COMPARATIVE CASE STUDY OF RAINFALL - RUNOFF MODELS OVER

THE NYANDO RIVER BASIN

BY

STEPHEN KIBE RWIGI

Department of meteorology

Faculty of Science

University of Nairobi

A thesis submitted in part fulfilment for the degree of

Master of Science in meteorology,

University of Nairobi.

May 2004

DECLARATION

This thesis is my original work and has not been presented for examination in any other university

Signature................................
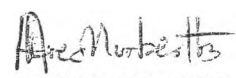
STEPHEN KIBE RWIGI   10/5/2004

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

Signature................................   17/05/2004

PROF. F. M. MUTUA

Department of meteorology

Faculty of Science

University of Nairobi

Signature................................   17/05/2004

DR. A. O. OPERE

Department of meteorology

Faculty of Science

University of Nairobi

DEDICATION

This work is dedicated to the four people most precious to me.

My wife Lucy

My son Mark

My daughter Sarah

My niece Jacqueline

TABLE OF CONTENTS

ABSTRACT

The main objective of this study was to assess the performance of some existing rainfall-runoff linear systems models and a conceptual model using data from the Nyando catchment. In order to estimate the optimum parameters for the various models, split samples of data were used; one sample for calibration and the other for verification.

The data used comprised daily areal average rainfall, daily average runoff and daily average evaporation. Regression analysis method was used to estimate missing rainfall and runoff data while seasonal mean method was used to estimate missing evaporation records.

Homogeneity of the data was tested using the mass curve method. Results of homogeneity test indicated that data from the catchment are generally homogeneous as shown by the high R^2 efficiency values.

The performance of each of three rainfall-runoff linear systems models, the Simple Linear Model (SLM), the Linear Perturbation Model (LPM) and the Linearly Varying Gain Factor Model (LVGFM), and a conceptual model called the Soil Moisture Accounting and Routing (SMAR) model, was assessed using data from the Nyando catchment. The linear systems models were applied in both non-parametric and also under the constraint of the gamma function impulse responses. Optimum parameters were obtained by the method of Ordinary Least Squares (OLS) and by Rosenbrock's search technique for non-parametric and parametric modes respectively.

Results obtained in the simulation mode indicate that there is a good agreement between the actual and estimated stream flow when the conceptual SMAR model is used. For the SLM there is a marked difference between actual and estimated stream flow especially in the case of low and high flow seasons, respectively. The performance of the other models falls somewhere in between the performance of these two models.

The conceptual SMAR model appears to be more superior to the linear systems models with a higher R^2 efficiency (71%) than those of the linear systems models. Among the linear systems models, the LVGFM performs best on the Nyando catchment at R^2 efficiency of about 69 %, followed by the LPM at R^2 efficiency of 55 % and the SLM at R^2 efficiency of 47 % in this order.

From these results, the SMAR model may be considered the best model for the Nyando catchment among all the models that were considered in this study. Among the linear systems models considered, the LVGFM is the best model for the catchment followed by the LPM and the SLM in this order.

LIST OF TABLES AND THEIR CAPTIONS

LIST OF FIGURES AND THEIR CAPTIONS

LIST OF ACRONYMS AND ABREVIATIONS

AR	Autoregressive Updating
CCD	Cold Cloud Duration
GFFS	Galway Flow Forecasting Systems
IUH	Instantaneous Unit Hydrograph
IVF	Index of Volumetric Fit
JAS	July August September Season
KMD	Kenya Meteorological Department
LBDA	Lake Basin Development Authority
LPM	Linear Perturbation Model
LVEMP	Lake Victoria Environment Management Project
LVGFM	Linearly Varying Gain Factor Model
MAM	March April May Season
MSE	Mean Square Error
OLS	Ordinary Least Squares
OND	October November December Season
SLM	Simple Linear Model
SMAR	Soil Moisture Accounting and Routing model
UCG	University College Galway
UNDP	United Nations Development Programme
UNEP	United Nations Environment Programme
WMO	World Meteorological Organisation

1.0   INTRODUCTION

Sustainable water resources management depends on reliable hydrological forecasts that can be used for important and sometimes sensitive decision-making support. The basic management process is to collect and assess facts so that informed decisions can be made on future strategy and on which actions to be taken based on the assessment of the available facts. To predict future conditions and allow for strategic planning, some form of a model is always required. Modern managers and designers of hydrological resource projects are thus accepting models as interactive decision support tools. This is probably because model forecasts permit decisions to be made with more confidence and hence allow for a more efficient use of water resources (Falconer, 1992). This makes model forecasts a major factor in economic development since water is a major player in any economic activity.

The use of computers in all aspects of hydrology has lead to increased emphasis on catchment modeling; where a catchment model is defined as a set of mathematical abstractions describing relevant phases of the hydrologic cycle with the sole objective of simulating the conversion of rainfall into runoff (Ponce, 1989). One of the major uses of a model is in deciding what action to take based on the model's predictions. For instance, the extent of losses due to hydrological catastrophes such as floods or droughts after a decision has been made based on a model forecast as opposed to what would have happened if the decision were not made at all; justify the usefulness of a forecast. Such forecasts would necessarily have to come from a hydrological model that is capable of simulating the catchment characteristics as accurately as possible. Water resources managers would use such model outputs as the main decision support tools. Thus hydrological models have a major role to play in all aspects of hydrological resource management.

Hydrological models are probably the most powerful tools available for clarifying the significant processes occurring in a natural hydrological system and for predicting the effects of any changes to the system. There is need, therefore, to model hydrological systems in order to solve the problems of decision making in both the design and operation stage of water resource projects. Hydrological forecasts therefore have a vital economic value.

The basic model inputs are rainfall, evaporation and runoff. Rainfall is the major source of most of the water resources in a catchment and it is therefore taken as the main

model input while runoff is taken as the model output. Rainfall is the main source of runoff, evaporation, infiltration as well as the ground water. The level of runoff in the rivers is a determining factor in the development of many hydrological projects. The proportion of rainfall that finally transforms to runoff determines this level. This proportion is always a function of the catchment's physical properties, which determine the level of abstractions such as evaporation, infiltration and percolation. Since runoff comes from rainfall, it is important to use a rainfall-runoff model to forecast future discharge from a catchment. In such a model rainfall data is used as the input in order to forecast the future discharge from a river.

Changes in the state of a river are related to weather changes occurring over the catchment area. Changes in hydrological processes are normally found to be less abrupt and always lagging behind those of underlying meteorological factors. The relatively slow response of hydrological processes to the changes in meteorological processes is what makes it possible to model runoff (WMO 1975). The purpose of river flow forecasting models is to enable us to provide reliable information for water resources management. Such forecasts may be useful in areas such as:

(i)   Issuing of flood warnings that help in planning for the evacuation of populations threatened by rising water levels

(ii)   Management of water storage reservoirs more effectively and efficiently

(iii)   Simulation of possible discharge series from the catchment during the design of various hydrological structures

(iv)   Prediction of effects of proposed works on a catchment on the flow regime at the catchment outlet.

Early warnings about high water levels make it possible for measures to be taken in order to strengthen the embankments, protect bridges, evacuate populations and their property from the flood area.

The success of many hydrological projects is highly dependent on reliability of forecasts available to operators of these projects. These forecasts' reliability will always depend on the forecasting model. The forecasting model should be easy to operate. Data required for the model should also be affordable and easily available. The rainfall-runoff models are easy to operate and their data requirements are affordable. Projects such as hydropower generation, irrigation schemes, fresh water supply, flood control, water quality control, water based recreation, fish and wildlife management require proper planning and

2

management in order for them to succeed. The planning and management of such projects relies on timely forecasts derived from detailed analyses of surface water flow. In addition, forward planning in case of floods or drought in a given catchment requires accurate and reliable forecasts. Such forecasts are derived from a good model that best responds to the basin hydrometeorologic characteristics.

