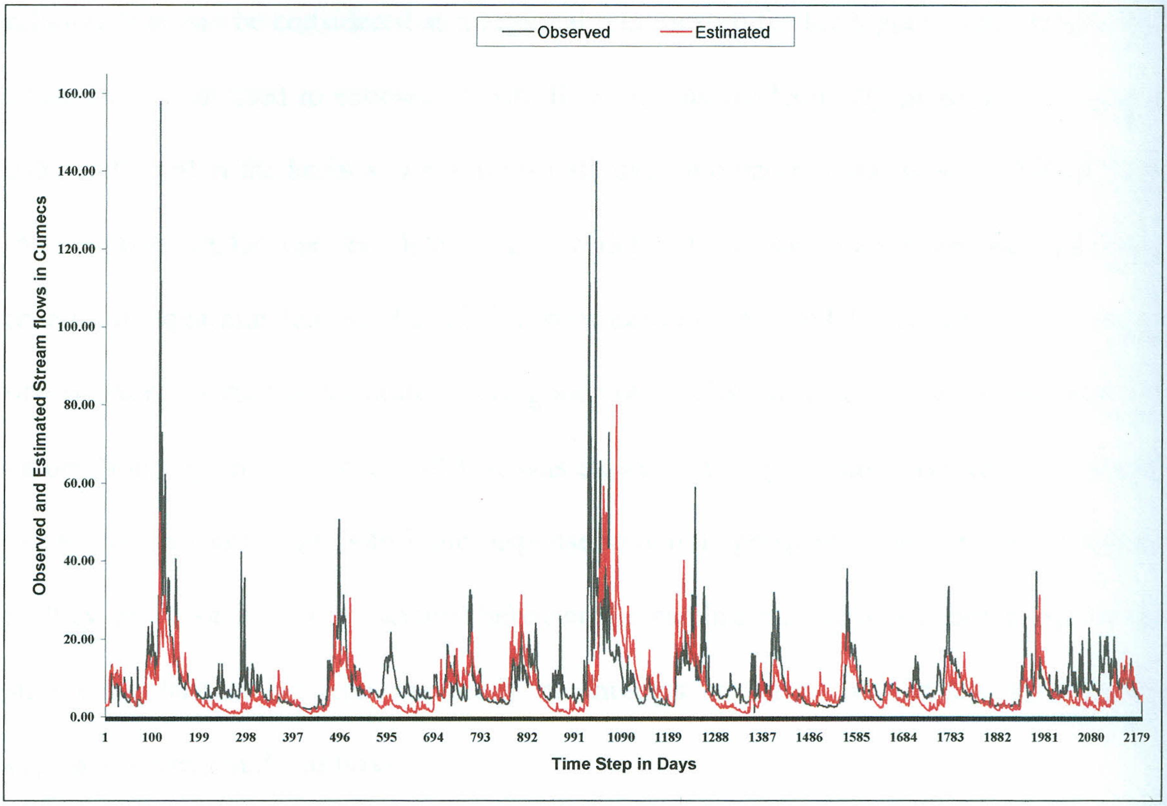


Figure 4.7 Observed and Estimated Stream flows for Sagana river at station 4AC03 for the period 1970-1975

.

4.9 Establishment of a Regional Parameter Set

Complete validation of the derived DRC-PCD relationships involved simulations of stream flows using the estimated DRCs for all the calibration catchments and a comparison with stream flow simulations using calibrated DRCs. Figure 4.8 shows a time series plot of observed and simulated daily stream flows using calibrated parameters and the regionalised parameters set estimated from the DRC-PCD relationships, for Maragua catchment (at station 4BE01). Time series plots for the other calibration catchments are shown in appendix 3.1 to 3.5. The results of stream flow estimations in both validation and calibration catchments which were successful indicate that the derived DRC-PCD relationships can be considered as a regional relationship for the Upper Tana basin in the sense that it can used to estimate stream flows in any inadequately gauged or ungauged catchment within the basin so long as rainfall and temperature data are available and that the requisite PCDs can be determined or derived. It was seen from the optimised regression equations that all the relationships defining the model parameters in terms of one or more landscape attributes were good, with a low degree of scatter and high R^2 values (ranging from 0.59 to 1.00). It was expected that physical attributes with strong influences on catchment hydrologic response may not appear in the regression equations as they may not vary significantly between the catchments. This was attributed to the strong similarities between the study catchments in terms of climate, topography, vegetation cover and soil types.

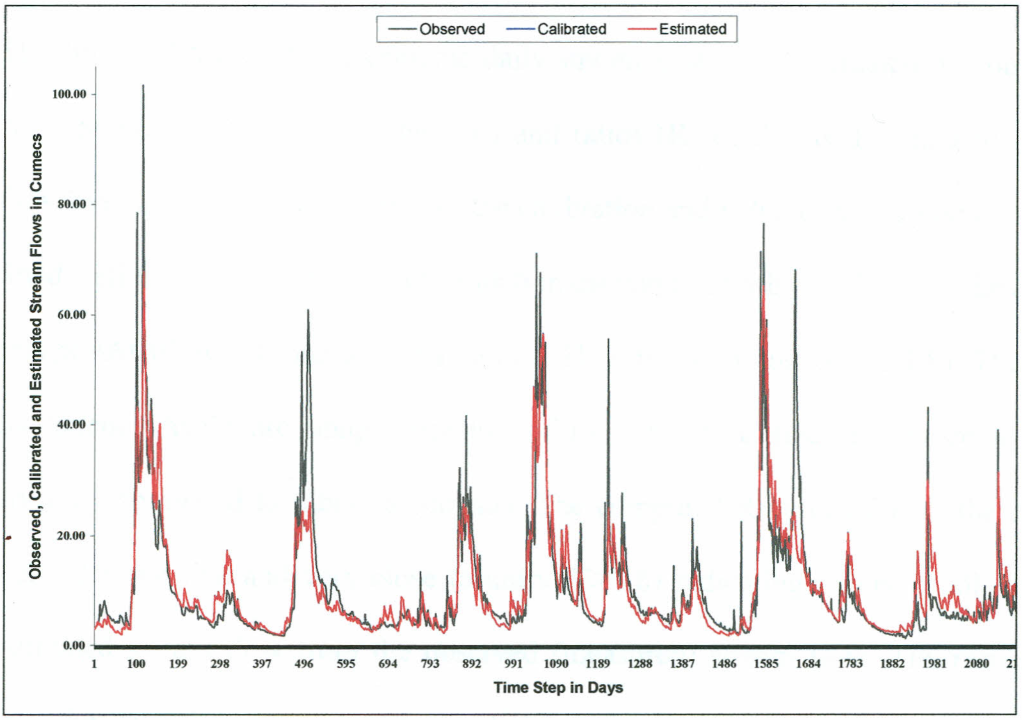


Figure 4.8 Observed, calibrated and estimated stream flows for Maragwa river at station 4BE01 for the period 1970-1975

Figures 4.6 and 4.7 show the observed and predicted stream flows for Gura (4AD01) and Sagana (4AC03) catchments. These stream flows were predicted using the estimated parameters only without calibrating the model to observed stream flow record of those catchments. The estimated stream flow for catchment 4AD01 compares very well with the observed values both visually and statistically. Although the estimated stream flow for catchment 4AC03 looks poor when looked at statistically (in terms of R^2), it compares well with the observed values except for the extreme peaks where the observed values look exaggerated. However, this pattern was also noted during calibration as discussed earlier this may be due to errors in the rainfall or stream flow records.

4.10 Model Performance

The ability of the model to estimate daily stream flows was evaluated by computing the Nash Sutcliffe (1970) efficiencies (E) and ratios (R) of the total estimated (simulated) flow and measured stream flows for the calibration and validation catchments. The model fitted well to both calibration and validation catchments with E>0.5 for all the catchments except 4AC03 which had an E value of 0.21 (table 4.13 and figure 4.9). The results of catchment 4AC03 are unique because of its very large size and hence hydrological diversity compared to other catchments. The computed R values for all the catchments were also positive and very close to unity (E>0.8) indicating a very small deviation of estimated stream flow from the observed (measured) stream flow. The results indicate that the model is capable of simulating to a reasonable degree of accuracy the hydrological response characteristics of these catchments.

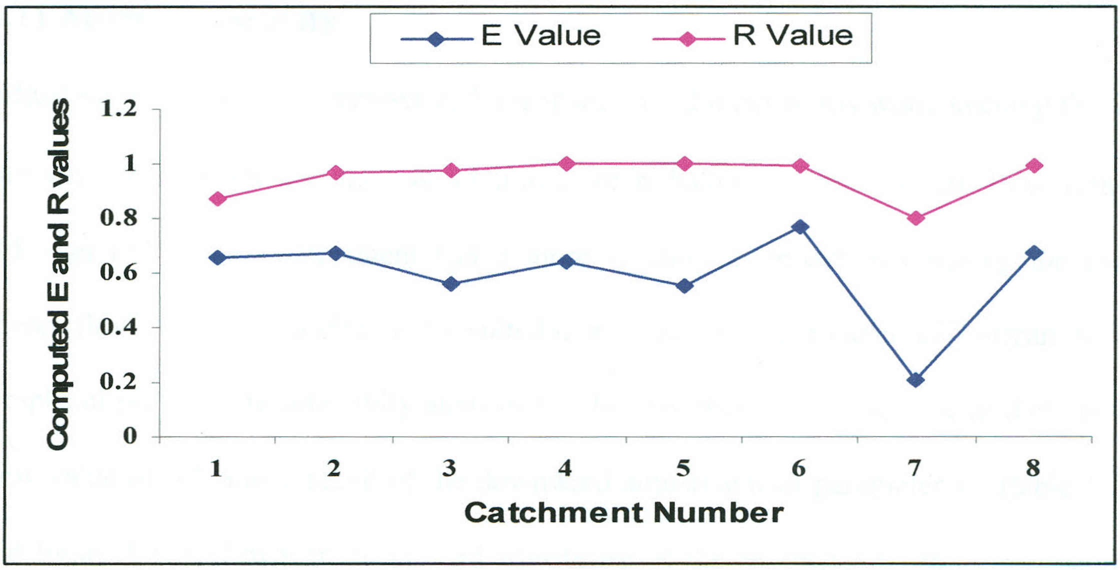


Figure 4.9 Computed E and R values for the calibration and validation catchments

Table 4.13 Efficiencies and Ratios based on estimated model parameters

Catchment ID	Efficiency (E)	Ratio (R)
*4CB07	0.66	0.87
*4CB04	0.67	0.97
*4CA02	0.56	0.98
*4BC02	0.64	1.00
*4BC04	0.55	1.00
*4BE01	0.77	0.99
**4AD01	0.67	0.99
**4AC03	0.21	0.80

* Calibration catchments **Validation catchments

4.11 Model Sensitivity

Adjusting the values of parameter τ_w both upwards and downwards while keeping that of parameter f constant had the overall impact of reducing the total stream flow values although the upward adjustment had a more significant impact in reducing the total stream flow volume. In addition, it resulted in a reduction of the mean daily stream flow. Graphical plots of the sensitivity analysis results also indicate an increase in daily peak flow value of 6.1% as a result of the downward adjustment of parameter τ_w (table 4.14 and figure 4.10). However, an upward adjustment of the parameter by the same margin (50%) reduced the daily peak flow value by a margin of 10.48%. It is worth noting however, that although this pattern was evident in all the peaks, the magnitude of the variations varied from peak to peak and tended to increase as the peak values increased. The IHACRES model therefore appears to be very sensitive to changes in parameter τ_w.

Table 4.14 Parameter adjustments and the resulting stream flow changes for Maragwa River at 4BE01 for the period 1970-1975

Parameter	**Parameter value**	**Total stream flow (m3/s)**	**Stream flow change (m^3/s)**	**Mean daily flow (m^3/s)**	**Peak flow (m^3/s)**
	29.5	22114.13	- 12.26	10.10	71.95
τ_w	59	22126.39	0	10.10	67.81
	88.5	21840.60	- 285.79	9.97	60.70
	1.1	22157.38	+30.99	10.12	65.72
f	2.2	22126.39	0	10.10	65.19
	3.3	22096.22	-30.17	10.09	64.58

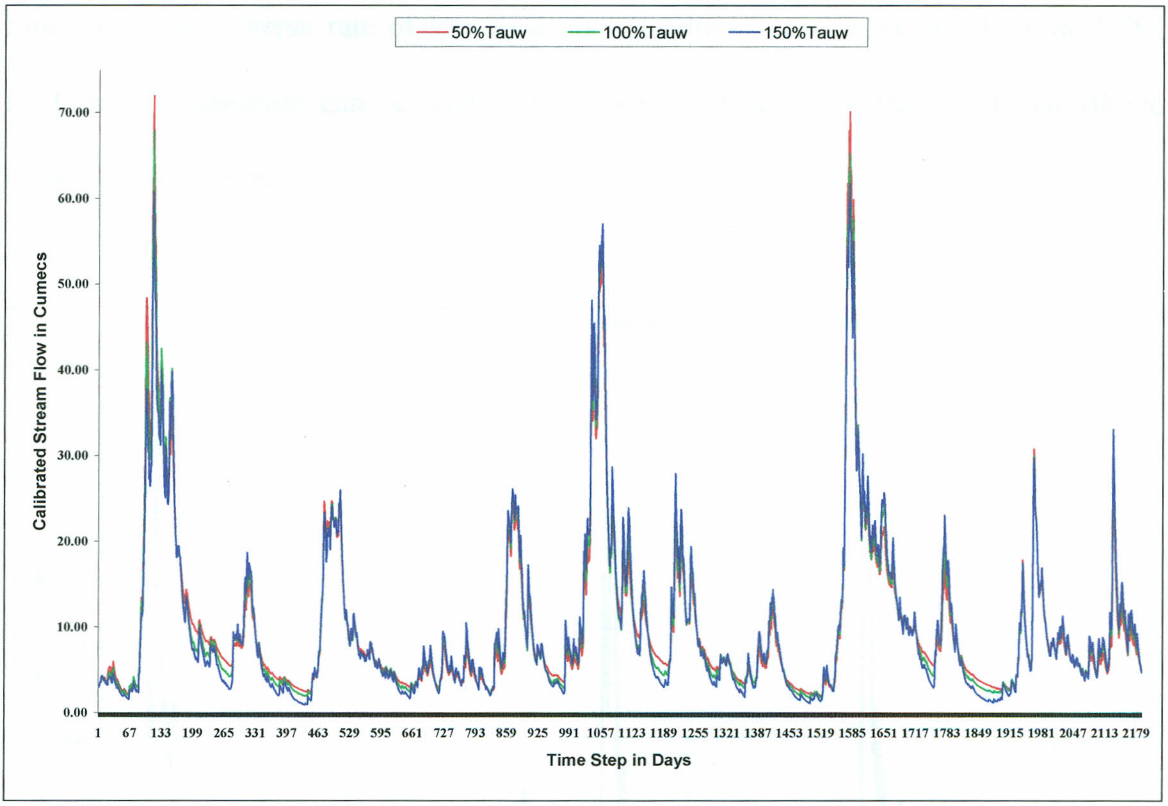


Figure 4.10 Effect on stream flow of adjusting τ_w while keeping f constant for Maragwa river at Station 4BE01 for the period 1970-1975

Keeping parameter τ_w constant and adjusting the values of parameter f both upwards and downwards had very little impact on both the total stream flow volume and the mean daily stream flow. Graphical plots of the results also show no clear distinction between the plots at 50%, 100% and 150% values of parameter f. The daily peak flow was also not affected significantly by adjustments in the values of parameter f with only a marginal decrease of 0.01% as a result of 50% increase in parameter value. The results of the sensitivity analysis show that the model IHACRES is sensitive to changes in parameter τ_w and virtually insensitive to changes in parameter f.

Since τ_w is the inverse rate of loss to evapotranspiration and to stream flow at 20°C, land/cover or land/use can be said to be a dominant driver of catchment hydrologic response in this area.

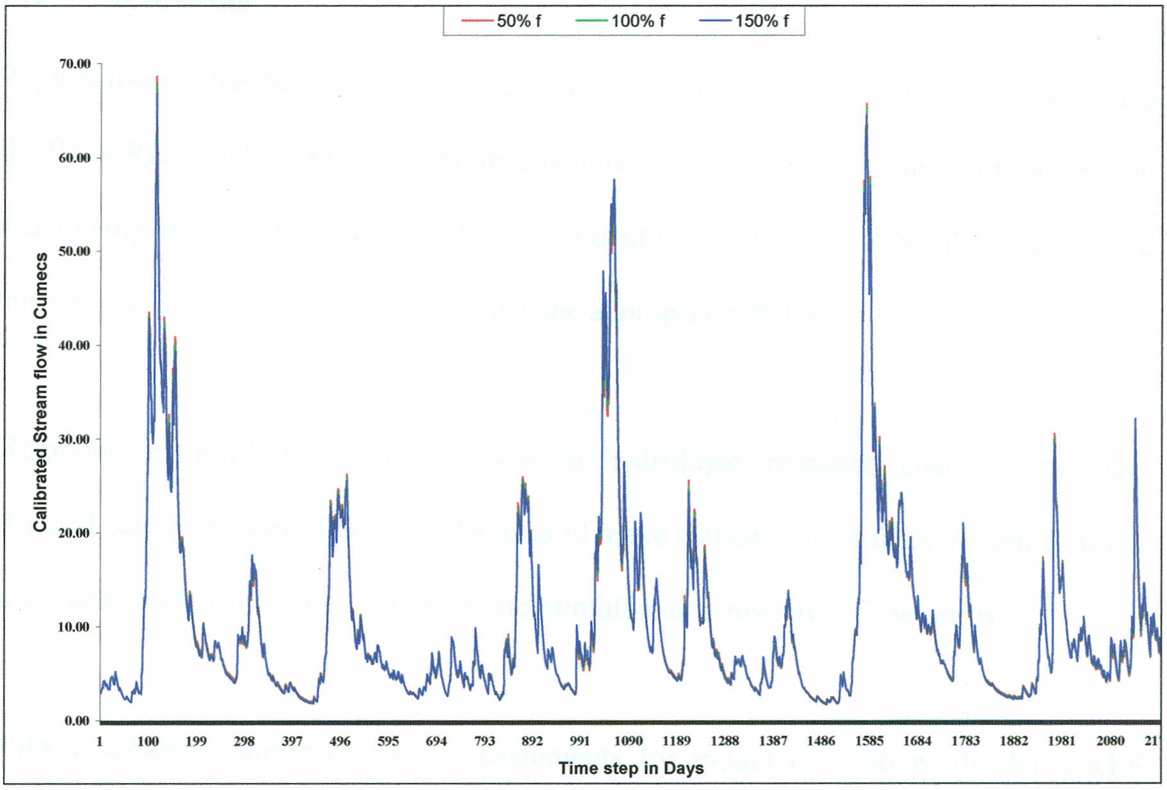


Figure 4.11 Effect on Stream flow of adjusting *f* while keeping τ_w constant for Maragwa River at 4BE01 for the period 1970-1975

CHAPTER FIVE

CONCLUSIONS AND RECOMMENDATIONS

5.1 Conclusions

Regionalisation has been successfully demonstrated for the Upper Tana River basin using the IHACRES model. However, the results must be viewed in the context of the amount and quality of data that was available to quantify the DRCs, as well as the limited information that was available to quantify the appropriate PCDs.