Since no single model structure is capable of producing consistently accurate forecasts for a wide range of catchments and climatic conditions, there is need to compare performance of various available rainfall-runoff models for suitability to the catchment in question. This is important especially where the models are developed using data from a catchment that is different from the one under study. Most models have the potential for success or failure depending on the matching of the chosen model to the catchment under study. Since models are only a representation of reality, each type of model has its own limitations and approximations. Comparing the performance of different existing rainfall-runoff models using data from the Nyando catchment in Lake Victoria Basin, will help in choosing the best model, out of the existing ones, that best fits the catchment behaviour in transforming input to output. This model will be useful in issuing forecasts for flood warnings, rational regulation of runoff, construction and operation of hydraulic structures such as bridges, culverts and spillways of dams.

Realizing that most hydrological models have been developed and applied in temperate zones, there is need to evaluate the performance of these models in tropical regions in order to select the one that best fits our needs (Adegu, 1999). This is normally done through recalibration of these models using data from the tropical regions.

1.1 Statement of The Problem

Nyando River is used as a source of food, domestic water, irrigation water and industrial water for sugar factories such as Miwani, Muhoroni and Chemelil, which are all located within the catchment. The growing population has increased the demand for water, food, and land for settlement. The demand for food and shelter has forced the people living in the catchment to farm and settle within the flood plains of Kano, which are found in the lower reaches of River Nyando. This means that more and more people are becoming vulnerable to the frequent flood events in the Kano plain which are known to cause a lot of agony to the area residents. When the floods occur, people are displaced from their homes

3

and there is loss of life for both human and livestock either through drowning or as a result of water borne diseases.

Living in this area becomes total chaos because during these floods, villages, farms and roads get completely inundated, thus disrupting farming activities as well as the transportation network. The timing of these floods and the subsequent evacuation of the inhabitants and their properties are quite difficult without an early warning system. Prior planning with the help of a reliable model output that is easily applicable, would go a long way in alleviating such calamities. Besides, the residents need water both for irrigation and domestic purposes. To achieve a rational use of this water, drawn mainly from river Nyando, it is important that a model, which is capable of forecasting changes in the flow regime of the Nyando River, the main watercourse in the catchment, be identified.

1.2   Justification

In view of the above problems in the area, there is need to develop a comprehensive hydrological modelling system for timely warning of floods and efficient management of the available water resources in the catchment. A model that can simulate flood discharge on the lower reaches of the Nyando catchment, using rainfall from the upper reaches as input, and with a lead time of at least one day would be useful in saving many lives of people and livestock living in the flood plains of the Nyando catchment. It would be particularly helpful in making decisions related to flood protection measures within the catchment as well as the rational use of water for irrigation. The output of this study will form a basis for an early warning system for floods in Nyando. The rational use of water from river Nyando would also be improved.

1.3   Objectives

The main objective of this study is to identify a suitable rainfall-runoff model, from among some existing models, for use in the Nyando catchment. The specific objectives are:

(i)   Determine the best calibration parameter values for the Nyando catchment using some of the existing linear systems models particularly the Simple Linear Model (SLM), the Linear Perturbation Model (LPM), the Linearly Varying Gain Factor Model (LVGFM), and a conceptual model particularly the Soil Moisture Accounting and Routing (SMAR) model using rainfall, evaporation and runoff data from the Nyando Catchment.

4

(ii) Assess the optimum performance of each of the linear systems models by comparing their outputs among themselves as well as comparing their performance to that of a conceptual model.

(iii) Find out and recommend, based on their performance, the most appropriate model for use in decision-making support such as in flood forecasting and warning in the Nyando Basin.

In the next section, a review of some of the relevant work that has been done in this field is presented.

1.4 Literature Review

Many attempts have been made to synthesize the unit hydrograph from rainfall records. This has been done using the instantaneous unit hydrograph (IUH) to formulate mathematical expressions for the shape of the unit hydrograph (Shaw, 1988). The instantaneous unit hydrograph is defined as the runoff caused by a unit volume of rainfall input generated instantaneously and uniformly over the whole catchment (Mutua, 1979). Nash (1957) postulated that the transformation by a catchment of an effective rainfall into surface runoff could be modeled by routing that rainfall down a cascade of equal linear reservoirs. Essentially the synthesis of the instantaneous unit hydrograph (IUH) is the basis of hydrological modelling and has given rise to many hydrological models.

The complex interactions of hydrological processes have made the development of a successful rainfall-runoff model a real daunting task. However, the challenge has been considered and there now exist in literature many rainfall-runoff models. Few of these can be applied indiscriminately beyond the conditions for which they were developed and none are totally satisfactory since in almost all cases, modelled runoff deviates from observed runoff (Ward and Robinson, 1990).

During the UNDP/WMO (1974) study of the Lake Victoria Basin, the available measured runoff data were found to be insufficient. Rainfall data were however, sufficiently available. It was then found necessary to generate runoff using the available rainfall data. Sugawara and Murayama (1956) rainfall-runoff model was used for the purpose of generating synthetic runoff records from the available rainfall data. This model was specifically applied in the Nyando catchment and the results were found to be acceptable. The model consists of linear cylindrical vessels in series and/or parallel analogous to basin runoff. This model utilises annual, seasonal, monthly, daily, hourly or smaller time scales of

5

rainfall as input to generate corresponding mean runoffs. For the case of Nyando catchment, monthly rainfall was used to generate the corresponding runoffs for the period 1961 to 1968. The objective of the model was to fill in missing runoff records and this was successfully achieved.

Adegu (1999) used the satellite Cold Cloud Duration (CCD) data to simulate rainfall data, which he then used to develop a rainfall-runoff model for the Nyando basin. Adegu used TAMSAT model to produce quantitative precipitation estimates. These estimates were then applied to the PITMAN model to simulate runoff. The results revealed that TAMSAT rainfall estimates and PITMAN model have a lot of potential in the Nyando catchment subject to availability of good quality data. Events of short and medium duration could be predicted using this model. It should be noted that Adegu used only one type of model (conceptual) to simulate runoff. However Adegu did not assess how well black box models would have performed with CCD data.

Lake Victoria Environmental Management Project (LVEMP 2002) commissioned a study over the whole of Lake Victoria Basin whose main objective was to provide advice and assistance to the Water Quality and Ecosystem Management component of LVEMP in the three East African countries over a period of 18 months from 1st August, 2000. The study incorporated many task forces. Among these was the Meteorology/Hydrology task force, which was charged with the task of coming up with a time series of all inflows and outflows into and out of the lake for the period 1950 to 2000. This was not an easy task as there were a lot of gaps in the discharge data unlike in the rainfall data records. The task force recommended the application of rainfall-runoff modelling in order to fill in the missing river discharge data records using the available rainfall records.

For this purpose the task force opted to use the Sacramento model to try and generate river discharge records from rainfall data. Though the model was found suitable, it was soon abandoned due to both technical and administrative problems. Among other technical problems, the model was found to be difficult to apply, as it requires characteristic values for at least sixteen parameters. The amount of work involved in calibrating the model was found to be enormous. The task force thus recommended the application of the NAM and SMAP models, which were found to be user-friendlier than the Sacramento model. The models were able to generate acceptable river discharge data records. Where rainfall records were found missing, the task force used correlation analysis and double mass curve method to fill in the missing rainfall data records. Here a double mass curve was evaluated

6

for the subject station, and the reference station and a trend line fitted to the curve. The equation of the trend line was used to fill in as many gaps as possible in the subject station records. The main emphasis of the project was water quality. Little attention was paid to flood related problems, which are a perennial nuisance in the basin.