Regional relationships developed between hydrologic response characteristics and physical catchment descriptors can be used to make estimations of daily stream flows of ungauged catchments from an appropriate rainfall and temperature time series.

With a suitable number of gauged catchments for model calibration, the IHACRES model can be regionalised and used to simulate rainfall-runoff processes in ungauged catchments of the Upper Tana River Basin.

The results from the current study have shown that it is possible to develop relationships for transferring hydrological data from one catchment to another in a given geographic and climatic region, and that the form of these relationships may vary from one geographic region to another.

5.2 Recommendations

Based on the results of this study, the following recommendations can be made:

The methodology should be applied to generate stream flow data in ungauged catchments of the basin where water resources projects are planned in the absence of adequate data.

The derived DRC-PCD relationships can be more stringently tested by using validation catchments that explore a wider parameter space.

A larger data set that explores a wider geographical space be used in order to improve the established DRC-PCD relationships and reduce the differences between the calibrated and estimated parameter values so that the results can be applied with greater confidence.

This regionalization approach be applied in other climatic regions in order to improve understanding of the physical controls on the hydrological response of catchments in different hydro-climates.

The methodology be applied using catchments from semi-arid areas in order to understand differences in hydrologic response characteristics between humid and semi-arid catchments

Rainfall data from satellite imagery be used to carry out similar studies in order to improve the relationships between model parameters and catchment characteristics.

REFERENCES

Abbott, M.B., Bathurst, J.C., Cunge, J.A., O'Connell P.E., and Rasmussen, J. (1986): An Introduction to the European Hydrological System. Journal of Hydrology, 87: 45-59

Acreman, M. (1994): The role of artificial flooding in integrated development of river basins in Africa. In: Kirby C. and White W.R. (Eds.): Integrated river basin development. John Wiley and Sons Ltd.

Agwata J.F. (2005): Characteristics of Hydrologic drought in the Upper parts of the Tana Basin. Unpublished PhD Thesis, Kenyatta University, Nairobi Kenya.

Baker H.B. (1967): Geology of the Mt. Kenya Area. Geological Survey of Kenya Report No. 79. Government Printer Nairobi 40 pp

Bates, B.C. (1994): Regionalization of Hydrological data. A review. Regionalization and scaling of hydrological data. Report 94/5, CRC For catchment hydrology, 61 pp

Bathurst, J.C. (1986): Physically-based distributed modelling of an upland catchment using the SHE model. Journal of Hydrology 87:79-102

Beven, K. J. (1989): Changing ideas in Hydrology: The case of physically based models. Journal of Hydrology 105: 157-172

Beven, K.R., Lamb P., Quinn R. and Freer J. (1995): "TOPMODEL" Computer Models of Watershed Hydrology. Singh V.J. (Eds.): Water Resources Publications, pp 727- 668

Blackie, J.R., and Eeles C.W. (1985): Lumped catchment models. In: Anderson M.G. and Burt T.P. (Editors). Hydrological forecasting. John Wiley and Sons Ltd,

Bobotti, O.K. (1996): Tana River Management. "Proceedings of a seminar on Integrated Water Resources Management in Kenya. World Bank, Nairobi pp 122-129

Bloschl, G. and Sivapalan M. (1995): Scale issues in hydrological modelling. A review, Hydrological Processes 9: 251-290.

Bosch, J.M. and Hewlett J.D. (1982): A review of catchment experiments to determine the effects of vegetation changes on water yield and evapotranspiration. Journal of Hydrology 55: 3-23

Brazil, L.E. (1988): Multilevel calibration strategy for complex hydrologic simulation models. PhD Dissertation, Department of Civil Engineering, Colorado State University, Fort Collins.

Burn, D.H. and Boorman, D.B. (1992): Estimation of hydrological parameters at ungauged catchments, Journal of Hydrology., 143, 429-454.

Burn, D.H. and Zrinji Z. (1994): Flood frequency analysis for ungauged basins. Journal of Hydrology 153: 1-21

Chow, V.T., Maidment, D.R. and Mays, L.W. (1988): Applied Hydrology. International Edition. McGraw-Hill Company, New York.

Clarke, R.T. (1994): Statistical Modelling in Hydrology. John Wiley and Sons Ltd, pp 426

Croke, B.W., Andrews F., Jakeman A.J., Cuddy S. and Luddy A. (2005): Re-design of the IHACRES rainfall-runoff model. Engineers Australia Proceedings of the 29th Hydrology and Water Resources Symposium, 21-23 February 2005, Canberra, Australia.

Diskin, M.H. and Simon, E. (1977): A procedure for the selection of objective functions for hydrologic simulation models. Journal of hydrology 34: 129-149.

Dooge, J.C. (1959): 'A general theory of unit hydrograph; Journal of geophysical research, 64(1):241

Duan, Q., Sorooshian, S. and Gupta, V.K. (1992): Effective and efficient global optimisation for conceptual rainfall-runoff models. Water Resources Research 28 (4): 1015-1031

Eckhardt, K. and Arnold, J.G. (2001): Automatic calibration of a distributed catchment model. Journal of Hydrology 251: 103-109.

Food and Agriculture Organisation (FAO) (1997): Digital Soil Map of the world. CDROM. Rome, Italy.

Funke, R., Schumann, A.H. and Schultz G.A. (1999): Regionalization of parameters in mesoscale hydrological models. In: Diekkruger B., Kirkby M.J, and Schroder U. Eds. Proceedings of a conference held at Braunschweig, March 1997. IAHS Publication No. 254: pp 171-179.

GoK (2000): Population Census, Central Bureau of Statistics, Government of Kenya, Nairobi Kenya.

Harper, D. and Brown T. (1998): The Sustainable management of Tropical catchments, John Wiley and sons Ltd.

Hornbeck, J.W. (1973): Storm flow from Hardwood forested and cleared watersheds in New Hampshire. Water Resources Research, 9(2):346-354.

Hughes, F.M.R. (1984): A comment on the Impact of development schemes on the flood plain forests of the Tana River basin of Kenya. The Geographical Journal, 150(2): 230-244

Jaetzold R. and Schmidt H. (1983): Farm Management hand book of Kenya, Volume II Part C, East Kenya- Ministry of Agriculture/GTZ, Nairobi, Kenya.

Jakeman, A.J. Littlewood, I.G. and Whitehead P.G. (1990): Computation of instantaneous unit hydrograph and identifiable component flows with application to two small upland catchments. Journal of Hydrology 117: 275-300.

Jakeman, A.J., and Hornberger, G.M. (1993): How much complexity is warranted in a rainfall runoff model? Water resources research 29(8) 2637-2649.

Jakeman, A.J., Chen, T.H., Post, D.A., Hornberger, G.M., Littlewood, I.G. and Whitehead, P.G. (1993): Assessing uncertainties in hydrological response to climate at large scale. Macro scale modelling of the hydrosphere. Proceedings of the Yokohama Symposium, IAHS Publication N0. 214, 37-47.

JICA (Japan International Co-operation Agency) (1992): The study in the National Water Master plan. (Sectoral report B, Hydrology), Ministry of Water Development, Nairobi, Kenya

Kalders, J. (1988): Evapotranspiration in Kenya, Irrigation Section, Ministry of Water Development, Government of Kenya, Nairobi, Kenya.

Kirkby, M.J. (1999): Translating models from hillslope (1ha) to catchment (1000km^2) scales.In: Diekkruger B., Kirkby M.J., and Schroder U. (Eds). Proceedings of a conference held at Braunschweig, March 1997. IAHS Publication no. 254 pp 1-12.

Kokkonen, T.S., Jakeman A.J., Young P.C., and Koivusalo H.J. (2003): Predicting daily Flows in ungauged catchments: Model regionalization from catchment descriptors at the Coweeta Hydrologic Laboratory, North Carolina. Hydrological processes 17: 2219-2238.

Kokkonen, T.S. (2002): Rainfall-runoff modelling – comparison of modelling strategies with focus on ungauged predictions and model integration. PhD Thesis, Helsinki University of Technology, Finland.

Leavesley, G.H. (1994): Modelling the effects of climatic change on water resources - A review. Climatic change 28: 159-177

Legesse, D., Vallet-Coulomb, C. and Gasse, F. (2003): Hydrological response of a catchment to climatic and land use changes in Tropical Africa: Case study of South Central Ethiopia. Journal of Hydrology: 275: 76-85.

Linsley, R.K. and Crawford N.H. (1960): 'Computation of a synthetic stream flow record on a digital computer'. Hydrological Sciences. Bulletin, IASH, Publication 51: pp 526-538

Linsley, R.K. and Crawford, N.H. (1966): Digital simulation in hydrology, Stanford watershed model IV, Technical report 39, Dept. of Civil Eng. Stanford University, Stanford, CA.

Littlewood, G., Down, K., Parker, J.R. and Post, D.A. (1997): IHACRES catchment scale Rainfall-stream flow modelling. PC v 1.0, user Guide http://www.ceh.ac.uk/.
On 20th February, 2005.

Littlewood, G., Jakeman, A .J., Croke, B., Kokkonen, T. and Post D.A. (2002):
Unit hydrograph characterization of flow regimes leading to stream flow estimation in ungauged catchments. IAHS decade for PUB. PUB meeting in Brasilia November 2002. Jason Hunt naturalchild.org/Jason.

Mazvimavi, D. (2002): Estimation of flow characteristics of ungauged catchments, case study of Zimbabwe. PhD Thesis, Wagenin University, Netherlands

Mimikou, M. (1984): Regional relationships between basin size and runoff characteristics. Hydrological Sciences Journal 29(1): 63-73

Mimikou, M. and Gordios, J. (1989): Predicting the mean annual flood and flood Quantities for ungauged catchments in Greece. Hydrological processes 15: 2281- 2295

Moore, I.D., Grayson, R.B. and Ladson, A.R. (1991): Digital terrain modelling: a review of hydrological, geomorphologic, and biological applications. Hydrological processes 5: 3-30

Mulvaney, T.J. (1851): On the use of self-registering rain and flood gauges in making observations of the relations of rainfall and flood discharges in a given catchment.
Proc. Inst. Civil Engineers, Ireland, 4, 18-31

Muthusi, F.M. (2004): Evaluation of the USGS stream flow model for flood simulation (Case study of the Nyando Basin) Unpublished Msc. Thesis, Jomo Kenyatta University of Agriculture and Technology, Nairobi, Kenya.

Mutisya, D. and Mutiso S. (1998): Social-economic aspects of subsistence farming and soil erosion in Tropical catchments, The Tana Catchment, Kenya, In: Harper, D and Brown T. (Editors) Sustainable Management of Tropical catchments, John Wiley and Sons Ltd.

Mutua, F.M. (1993): The use of Akaike information criteria in the identification of optimum flood frequency model. Proceedings of the 1st International conference of the African Meteorological Society, Nairobi Kenya. pp 871-881

Mwakalila, S. (2003): Estimation of stream flows of ungauged catchments for river basin management. Physics and Chemistry of the Earth 28: 935-942

Nash, J.E. (1957): The form of instantaneous unit hydrograph, International Association of Science and Hydrology. Publication 1, 45(3): 114-121

Nash, J.E. and Sutcliffe J.V. (1970): River flow forecasting through conceptual models Part 1- A discussion of principles. Journal of Hydrology 10: 282-290.

Nathan, R.J. and McMahon T.A. (1990): Identification of homogenous regions for purposes of regionalization. Journal of Hydrology 121:217-238

Nathan, R.J. and McMahon T.A. (1992): Estimating low flow characteristics in ungauged catchments. Water Resources Management 6: pp 85-100.

Nyadawa, M.O. (1997): Conceptual approach to runoff prediction in Tropical basins in Kenya. Unpublished PhD. Thesis Gumna University, Japan.

Onyando, J.O. (1994): Estimation of runoff volumes and peak rates for rural catchments in Kenya. Msc. Thesis, University of Nairobi, Kenya.

Onyando, J.O. and Sharma, T.C. (1995): Simulation of direct runoff volumes and peak rates for rural catchments in Kenya, East Africa. Hydrological Sciences Journal 40(3): 367-380

Onyando, J.O. (2000): Rainfall-Runoff models for ungauged catchments in Kenya. PhD. Thesis, Ruhr University Bochum, Germany.

Onyando, J.O. (2002): Regionalization of a conceptual rainfall- runoff model for ungauged catchments. Journal of Engineering in Agriculture and Environment Vol. 2:13-24.

Otieno F.A. and Maingi S.M. (1993): Sedimentation problems of the Masinga reservoir. Proceedings of the Fourth National Workshop on Land and Water management towards sustainable land use. Nairobi, Kenya.

Oyebande, L. (2001): Water problems in Africa. How can science help? Hydrological Sciences journal, Vol. 46: No. 6 947-961

Perrin, C., Michel, C. and Andreassian, V. (2001): Does a large number of parameters enhance model performance? Comparative assessment of common catchment model structures on 429 catchments. Journal of Hydrology 242: 275-301.

Pilgrim, D.H. (1983): Some problems in transferring hydrological relationships between small and large drainage basins and between regions. Journal of hydrology 65: 49-72

Post, D.A., and Croke B.F. (2002): Predicting hydrologic response from physio-climatic attributes: An application to ungauged sub-catchments of the Burdekin River, North Queensland. In: A. Rizzoli et al (Editors). Proceedings of the 1st Biennial meeting of the International Environmental Modelling and Software Society, University of Lugano, Switzerland, Vol. 1, 334-339

Post, D.A. and Jakeman A.J. (1996): Relationships between catchment attributes and hydrological response characteristics in small Australian catchments
Hydrological processes 10: 877-892.

Post, D.A. and Jakeman A.J. (1999): Predicting the daily stream flow of ungauged catchments in S.E Australia by regionalizing the parameters of a lumped conceptual rainfall- runoff model. Ecological Modelling 123: 91-104

Refsgaard, J.C. (1996): Parameterization, Calibration and validation of Distributed Models. Journal of Hydrology, 198: 69-97

Refsgaard, J.C. and Knudsen J. (1996): Operational validation and inter comparison of different types of hydrological models. Water resources research 32: No.7, 2189-2202

Refsgaard, J.C., Seth S.M., Bathurst J.C., Erlich M., Storm, B., Jorgensen, G.H. and Chandra, S. (1992): Application of the SHE model to catchments in India, Part 1, General results, Journal of Hydrology 140: 1-23.

Reimers, W. (1990): Estimating hydrological parameters from basin characteristics for large semi-arid catchments. Proceedings of the Ljubljana Symposium, IAHS Publication N0. 191: pp 187-194

Riggs, H.C. (1990): Estimating flow characteristics at ungauged sites. Proceedings of the Ljubljana Symposium, April 1990. IAHS Publication No. 191: 159-169

Scertzer, D., Uhlenbrook, S. and Zehe, E. (2003): IAHS decade on Predictions on Ungauged Basins (PUB), 2003-2012: Shaping an Exciting future for the Hydrological Sciences. Hydrological Sciences Journal 48(6):857-880.

Schaake, J.V., Duan, Q., Koren, V.I. and Cong S. (1997): Regional parameter estimation of land surface parameterisation for GCIP large-scale area southwest, Paper presented at 13TH conference on Hydrology, American Meteorological Society, Long Beach, California, February, 1997,

Sefton, C.E., and Howarth, S.M. (1998): Relationships between dynamic response Characteristics and physical catchment descriptors of catchments in England and Wales. Journal of Hydrology 211: 1-16.

Seibert, J. (1999): Regionalisation of parameters for a conceptual rainfall-runoff model. Agriculture and Forest Meteorology, 98-99:279-293.

Servat, E. and Dezetter, A. (1993): Rainfall-Runoff modelling and water resources assessment in North –Western Ivory Coast , Tentative extension to ungauged catchments, Journal of Hydrology, 148: 231-248

Shaw, E.M. (1988): Hydrology in Practice, Second Edition. VNR International Company Ltd., London. PP 165-191

Shaw, E.M. (1996): Hydrology in practice. Third Edition, Chapman and Hall, London.

Sherman, L.K. (1932): Stream flow from rainfall by the unit graph method, engineering news record, 108: 501-505.

Sivapalan, M., Takeuchi, K., Franks, S.W., Gupta, V.K., Karambiri, H., Lakshmi, V., Liang, X., McDonnell, J.J., Mendiondo, E.M., O'Connel, P.E., Oki, T., Pomeroy, J.W., Scertzer, D., Uhlenbrook, S. and Zehe, E. (2003): IAHS decade on Predictions on Ungauged Basins (PUB), 2003-2012: Shaping an Exciting future for the Hydrological Sciences. Hydrological Sciences Journal 48(6):857-880.

Solomon, S. (1967): Relationship between precipitation, evapotranspiration and runoff in tropical-equatorial regions, Water Resources Research, 3(1):163-72

Sorooshian, S. and Gupta, V.K. (1983): Automatic calibration of conceptual Rainfall-Runoff Models: The question of parameter observability and Uniqueness. Water Resources Research, 19 (1): 251-259.

Sorooshian, S., Duan, Q. and Gupta, V.K. (1993): Calibration of Rainfall- Runoff Models. Application of global optimization to the Sacramento Soil Moisture Accounting Model. Water Resources Research, 29: 1185-1194.

Sorooshian, S., and Gupta, V.K. (1995): Model calibration. In: Singh V.P (Eds) Computer models of watershed hydrology. Chapter 2 pp 23-68

Stoneman, G.L. and Schofield N.J. (1989): Silviculture for water production in Jarrot forest of Western Australia- An evaluation 27: 273-293

TAMS, (1980): National Water Master Plan (stage 1) volume V, Ministry of Water Development, Nairobi, Kenya.