Mutua (1979) used Wiener-Hopf theory of optimum dynamic linear systems to develop a rainfall-runoff model for Nzoia river catchment. The model is based on stable unit hydrographs of the Nzoia catchment determined from coincident records of rainfall as input and runoff as output. Mutua (1979) used two-hourly and eight-hourly unit rainfall hydrograph to develop the model. The two-hourly unit rainfall hydrograph model was applied to the upper catchment while the eight-hourly unit rainfall model was applied to the whole catchment. Mutua (1979) developed a criterion for selecting storms, which could effectively be monitored by the automatic rain gauges in the upper sub-catchment. These were used to determine the two-hour unit hydrograph. Mutua (1979) found that two-hourly stream flows could be simulated if the number of automatic gauges could be increased in the lower sub-catchment. The eight-hourly unit hydrograph simulated responses compared well with observed stream flow. The upper sub-catchment, with two-hourly unit hydrograph rainfall was found to give unrealistic and unstable runoff response functions. This was possibly due to sparse distribution of existing self-recording rain gauge stations. Mutua (1979) thus used the diurnal distribution patterns of rainfall in the basin to synthesize eight-hourly unit hydrograph for the whole catchment.

Nzoia is known to be a homogeneous catchment in terms of slope. It has a slope that is almost uniform in most areas as opposed to Nyando where we encounter plains on the western parts and hills on the eastern parts. This makes Nzoia suitable for modelling using lumped models. The same may not be said of the Nyando Basin.

In his study of the upper Athi river catchment, Opere (1991) developed a simple model for the catchment based on the response of discharge to the catchment rainfall values. The model divides the catchment into sub areas centred on lines of constant time of travel of water to the outlet. Opere (1991) used areally averaged rainfall to plot isochrones using cross-correlation and coherence functions. The functions were used to derive the mean kernel function for the simple model. Opere (1991) found that the maximum response was centred within 1 to 5 days. Using separate data Opere (1991) confirmed that the model performed well within limits of experimental errors. Since the model is based on average response to the catchment rainfall values, Opere (1991) suggested use of hourly data to

improve the model. Opere's work was focused on Athi river catchment where no serious flooding problems have been reported. This may not be very helpful to the Nyando residents.

Mutulu (1984) studied both Yala and Sondu catchments concurrently. Mutulu (1984) used daily areal average rainfall and runoff series for Yala and Sondu basin to develop a transfer function model. This model was found to reproduce the main features of the observed hydrographs quite well. Mutulu (1984) suggested a possibility of using multivariate input transfer functions to improve the model further. The shortcomings of this model are mainly the fact that it requires both rainfall and runoff data for previous days in order to predict the future. Mutulu's model was limited to up to two days lead-time forecasts. The model used few years of data and large departures were observed especially during flood events; 20 and 30 years of data could improve the performance of this model.

Kato (1982) has used Wiener-Hopf theory to develop a rainfall-runoff model for Ruvu River basin in Tanzania. Kato (1982) found that stream flow and rainfall have almost similar statistical properties indicating a close similarity between rainfall and stream flow series. Three different unit response functions corresponding to the three sub catchments were computed. Different rainy seasons were chosen to assess the accuracy of these unit response functions.

The estimated stream flow was found to be closer to the actual stream flow for smaller sub-catchments. For large sub-catchments, there were marked deviations from the actual observations especially when rainfall was extremely heavy. These obvious discrepancies in the model output results could be taken as indicators that the model could be limited to the size of catchment. This appears to be especially the case when one uses a linear approach to estimate stream flow using rainfall as input. Kato (1982) thus suggested that the rainfall intensity be classified on the basis of light, moderate and heavy rainfall. This would lead to the development of three unit response functions on the basis of rainfall intensity; light, moderate and heavy.

Wanjohi (1999) has studied water quality and flow modelling over the Nyando basin. The water quality model was found to have potential for operational use as a decision support tool for water management in the area. It can be used, as a low-cost alternative tool to indicate what pollution control measures should be taken. DUFLOW package was used to design the flow and quality of the model. Flood protection measures were simulated for the lower parts of Nyando that included dykes, by pass and wetland clearing.

Construction of a by-pass and dykes were found to be satisfactory alternatives as flood protection measures.

1.4.1 Overview of Hydrological Modelling

A hydrological model is a simplified representation of a real hydrological system. Models, being representations of hydrological systems, are generally less complex than the real hydrological systems and therefore cannot always represent all the minute details of the system. They are, nevertheless, necessary tools of studying the real hydrological systems, whose output results provide us with new information about the processes taking place in the real system; thus helping us to acquire detailed knowledge about system. Models are also essential tools in performing complex analysis as well as in making informed decisions based on their output results.

Hydrological models may be classified into two main groups, namely, material and formal models. A material model is a simplified physical representation of the prototype, simpler in structure but with properties resembling those of the prototype. On the other hand a formal model is a mathematical abstraction of an idealized situation that preserves the important structural properties of the prototype. Formal models are also referred to as mathematical models. They are a formulation of the past hydrological events that can be integrated into the future (Ponce, 1989). In terms of cost of application, material models are more costly than mathematical models. Mathematical models are readily available, highly flexible and comparatively inexpensive. They are therefore preferred over the material models, which are not readily available.

Mathematical catchment models may further be classified into two main groups, namely, Deterministic and Stochastic models. Deterministic models seek to simulate the physical processes in the catchment involved in the transformation of rainfall into runoff. These models imply a cause and effect relationship between chosen parameter values where results are obtained from the solution of certain prescribed equations.

River flow at the outlet of a catchment depends on various independent variables such as rainfall, evaporation, types of soils and vegetation cover among others. The process of linking such variables and the river flow is deterministic in nature. Thus the classical approach would be to develop a deterministic model based on the physics of transformation. This approach, however, fails in hydrological forecasting due to the complexity of the boundary conditions (Nash and Sutclife, 1970).

9

Stochastic models on the other hand take into account the chance of occurrence or probability distribution of the hydrological variables. Hence a stochastic model accounts for randomness such as the magnitude and timing of various processes in the hydrological cycle. This is not the case in deterministic models. For purposes of river flow forecasting, deterministic models are normally preferred. These are themselves divided into two main categories namely the black box and conceptual models.

Hydrologists have long recognized the complexity of the problem and hence resorted to the use of simplified hydrological models. Such simplified models may be described as conceptual or empirical models depending on whether they are or are not capable of physical interpretation.

Kachroo (1992) observed that while it is desirable that a model should represent as closely as possible the actual physical processes occurring within the catchments, it is essential that the model should also represent accurately the transformation of the input into an out-put. The primary utility of a model is reflected in the extent to which it satisfies this practical objective. This is the model efficiency and is one of the principle requirements from a model. The second requirement is that of model consistency. In this requirement the level of accuracy and the estimates of parameter values should persist through different samples of data. The third requirement is that of model versatility. This requires that the model should be accurate and consistent when subjected to diverse applications involving model evaluation criteria not directly based on the objective function used to calibrate the model.

Over and above these requirements the model should also be simple to calibrate, simple to operate and still give acceptable results for the intended applications.

1.4.1.1   Linear Systems Models

These are models, which are formulated in terms of linear equations and processes. They may also be described as empirical black box models since they mathematically describe the rainfall-runoff relation without regard to the physical processes that relate them. These models rely less on the physics of transformation of rainfall into runoff. They depend more on the observation of factual relations in the existing records of the catchments. Though simple in nature, empirical black box modelling approach may, and sometimes does give results as good as those of models that are considered to be more superior in terms of representing the physical processes governing the transformation of rainfall into runoff.

10

Reporting on the 'Workshop on Mathematical Models in Hydrology', held in Pisa, Italy, in 1974, Clarke (1977) noted that 'simple black box models remain competitive with supposedly physical-based, large-number-parameter models'. He went on to state that the validity of this conclusion lies on how intrinsically good the models are, the adequacy of the available data and the extent to which the assumptions of the models are met in the catchments under study. Traditional methods of curve fitting, multiple regression and unit hydrographs are examples widely used in empirical hydrology (Liang, 1995).