Thomas, M.K., Liniger, H.P. and Gichuki, F.N. (1993): Development of a stream flow model for rural catchments in Kenya. Proceedings of the Fourth National Workshop on Land and Water management towards sustainable land use. Nairobi, Kenya, 191-202

Vandewiele, G.L. and Elias, A. (1995): Monthly water balance of ungauged catchments obtained by geographical regionalization. Journal of Hydrology, 170: 277-291

Vertessy, R.A., Hatton, T.J., O'Shaughnessy, P.J. and Jayasuriya, M.D.A. (1993): Predicting water yield from a mountain ash forest catchment using a terrain analysis based catchment model. Journal of Hydrology 150: 665-700.

Vogel, R.M. and Kroll C.N. (1992): Regional geohydrologic-geomorphic relationships for estimation of low-flow statistics, Water Resources Research, 28(9): 2451-2458.

Weeks, W.D. and Hebbert R.H.B. (1980): A comparison of Rainfall-Runoff models. Nordic Hydrology, 11:7-24

Wheater, H.S., Jakeman A.J. and Beven K.J. (1993): Progress and directions in Rainfall Runoff Modelling. In: Jakeman A.J., Beck M.B. and McAleer MJ (Editors).Modelling change in Environmental systems, John Wiley and sons Ltd 101-132.

World Bank (2003): Towards water secure Kenya. Water resources sector memorandum of Kenya, Draft vol. 2, Nairobi, Kenya

World Meteorological Organisation (1994): Guide to Hydrological Practices, Publication No. 168. Geneva, Switzerland, 421-429

Woolhiser, D.A. and Brakensiek, D.L. (1982): Hydrologic system synthesis. In: Haan C.T., Johnson, H.P. and Brakensiek D.L, (Editors).Hydrologic Models of Small Watersheds. An ASAE Monograph No. 5, pp 1-15

Young, P.C. (2002): Databased Mechanistic and Top- Down Modelling. In: A. Rizzoli *et al.,* (Editors). Proceedings of the 1st Biennial Meeting of the International Environmental Modelling and Software Society, University of Lugano, Switzerland, Vol. 1: 363-374.

APPENDICES

Appendix 1.1 List of Rainfall Stations used in the Study

Station ID	Station Name	Longitude (Deg)	Latitude (Deg)	Start Year	End Year	% of Data Missing
9036233	Kimakia forest station	36.75	-0.77	1970	1975	0.0
9036259	Gatare forest station	36.76	-0.71	1970	1975	0.0
9037005	Gethumbwini coffee Estate	37.03	-0.97	1970	1975	0.0
9037096	Sagana fish culture farm	37.20	-0.67	1970	1975	1.4
9037015	Ragati forest station, Nyeri	37.17	-0.38	1970	1975	8.3
9036271	Kiandongoro Gate station	36.75	-0.48	1970	1975	0.0
9036291	Kiriaini Chief's camp	36.95	-0.60	1970	1975	1.4
9037100	Kanyaba Chief's camp	37.08	-0.47	1970	1975	2.8
9036104	Muriranja's Hospital	36.97	-0.75	1970	1975	1.4
9036292	Ichichikarura village station	36.83	-0.70	1970	1975	4.2
9036164	S. Kinangop forest station	36.68	-0.72	1970	1975	1.4
9036188	Sasumua Dam station	36.67	-0.75	1970	1975	4.2
9036220	Karuga farm station	36.90	-0.92	1970	1975	0.0
9037047	Bendor Estate, Thika	37.03	-0.97	1970	1975	1.4
9036072	Mweiga Estate, Nyeri	36.92	-0.35	1970	1975	0.0
9036268	Amboni Agric. camp	36.87	-0.33	1970	1975	0.0
9036274	Aberdare Part tree Tops	36.90	-0.35	1970	1975	1.4
9036144	Blue line farm station	36.95	-0.33	1970	1975	2.8
9036099	Kagochi, Karatina station	37.15	-0.38	1970	1975	0.0
9037151	Gumba Chief's camp	37.13	-0.60	1970	1975	6.9
9036017	MoW Nyeri station	36.95	-0.42	1970	1975	2.8
9036223	Nyeri Prisons station	37.32	-0.95	1970	1975	0.0
9036288	Nyeri Met. Station	36.97	-0.43	1970	1975	0.0
9037120	Kabaru forest station	37.15	-0.28	1970	1975	4.2
9037069	Karatina Hombe forest stn	37.12	-0.35	1970	1975	2.8
9037158	Sagana State Lodge	37.07	-0.37	1970	1975	0.0

Appendix 1. 2 List of Temperature stations used in the study

Temperature Station ID	Longitude in Degrees	Latitude in Degrees	Mean Monthly Temperature °C
9037158	37.07	0.37	17.41
9036157	36.83	0.45	13.84
9036164	36.68	0.72	11.59
9036288	36.97	0.43	17.73
9036233	36.75	0.80	12.89
9037096	37.20	0.67	21.15
9036260	36.87	0.13	17.05

Appendix 1. 3 Annual rainfall values for selected rainfall stations

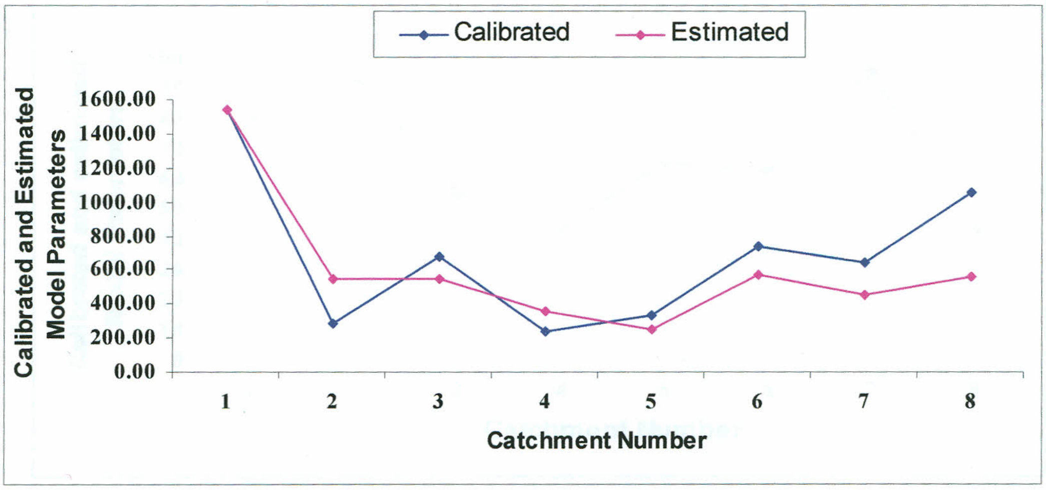
	Annual Rainfall (mm)					
Station ID	**1970**	**1971**	**1972**	**1973**	**1974**	**1975**
9036233	2223	1756	2272	1737	2113	1554
9036259	2399	2046	2619	1828	2444	1662
9037005	1236	980	771	791	1002	721
9037015	1492	1232	1442	1510	927	1399
9036291	1186	1049	2024	1796	1601	1180
9036164	1644	1255	1622	1411	998	1092
9036188	1512	1537	1752	1176	1433	1424
9036220	1628	1335	1371	1049	1705	1032
9036072	1080	818	882	752	800	727
9036274	1047	888	1035	883	910	728
9036144	794	582	644	624	657	592
9036099	1160	850	1355	985	909	1024
9037151	1207	887	1485	1086	1160	1304
9036223	667	441	543	511	474	599
9036288	925	601	1114	909	814	817
9037158	759	514	823	663	789	737

Appendix 1.4 Agro-ecological zones and land use types

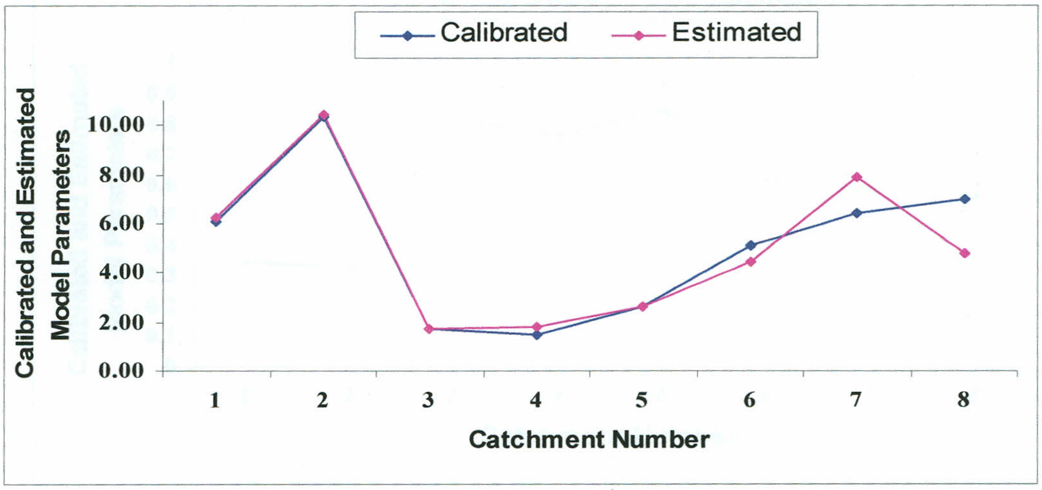
ZONE	DESCRIPTION	LAND USE TYPE
LH0	Lower Highland-Humid	Forestry
LH1	Lower Highland-Humid	Tea and Dairy farming
LH2	Lower Highland-Sub-Humid	Maize, Wheat and Pyrethrum
LH3	Lower Highland-Semi Humid	Barley and Wheat farming
LH4	Lower Highland-Transitional	Cattle, Sheep and Barley farming
LH5	Lower highland-Semi Arid	Lower highland Ranching
LM3	Lower Midland-Semi Humid	Cotton Farming
LM4	Lower Midland-Transitional	Maize, Millet sorghum and cotton
LM5	Lower Midland-Semi arid	Livestock, millet , Sorghum and beans
UH0	Upper Highland- Humid	Forestry
UH1	Upper Highland-Humid	Dairy and Sheep Farming
UH2	Upper Highland –Sub-Humid	Wheat and Pyrethrum Farming
UH3	Upper Highland- Semi Humid	Wheat and Barley farming
UM1	Upper Midland-Humid	Coffee and Tea farming
UM2	Upper Midland-Sub Humid	Main coffee growing zone
UM3-4	Upper Midland Marginal Area	Coffee and maize farming
UM3	Upper Midland-Semi- Humid	Marginal coffee growing
UM4	Upper Midland- Transitional	Sunflower and Maize farming
TAI and TAII	Tropical-Alpine Moor and Heath lands.	National parks and limited grazing potential

The Table is in reference to Figure 1.5 (Source: Jaetzold and Schmidt, 1983)

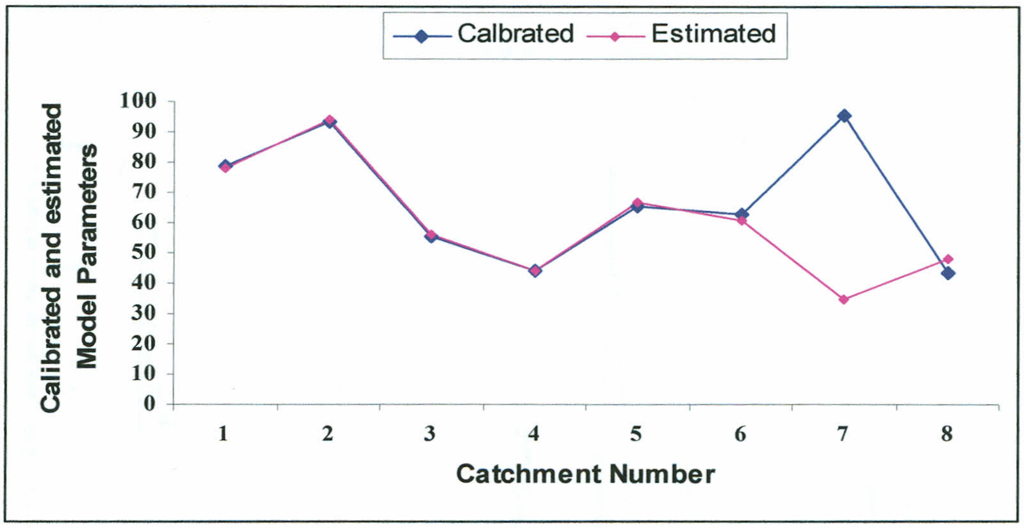
Appendix 2.1 Calibrated and estimated values of parameter 1/C



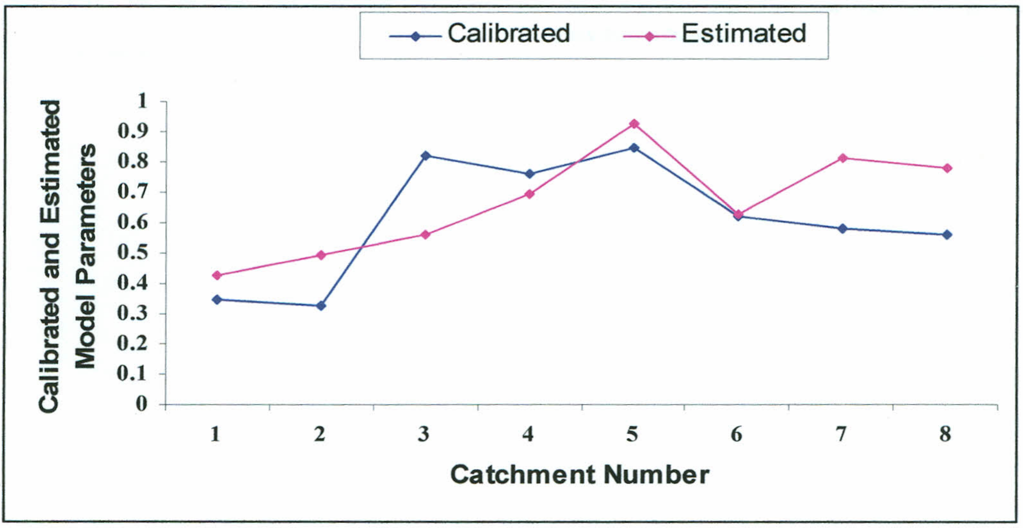
Appendix 2.2 Calibrated and estimated values of parameter τ_q



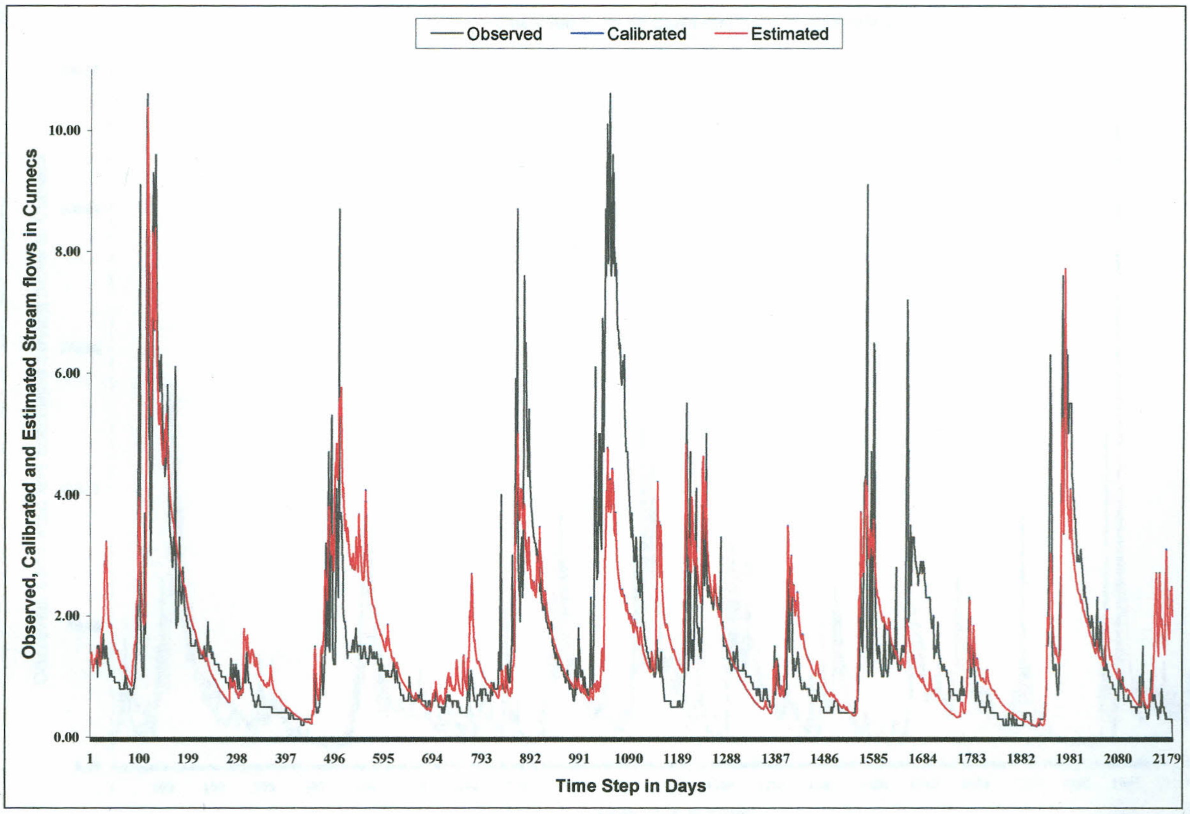
Appendix 2.3 Calibrated and estimated values of parameter τ_s