Many of the empirical black box models owe their origin to the Sherman (1932) non-parametric unit hydrograph method for modelling the storm-runoff component of the discharge hydrograph. The unit hydrograph represents the response of a catchment to a unit volume of rainfall of a defined duration. Thus empirical black box models may be considered to be operating on the principles of linearity where a linear system is defined (Dooge, 1973) as one that satisfies the property of superposition. This means that the system's response characteristics are additive; where additive means that the sum of the output caused by a sum of inputs is equal to the sum of outputs produced by each of the inputs individually. The system should also be time-invariant; where its parameters do not change with time. For such a system the form of output depends only on the form of input and not on the time at which the input is applied. These models require only synchronous rainfall and runoff data as input and output respectively.

In this study the performance of three linear systems models were compared for data collected from the Nyando river catchment. The linear rainfall-runoff models, which were used, are the Simple Linear Model (SLM), the Linearly Varying Gain Factor Model (LVGFM) and the Linear Perturbation Model (LPM).

1.4.1.2   Conceptual Models

Conceptual models are generally a simplified representation of the physical processes in a hydrological system. They are normally based on mathematical descriptions, which simulate complex processes by relying on a few key conceptual parameters. The models attempt to simulate the dominant physical mechanisms and processes responsible for the conversion of rainfall to runoff in a simple yet practical way where they describe, in a simplified way, the movement of water in a basin in both space and time and also account for the storage (Maidment, 1993) They are developed by purely rational consideration involving the interplay of inductive and deductive reasoning.

The parameters of a conceptual model are either physically measurable or can be interpreted in physical terms. The effects of some physical changes of the river basin systems are reflected in certain parameters, which could be changed without the need to recalibrate the model. This may be particularly important for a river basin such as Nyando where the land use is changing rapidly due to urbanization or other factors such as increase in population leading to increased demand for land for settlement and farming activities.

Some conceptual models are based on the unit hydrograph theory. The essence of the unit hydrograph theory is the assumption of linearity in the relation between the storm and the effective runoff (Mutua, 1979).

These models attempt to explain the physical processes involved in transforming rainfall into discharge. They require additional data such as evaporation unlike the linear systems black box models where only rainfall and runoff data are required. Conceptual models can often be complex in structure, having many parameters each with a physical interpretation. Most of these parameters are calibrated by optimization methods based on the optimum matching of the model output to the observed stream flow records (Kachroo, 1992).

The generation of runoff by these models is expressed by a series of prescribed operations. The model structure involves complex systems of equations that are based on theoretical concepts governing the hydrological processes in a basin. However it should be noted that the complexity of a conceptual model is not an indicator of the model's efficiency in terms of its ability to match the model output with the observed discharge. Previous studies have shown that less complex models sometimes perform better than the more complex ones (Loague and Freeze 1985).

This group of models is more scientifically based than the black box models. One of the most widely used conceptual models is the Soil Moisture Accounting and Routing (SMAR) model. Unlike the systems analysis type of models, the SMAR model takes into account the effects of evaporation in determining the volumes of runoff. Unlike rainfall the effects of evaporation are not immediate. Over a period of time evaporation creates a soil moisture deficit and thus controls the generation of runoff from a subsequent storm. A conceptual model takes care of the delayed effects of evaporation.

CHAPTER 2

2.0 STUDY AREA, ITS PHYSICAL CHARACTERISTICS AND CLIMATE

The study area is located in the western side of Kenya and includes the whole of that area drained by the Nyando River and its tributaries. This forms the Nyando River catchment, which is drained into Lake Victoria. It is found within the Lake Victoria Drainage Basin, which is classified as the hydrological drainage area No.1 in Kenya. It covers an area of about 3580 square kilometres.

The Nyando River catchment spans across four districts namely Kericho, Kisumu, Nandi and Uasin Gishu. River Nyando is the principal watercourse in the catchment with two main tributaries namely: the Nyando itself originating from Mau and Londiani forests in Kericho area, and Ainomotua originating from the northern side of the Nandi escarpment. The area of interest is found within latitudes $0^\circ 7' N$ and $0^\circ 25' S$ and longitudes $34^\circ 34' E$ and $35^\circ 43' E$ as shown in Figures 2.1 and 2.2.

2.1 Nyando Basin Sub-Catchments

The Nyando river Basin has three main sub-catchments that represent the complete hydrological regime of the study area. These are:

(i)   The Nyando sub-catchment,

(ii)   The Ainamotua sub-catchment,

(iii)   The Cheronget – Kabletach – Asawo sub-catchment.

2.1.1 Nyando Sub-Catchment

This is the largest sub-catchment and is found on the upper side of the confluence of Nyando and Ainamotua as shown in Figure 2.2. It covers an area of about 1680 square kilometres, constituting about 47 % of the total area of the Nyando River catchment. It lies south of the equator and is separated from Ainomotua sub-catchment by the Tinderet volcanic dome. It generally slopes from east to west at an average slope of about 1.38 %. The Nyando tributary, which is the principal watercourse in the sub-catchment, originates from the Western Mau forest at an elevation of about 3000 m above sea level. It flows for about 100 km from the head reaches before joining Ainomotua just over 50 km upstream of the river mouth to form the main River Nyando. In the 100 km stretch, many smaller streams join the Nyando tributary.

2.1.2 Ainomotua Sub-Catchment

This is the second largest sub-catchment and covers an area of about 947 square kilometres, which comprises about 26.5 % of the total Basin area. The equator divides it into two roughly equal portions as shown in Figure 2.2. It generally slopes westwards at an average slope of about 2.36 %. The Ainomotua tributary is the principal watercourse in the sub-catchment. Though smaller in area, the sub-catchment is the largest contributor of water to the main River Nyando: contributing about 50 % of the water annually.

The Ainomotua tributary originates from the slopes of Tinderet and Timboroa forests at an elevation of about 3000 m above sea level. Many smaller tributaries join it before it joins the Nyando tributary to form the main River Nyando.
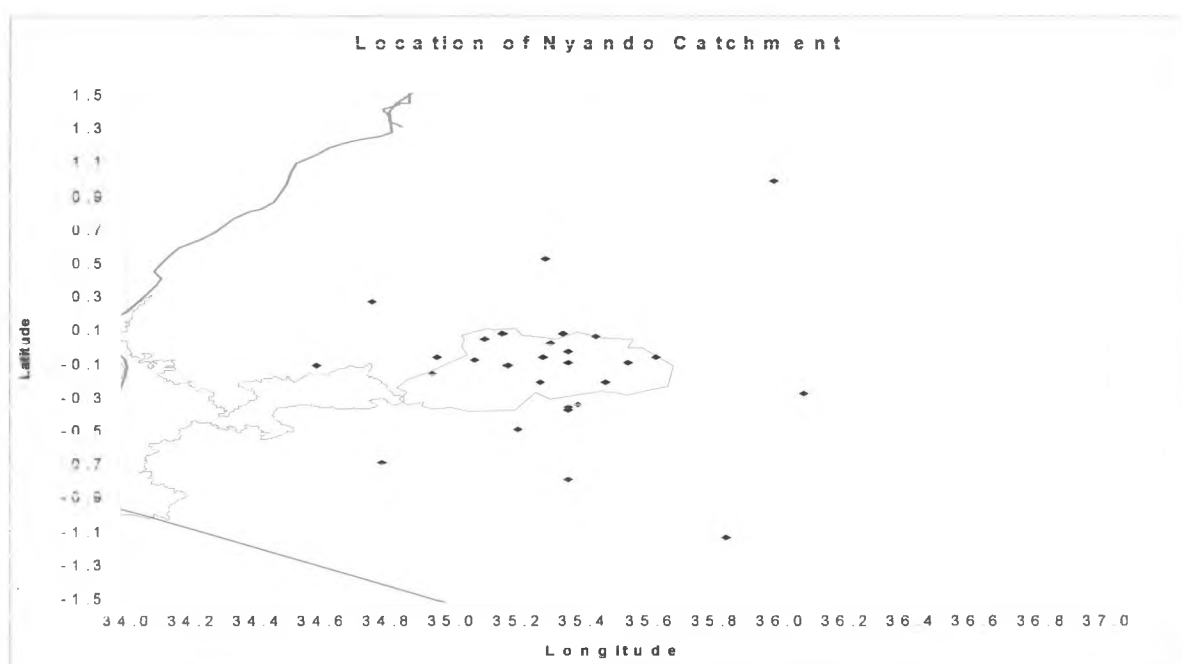
Figure 2.1: Location of Nyando Catchment within Kenya and rain gauge stations.
(Source: Adegu, 1999)