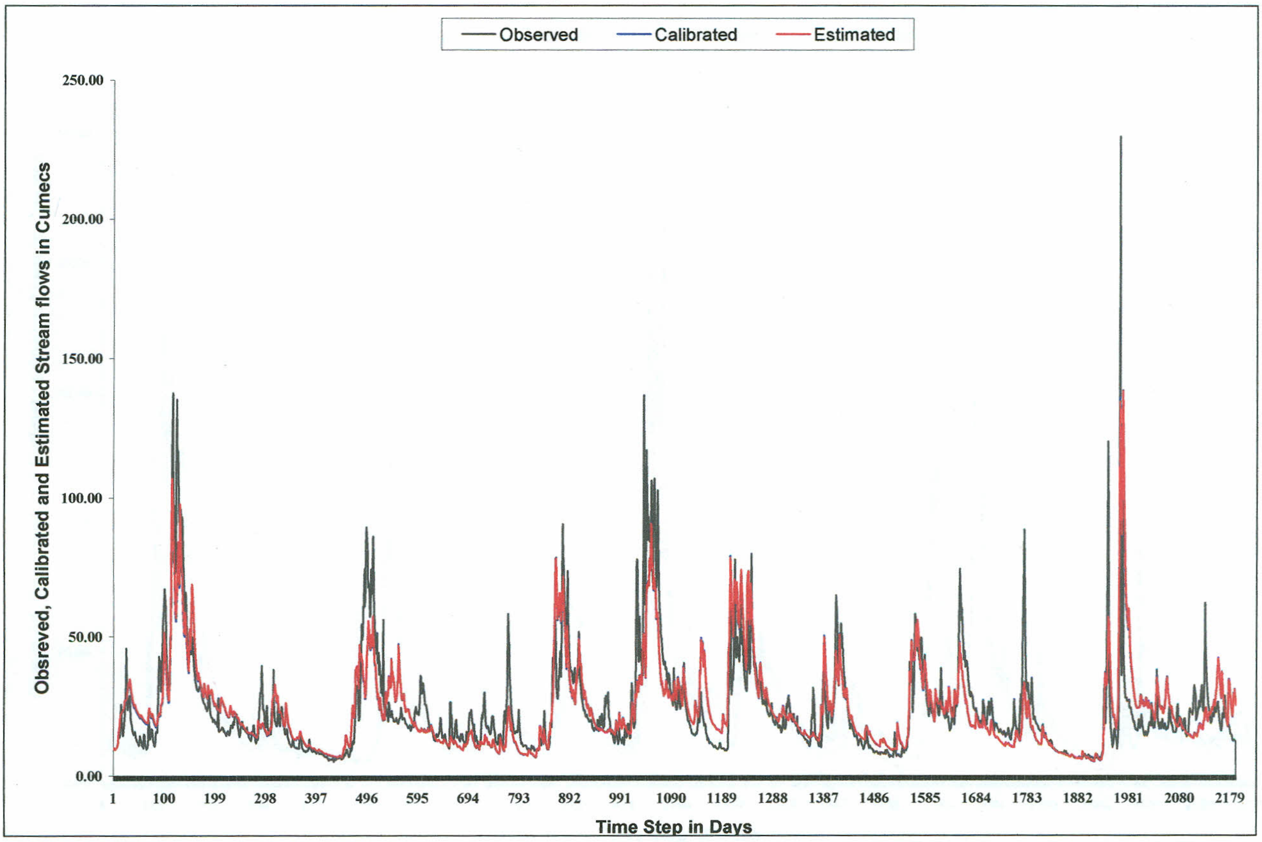
Appendix 2.4 Calibrated and estimated values of parameter v_s



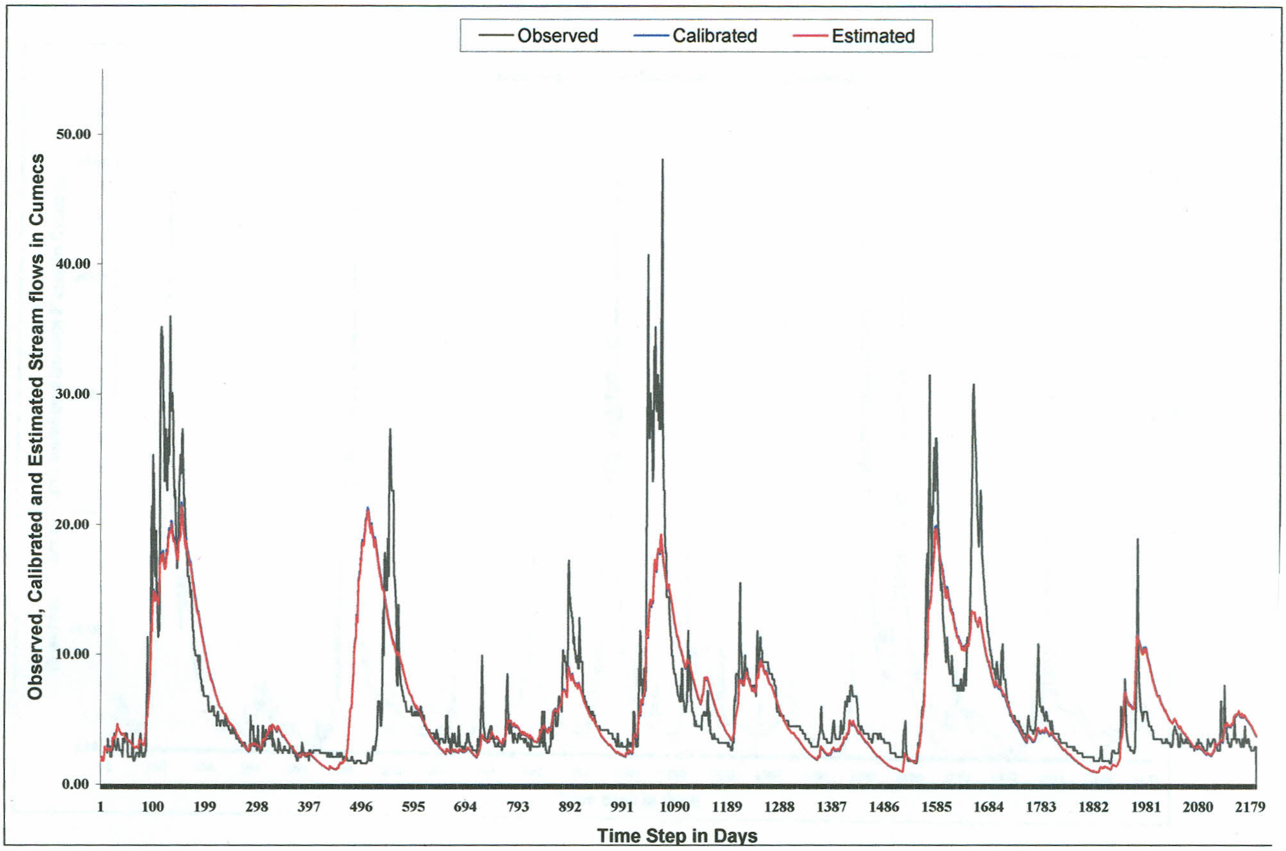
Appendix 3.1 Observed, calibrated and estimated flow for catchment 4BC04



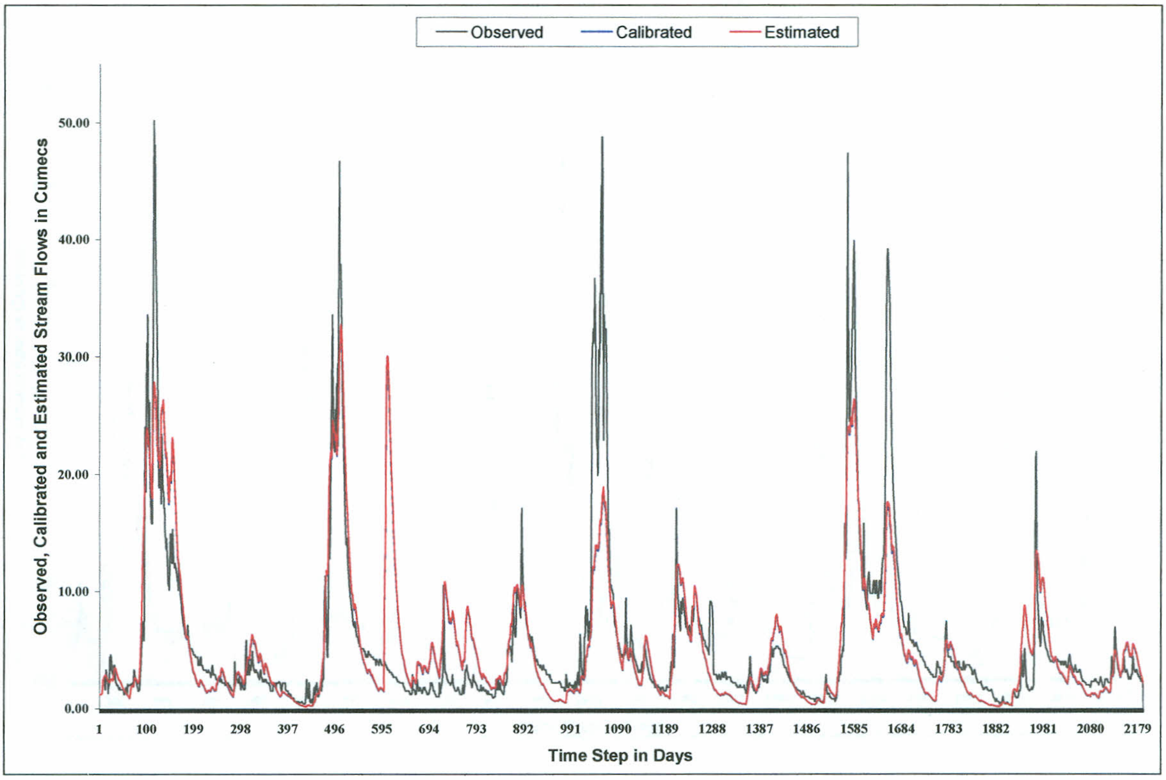
Appendix 3.2 Observed, calibrated and estimated flow for 4BC02