2.1.3 Cheronget- Kabletach – Asawo Sub-Catchment

This sub-catchment lies in the southern portion of the Nyando river Basin. It covers an area of about 767 square kilometres, which is about 21.4 % of the total area of the catchment. It is therefore the smallest of the three sub-catchments as shown in Figure 2.2. This sub-catchment is unique from the other two in that it slopes from southeast to northwest and is swampy in the lower reaches. The Cheronget, Kabletach and Asawo

streams drain it. These are the main streams in the sub-catchment, which are actually tributaries of the main Nyando. The contribution of these streams to the main river is very small. This may probably be attributed to the fact that most of their flow is lost in the Miruka swamp. Thus the main Nyando River may be considered as principally formed by the Nyando and Ainomotua tributaries.
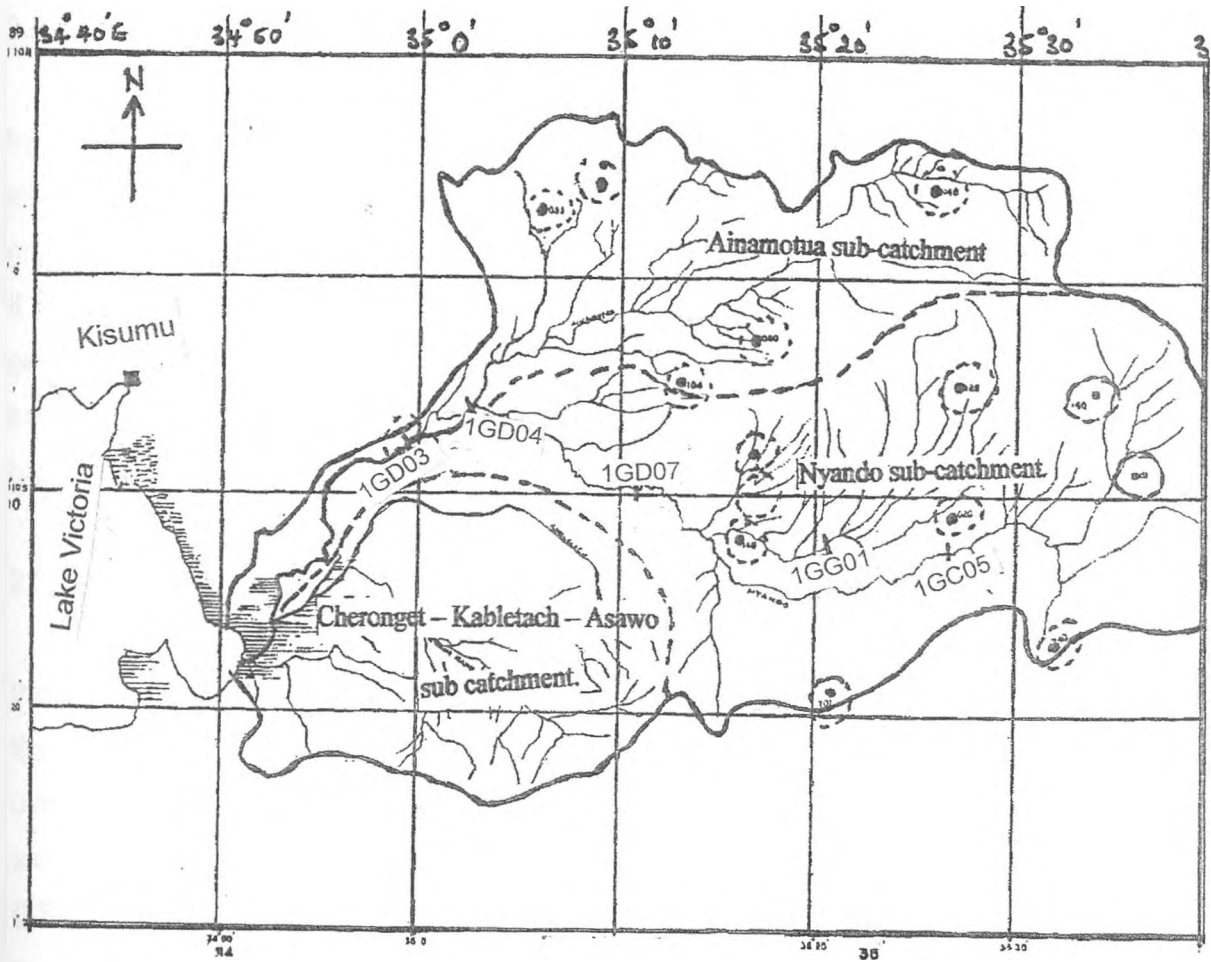
Figure 2.2: The Nyando Basin sub catchments, Rainfall and River gauging stations. (Source: UNDP/WMO 1974)

River Nyando spans a total length of about 150 km from the source on the Eastern end of the basin to the mouth on the western end of the basin. The highest point in the basin is about 3000 m above sea level at the Mau and Tinderet forests and the lowest is 1134 m above sea level at Lake Victoria level. The upper reaches of the Nyando basin have a higher average gradient than the lower reaches. The low potential gradient in the lower reaches of the Nyando river basin brings about impeded flow leading to the frequent

flooding experienced in the Kano plain. The flooding problem is not an issue on the upper reaches of the basin owing to the large potential gradient. However, high potential gradient is a possible factor contributing to soil erosion in the upper reaches of the basin. The eroded soil brings about siltation in the lower reaches of the river. This worsens an already serious flooding problem.

2.2 Physiographic and Hydrometeorologic Characteristics of The Basin

The quality and quantity of runoff generated in a drainage basin are affected by the hydrologic input to the basin and the basin's physical, vegetative and climatic features. Not all the water that occurs in a basin as rainfall transforms to runoff. Rainfall striking the ground may go into storage on the surface or in the soil and into the ground water reservoir. It is therefore important for the hydrologist to have a good understanding of the topography, geology, soils and land use patterns of the drainage basin under study. In this section a brief discussion of landform, geology, soils, land use and climate of the Nyando catchment is presented.

2.2.1 Landform

The landform varies from low plains near the lakeshore to plateaus and mountains to the east. This comprises the Kano plain on the western side and the Londiani Mountains on the eastern side. Land elevation varies from 1134 m above sea level on the lakeshore to over 3000 m above sea level at the summit of Londiani Mountains. The basin generally slopes from east towards west with relatively flat areas as you approach Lake Victoria (LBDA, 1987).

The general shape of the landscape, the steepness of the slopes and the total relief affect the way rainfall reaches the streams in any drainage basin. It should be expected that the steeper the slopes, the quicker will the runoff reach the main river draining the basin. As in most stream channel profiles, river Nyando exhibits the characteristic of decreasing slope proceeding in a downward direction (WMO/FAO, 1963). The relatively flat gradient of the lower reaches of river Nyando results in slowed rate of flow of the river.

Nyando basin is generally pear-shaped ovoid, which is a typical shape of many river basins (LVEMP, 2002). A catchment shape is known to affect the catchment response, which refers to the relative concentration and timing of runoff. Long narrow basins would be expected to have attenuated catchment response, whereas rotund basins would be

expected to have a highly peaked catchment response. Pear-shaped ovoid catchments, as is the case for the Nyando, are expected to exhibit a middle ground situation.

2.2.2 Geology And Soils

Most of the physical characteristics of a basin are influenced by geology. Geological factors also largely determine the storage time during which water is held between the rainfall and the eventual runoff as stream flow.

Nyando River Basin has the same geology as that of the main Lake Victoria basin. Geological formations in the basin vary from recent quaternary sediments to old rocks of Archean age. The most common formation is tertiary volcanic rock occurring in most of the eastern parts and extends to parts of the lakeshore area. The Bukoban system of Precambrian age covers Kisii and the surrounding areas including the Nyando river basin.

Rocks of Archean age of the Kavirondian and Nyanzian system are also common in the basin. The structural framework of rock material determines the rate at which water moves through the rock. It also determines the direction of movement of water and therefore mixing of waters from different aquifers (Opere, 1998)

The characteristics of the soils and rocks largely determine the storage system into which rainwater will enter. The soils in the Nyando basin are predominantly clays but vary greatly in texture, composition and structure. Soils derived from the quaternary volcanic rocks, which are generally fertile, are found in the higher rainfall areas on the eastern side of the catchment. Those soils derived from very ancient granite are reasonably fertile and tend to be in areas of low rainfall within the catchment. Potential runoff largely depends on rainfall intensity on one-hand and soil characteristics on the other. Soil characteristics play an important role in determining flood peaks. The more impervious soils like the black cotton soil produce high surface runoff. The structural framework of a given type of soil determines the rate at which water moves through the soil (Opere, 1998)

The lower reaches of the Nyando basin comprise generally medium to heavy clay soils. These soils are known to have impeded drainage and to exhibit poor structure. This is probably why the lower reaches of the basin are swampy. Soils in the upper reaches are well drained as opposed to soils in the lower reaches.

Runoff is one of the most difficult parameters to determine due to the many factors affecting its calculations. These include infiltration, drainage, permeability of soil and its structure. The soil type influences vegetative cover which in turn tends to retard overland

A diagram showing the expected variation throughout the year of rainfall, evaporation and discharge usually expresses the hydrological climate of a catchment most easily (Kachroo, et al 1992). This has been done for the Nyando catchment and the hydrological diagram is presented in Figure 2.3.
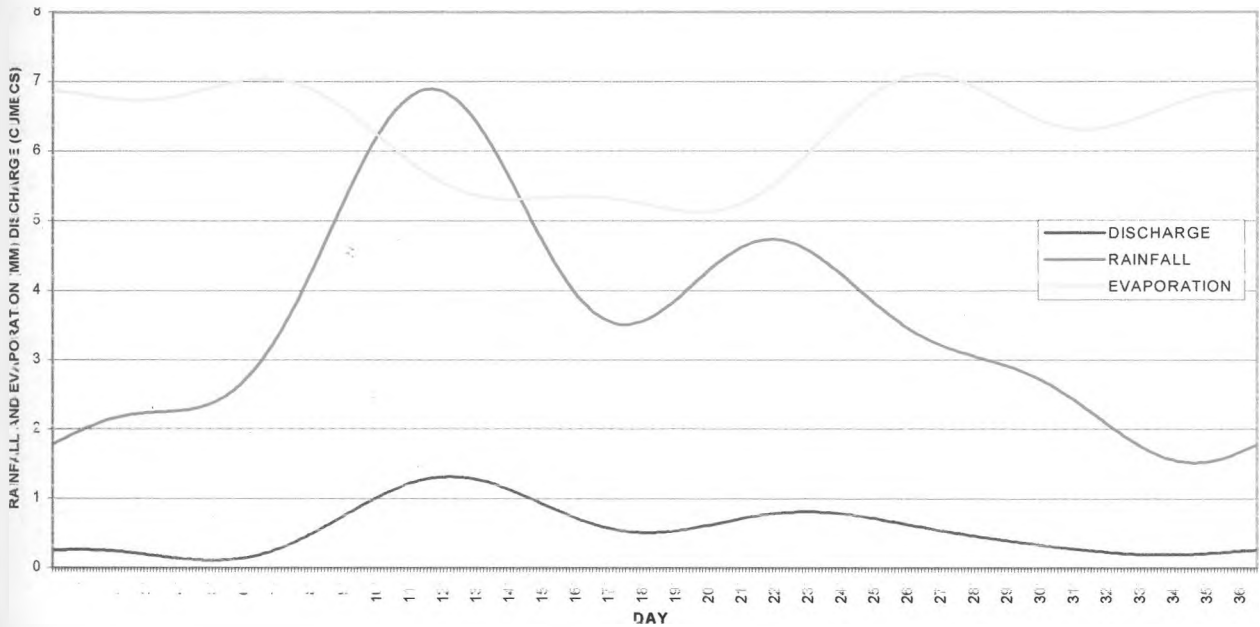
Figure 2.3: Hydrological diagram showing the seasonal mean variation of daily rainfall, discharge and evaporation over the Nyando catchment.

Figure 2.3 presents the daily seasonal hydrological diagram for the Nyando catchment. The figure presents the flow regime of river Nyando from January to December. The figure shows that there are two high flow seasons. The first one is centered in May and the other in August. The peak flows are observed to coincide with the two main rainfall seasons. It is clear from the figure that the flow regime follows the pattern of rainfall but always lagging behind by a few days.

The figure also shows how seasonal daily evaporation varies within the year. It is observed from the figure that the highest rate of evaporation occurs in March and the lowest in July. Interestingly the highest rate of evaporation occurs one month before the highest peak rainfall in April while the lowest occurs one month before the next peak rainfall in August. On average, it is observed that the rate of evaporation is high throughout the year. The seasonal variation of evaporation is not significantly pronounced, as it is with rainfall.

19

Rainfall is normally used as the descriptor of climate within the tropics. The variation of rainfall from January through December is usually considered a sufficient descriptor of tropical climate. Other meteorological variables do not exhibit significant variation throughout the year. The climate of the Nyando catchment therefore may most easily be described by a rainfall graph showing the expected variation from January to December. For this purpose monthly averaged rainfall graphs were plotted from seven selected stations within the catchment. These stations include: a station from each sub-catchment, the station at the highest, and the other at the lowest, elevation, and the station with the highest, and the other with the lowest, normal annual rainfall. Table 2.1 shows the average monthly rainfall from which the graphs were drawn.

Table 2.1: Monthly average rainfall at 7 typical rainfall stations within the Nyando catchment.

Month	Ahero	Kipkurere	Tambach	Kipkorech	Lumbwa	Nandi H	Koru	Mean
Jan	71.5	38.1	31.6	79.3	58.3	80.4	83.0	63.2
Feb	84.1	40.7	31.1	70.2	49.8	67.7	81.5	60.7
Mar	130.5	72.9	80.2	137.4	94.8	143.4	153.2	116.0
Apr	181.5	157.0	158.1	243.1	178.8	205.6	223.8	192.6
May	138.0	131.9	131.0	230.4	159.8	184.8	162.7	162.6
Jun	73.7	107.4	80.9	155.8	116.0	139.5	108.5	111.7
Jul	65.3	177.5	108.0	134.3	171.3	160.9	110.2	132.5
Aug	90.9	183.5	111.2	180.3	162.8	175.0	103.8	143.9
Sept	67.6	88.0	38.8	137.4	98.5	139.2	89.6	94.2
Oct	96.0	89.7	94.0	147.4	73.5	129.9	104.4	105.0
Nov	103.9	73.7	113.4	136.7	87.2	146.9	113.4	110.7
Dec	90.2	36.6	37.1	60.7	48.0	71.8	73.7	59.7

Figure 2.4 shows average monthly rainfall graphs at seven selected rainfall stations in the Nyando catchment. From the figure we see that there are two major peaks for each station. The first major peak is centered in April and normally occurs in March-April-May (MAM) season. The second major peak is centered in August and normally occurs in July-August-September (JAS) season. A weak peak, centered in November is also recognizable and normally occurs in October-November-December (OND) season. From the graphs, it is observed that rainfall in the Nyando basin exhibits a trimodal pattern.