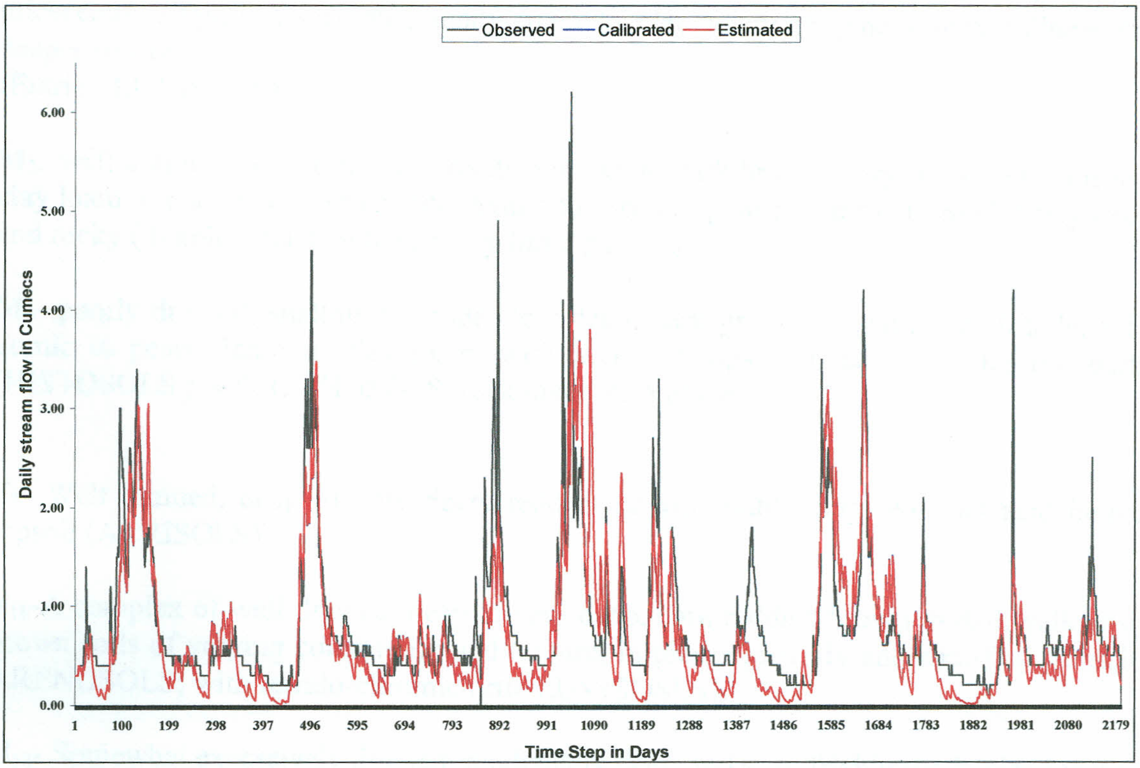
Appendix 3.3 Observed, calibrated and estimated stream flow for 4CA02



Appendix 3.4 Observed, calibrated and estimated stream flow for 4CB04



Appendix 3.5 Observed, calibrated and estimated stream flow for 4CB07



Appendix 4.1 Description of major soil types found in the study area

A_5-Well drained to poorly drained, very deep, and brown to dark brown, friable, micaceous, slightly calcareous, sandy loam to clay loam; in places with saline-sodic deeper subsoil
(Eutric FLUVISLOS)

M_2- well drained, very deep, dark reddish brown to dark brown, very friable and smeary, clay loam to clay, with a thick, acid humic top soil; in places shallow to moderately deep and rocky (humic ANDOSOLS, partly lithic phase.)

M_9- poorly drained, shallow to moderately deep, dark greyish brown, very friable, acid humic to peaty, loam to clay loam, with rock outcrops and ice in the highest parts (HISTOSOLS) with LITHOSOLS, rock outcrops and ice.

F_7- Well drained, deep to very deep, reddish brown friable clay, with an acid humic topsoil (ACRISOLS)

F_{16}-A complex of well drained, deep to very deep, dark reddish brown to dark yellowish brown soils of varying consistence and texture; in places gravelly and stratified (ferralic ARENOSOLS; with ferralo-chromic/ortic LUVISOSLS)

H_{12}- Somewhat excessively drained, moderately deep, red very friable sandy clay loam to sandy clay, in places rocky (CAMBISOLS; FERRALSOLS and rock outcrops)

H_{15}- Complex of excessively drained to well drained, shallow, dark red to drown, friable sandy clay loam to clay; in many places rocky, bouldery and stony in places with an acid humic topsoil (REGOSOLS, LITHOSOLS and CAMBISOLS and ROCK OUTCROPS)

L_1-Well drained, very deep, dark red, very friable clay (nito-rhodic FERRALSOLS)

L_2-Poorly drained, deep, black to dark grey, very firm, cracking clay **(**pellic Vertisols and verto-luvic PHAEOZEMS

L_3- Well drained, moderately deep to deep, dark brown friable to firm clay in places with humic topsoil (NITOSOLS)

L_4-Complex of well drained, shallow to very deep, dark red, friable clay; in many places rocky and bouldery (nito-rhodic FERRALSOLS and chromic CAMBISOLS, lithic and/or bouldery phase)

L_{11}-Poorly drained **,** very deep , dark grey to black, firm to very firm, bouldery and stony, cracking clay; in places with a calcareous, slightly saline deeper subsoil (pellic VERTISOLS, stony phases and partly saline phase)

L$_{12}$- Poorly drained, deep, black to dark grey very firm cracking clay (VERTISOLS)

L$_{17}$-Complex of:

Moderately well drained, shallow, yellowish red to dark yellowish brown, friable, gravelly clay over petroplinthite or rock (50-70%), (IRONSTONE SOILS; with LITHOSOLS)

Poorly drained, deep to very deep, dark brown to very dark greyish brown, mottled, firm to very firm, cracking clay; in places moderately deep to very deep over petroplinthite (Undifferentiated VERTISOLS and vertic GLEYSOLS)

L$_{21}$-Poorly drained, deep very dark greyish brown, mottled, firm clay, abruptly underlying a thick top soil of friable silty clay loam (solodic PLANOSOLS)

V$_2$ – Complex of well drained to poorly drained, shallow to moderately deep, dark reddish brown to very dark greyish brown, firm slightly to moderately calcareous, rocky stony or gravely clay.

R$_1$- Well drained, extremely deep, dark reddish brown to dark brown, friable and slightly smeary clay, with acid humic topsoil (ando-humic NITISOLS: with humic ANDOSOLS).

R$_2$- Well drained, extremely deep, dusky red to dark reddish brown, friable clay, with an acid humic topsoil (humic NITISOLS)

R$_3$- Well drained, extremely deep; dusty red to dark reddish brown, friable clay, with inclusions of well drained, moderately deep, dark red to dark reddish brown, friable clay over rock, pisoferric or petroferric material (eutric NITISOLS with nito-chromic CAMBISOLS AND chromic ACRISOLS, partly pisoferric or petroferric phase)

R$_5$- Well drained, moderately deep to very deep , dark reddish brown , friable to firm, cly (Nito-ferric LUVISOLS; with humic NITISOLS).

Ux4-Well drained, very deep, dark reddish brown to dark brown, very friable and smeary, silty clay loam, with a humic top soil (mollic ANDOSOLS)

Ux5- Well drained, very deep, dark reddish brown to very dark greyish brown, friable and slightly smeary clay, with humic topsoil (ando-luvic PHAEOZEMS)

UM15- Well drained, very deep, dark red, friable to firm, clay (nito-rhodic FERRALSOLS)

UM17-Well drained, moderately deep to deep, dark reddish brown to brown, friable to firm, sandy clay loam; in places with an acid humic top soil (ferralo-orthic ACRISOLS; with dystric and humic CAMBISOLS and humic ACRISOLS)

UM19- Well drained, moderately deep to very deep, dark reddish brown to dark yellowish, friable to firm, sandy clay to clay; in many places with a top soil of loamy sand to sandy loam (ferralo-chromic/orthic/ferric ACRISOLS with LUVISOLS and FERRALSOLS)

UM20- Well drained, moderately deep to deep, dark red to yellowish red, friable, sandy clay loam to clay (rhodic and orthic FERRALSOLS; with ferralo-chromic/orthic/ferric ACRISOLS)

UP4-Complex of well drained to poorly drained**,** shallow to very deep, dark red to black, friable to firm, cracking clay; in places sodic(pellic VERTISOLS; with verto-eutric NITISOLS, verto-eutric PLANOSOLS and orthic SOLONETZ, partly lithic phase)

Appendix 4.2 Classification of soil depths for soil types found in the study area

The effective soil depth is described using the following Key:

Depth range	Description
0-50 cm	Shallow
50-80 cm	Moderately deep
80-120 cm	Deep
120-180 cm	Very deep
More than 180 cm	Extremely deep

Source: Sombroek *et al.*, 1980

Appendix 4.3 Description of major land cover types found in the study area

AG$_1$- Rain fed herbaceous crop

AG-1B Scattered (In natural vegetation or other) Rain fed herbaceous crop (Field density 10-20 % of polygon area.).

AG-1C Isolated (in natural vegetation or other) Rain fed herbaceous crop (Field density 10-20 % of polygon area.

AG-2 Irrigated herbaceous crop

AG3-Rainfed shrub crop

AG3B-Scattered (In natural vegetation or other) rain fed shrub crop (Field density 20-40% of polygon area.)

AG-4 Forest plantation –Undifferentiated

AG-6 Rice fields

FR$_1$- Multi-layered trees (broad leaved evergreen)

FR$_2$- Closed trees

FR$_3$- Open trees (65-40% crown cover)

FR$_4$ – Very open trees (40-15 % crown cover)

FR$_5$- Closed to open woody vegetation (Thicket)

FR$_6$- Closed Shrubs

FR$_7$- Open Shrubs (45-40% crown cover)

RL$_2$-Shrub savannah

RL$_6$- Closed herbaceous vegetation on permanently flooded land.

UR- Urban and associated areas –rural settlements.

Source: FAO, 1997