We can therefore safely say that there is rainfall in the Nyando catchment throughout the year with April being the wettest month and February the driest. Most of this rain falls in two distinct seasons as shown in Figure 2.3
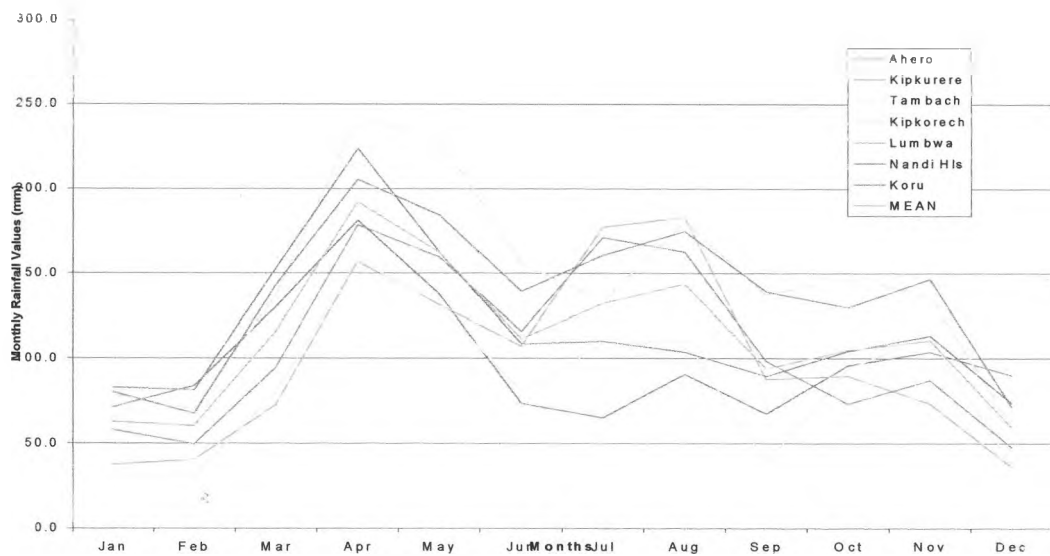
Figure 2.4: Average monthly rainfall at seven typical rainfall stations in the Nyando catchment.
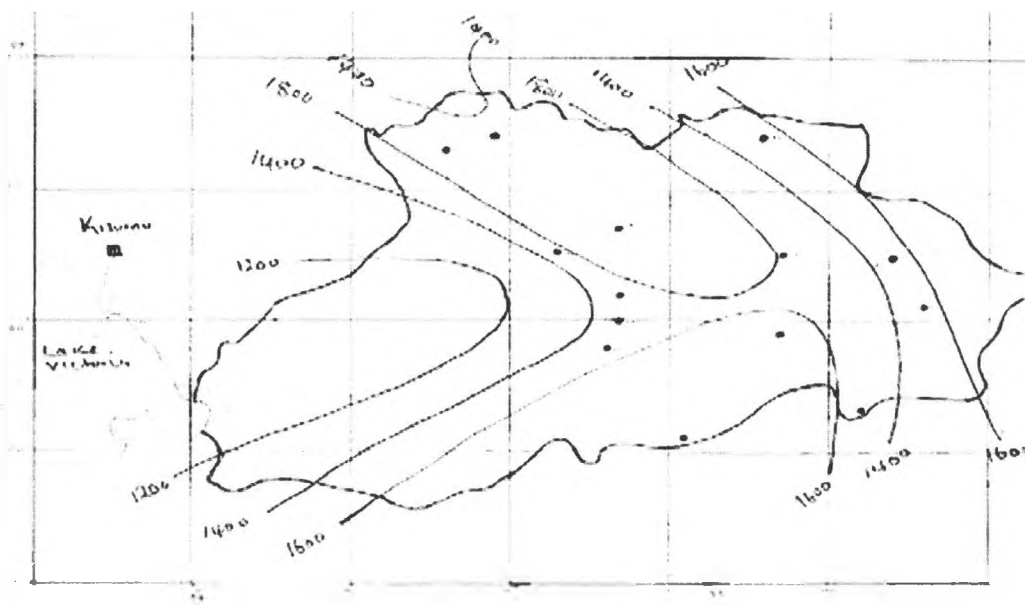
Figure 2.5: Annual rainfall distribution in mm over the Nyando catchment.
(Source: LBDA 1986)

The first peak shows the long rains season and the second peak shows the short rains season. The average annual rainfall in the catchment is about 1400 mm.

There is no distinct dry period and thus we can regard this as a moderately humid catchment. Effects caused by evaporation are thus minimum and may be ignored when

deciding on the best model to apply on the catchment.

From the study of the normal annual map of the Nyando river basin (Figure 2.5) it is observed that over the higher reaches of Ainomotua sub-catchment the normal annual rainfall varies from about 1400 mm to over 1800 mm. Going downstream the rainfall decreases gradually to below 1300 mm annually around the region of confluence between the Nyando and the Ainomotua.

Rainfall then decreases north of the Nyando-Sondu divide towards lake Victoria west of Londiani forest to well below 1100 mm per annum around the lakeshores. Low annual rainfall of about 1200 mm to 1300 mm is experienced over the area downstream of Nyando-Ainomotua confluence. Over the central parts of the Nyando catchment the normal annual rainfall varies from about 1400 mm to about 1500 mm. Rainfall then increases northwards and southwards towards the Nyando escarpment and the Nyando-Sondu divide respectively reaching normal annual values of well above 1800 mm. Rainfall is also observed to decrease from the central parts of Nyando catchments going either to the west towards Londiani forest with annual values falling to below 1200 mm (UNDP/ WMO, 1974).

3.0   DATA MANAGEMENT

The success of any research is mainly dependent on the handling and quality of data. It is therefore important to devote sometime to establish the suitability of data before carrying out any research. This chapter presents a brief description of data and methods used in determining the suitability of data for this study.

3.1 Database and Type

The data sets required for this study are: daily rainfall, evaporation and runoff. A brief discussion of availability, quality, statistics and management of these data sets is presented in this section.

3.1.1   Rainfall Data

Rainfall data sets are normally obtained by measurements using rain gauges. The data sets can be assembled using hourly, daily, monthly or yearly intervals. For the purpose of this study, daily interval data was used, which are the point rain gauge measurements accumulated over a period of one day.

Daily rainfall data sets for the Nyando catchment were available form the Kenya Meteorological Department (KMD). They were available in many rain-gauging stations within the catchment; more than those that were used in this study. Out of the large number of stations, 20 were initially selected. After preliminary analysis, 18 stations were selected out of these. These were the stations that had long spells of continuous and few missing data records relative to the other stations. The stations were selected in such a way that they were representative over the area in the sense that they covered the entire catchment as well as sub-areas with different rainfall characteristics such as high-altitude areas with high rainfall and low plains with low rainfall. The selected stations are located upstream of Ogillo; river gauging station number 1GD03.

The number of years with data varied from station to station, but all the 18 selected stations had data sets for the period 1968 to 2000. Table 3.1 below shows the 18 stations, their location, year of records and the percentage of missing data from 1980 to 2000, the period considered for the purpose of this study.

It is observed from the table that except for gauging stations G1 and G4, all the other stations have less than ten percent of the missing data records. Station G1 may not be used

as a source of data owing the large number of missing data records. Twenty eight percent of missing data is way above the allowed value of ten percent and below.

The rain gauge station-coding used in this study is only for the purpose of convenience.

Table 3.1: Rainfall stations used for the purpose of this study: their location, years of records and the percentage of missing data. The station coding, Gi, is used here only for the purpose of convenience.

Station No.	Station Code	Station Name	Latitude	Longitude	Years of records	Missing Data (%)
8935001	G1	Songhor, Kaabirir	2' N	35°18' E	1962 – 2000	28.6 %
8935033	G2	Nandi Hills Savani	3' N	35°6' E	1960 – 2000	8.0 %
8935148	G3	Kipkurere Forest	5' N	35°25' E	1960 – 2000	6.4 %
8935159	G4	Ainabkoi Forest	5' N	35°5' E	1962 – 2000	10.2 %
8935161	G5	Nandi Hills Tea Est.	5' N	35°9' E	1962 – 2000	1.0 %
9034086	G6	Ahero Irrigation	9' S	34°56' E	1960 – 2000	1.9 %
9035020	G7	Kipkelion Railway	12' S	35°28' E	1960 – 2000	7.6 %
9035075	G8	Kaisugu House	20' S	35°23' E	1962 – 2000	3.7 %
9035148	G9	Koru Bible School	12' S	35°16' E	1962 – 2000	4.3 %
9035151	G10	Londiani Entomology	8' S	35°35' E	1962 – 2000	3.6 %
9035155	G11	Londiani Forest Stn	3' S	35°37' E	1962 – 2000	3.7 %
9035188	G12	Tinga Monastry	12' S	35°27' E	1964 – 2000	4.4 %
9035201	G13	Kipkorech Estate	19' S	35°20' E	1968 – 2000	8.4 %
9035220	G14	Koru Homaline co.	10' S	35°17' E	1962 – 2000	3.2 %
9035226	G15	Londiani Forest	5' S	35°21' E	1960 – 2000	8.4 %
9035240	G16	Keresoi forest	17'S	35°32'E	1961 – 2000	3.6 %
9035256	G17	Malagat Forest	5' S	35°32' E	1962 – 2000	6.5 %
9035263	G18	Tinderet Tea Est.	21' S	35°21' E	1964 – 2000	4.6 %

3.1.2 Runoff Data

Runoff measurements usually require an elaborate stream flow level gauging procedure. Measurement and interpretation of stream flow level is one of the chief occupations of hydrologists concerned with the study and control of surface water. The gauging procedures are time consuming besides being costly to implement.

Daily stream flow data sets for River Nyando were available from the Department of Water Resources Management, which operates a network of about 24 water-level recording stations within the catchment. Of the 24 stations, only three are located on the main river. These are 1GD01, 1GD03 and 1GD04. Station 1GD01, which is located at the catchment outlet, is no longer functional. Gauging station number 1GD03 was chosen to be the source of runoff data. For the purpose of filling in missing records, runoff data sets from stations 1GD04, 1GD07, 1GC04, 1GC05 and 1GD07 were used. All these stations had data sets for different time periods but they all had data from 1969 to 1990. Station 1GD03 had data up

to 1997 and had few missing data records relative to other stations.

Station 1GD03 was chosen because of its strategic location. It is on the upper side of the flood plains as well as being almost at the outlet of the catchment. There is an area of about 2625 square kilometres upstream of the station. This represents about 74 % of the Nyando catchment area, which is drained through this station. The station was also found to have continuous daily discharge data from 1983 to1995 with only 118 days of missing records. This is only about 1.8 % of the available records. This was deemed to be adequate for the calibration and verification of the model.

3.1.3 Evaporation Data

Daily evaporation data were available from Kenya Meteorological Department. There were no evaporation records available from within the catchment since evaporation is only recorded at synoptic meteorological stations. There was none within the catchment. Evaporation data sets were therefore obtained from Kisumu and Kericho stations both of which are found outside the Nyando catchment. The number of years with data varied between the stations, but both stations had data records for the period 1983 to 1995

3.1.4 Limitations of Data

Data availability and adequacy were the main limitations. Due to economic constraints in third world countries such as Kenya most hydrological data are hard to find in adequate quantities. Missing data sets were therefore common occurrences. Suitable methods had to be found to fill in these missing data sets. Filling in the missing data for both rainfall and runoff was a difficult task because of the number of stations involved especially in the case of rainfall data. Runoff data sets were scattered and limited while evaporation data were only available at stations neighbouring the catchment namely Kisumu and Kericho Meteorological stations. Daily evaporation data sets were also not computerized. The task of keying in these data was enormous and time consuming.

3.1.5 Conversion of Daily Data To UCG Format

Before the data sets were fitted to the model they were converted to the University College Galway (UCG) format since the model could not take them in their original format. This is a standard format acceptable to the Galway Flow Forecasting System (GFFS) program.

3.2 Estimation of Missing Data

Continuity of data is a vital requirement in any research. Incomplete records of hydrological data sometimes occur possibly due to operator error or equipment malfunction. In such cases, it is often necessary to fill in the missing records using estimated values. This section deals with some of the methods used to estimate missing data records.

Like many other parts of the developing world, there are problems in Kenya in sustaining operational equipment for stream flow and rainfall measurements. Further, the personnel to take readings are either too few or unreliable (Opere 1998). These factors often lead to missing records where automatic recorders are unavailable. The estimation of missing records therefore, is a fundamental step and forms an integral part of this study.

Missing data may be estimated by many methods, which include the arithmetic mean, auto-correlation, regression analysis, isopleth's analysis, normal ratio, inverse distance, and seasonal mean, among other methods. In this study regression analysis method was used to estimate the missing records of rainfall and runoff while seasonal mean method was used to estimate missing evaporation records. Regression method was used because it involves more than one station in filling in the missing data records. This ensures that the estimated value of the missing record is more representative. This was more suited to the rainfall and runoff data since more than one linearly correlated stations were available. Where only the station with the missing data is available, like in the case of evaporation data, a seasonal mean value gives a better estimate of the missing data record.

3.2.1 Linear Correlation, Regression Analysis And Double Mass Curve Method

Correlation and regression procedures are widely used in hydrology and other sciences. The premise of the method is that one variable is often conditioned by the value of another or several other variables.

Before applying this method, it was important to first compute the linear correlation coefficients ($r_{x,y}$) between pairs of stations. The correlation coefficient is a value between -1 and $+1$ that indicates the strength of linear relationship between two quantitative variables. A linear correlation coefficient of $+1$ indicates a perfect positive relationship, a value of 0 indicates no relationship, and a value of -1 indicates a perfect negative relationship. If a data set were such that all of the x,y pairs plotted form a straight line, then $r_{x,y}$ would be equal to $+1$ or -1 exactly. For $+1$ the plot would be upward sloping indicating that data set x is the same as data set y while for -1 the plot would be downward sloping indicating that

data set x is the negative of data set y.

The available data was used to compute correlation coefficients between the selected stations. This was done for the purpose of identifying the level of the linear relationship between the selected stations. Values of $r_{x,y}$ close to +1 imply a near perfect linear direct relationship while those close to zero suggest lack of a strong linear relationship.

The stations that had highly correlated data to those of the station with the missing data records were selected for the purpose of filling in the missing records.

The sample linear correlation coefficient ($r_{x,y}$) was computed from Equation 3.1.

$$r_{xy} = \frac{Cov(x,y)}{\delta_x \delta_y} \text{------------------------------------} 3.1.$$

In Equation 3.1, $r_{x,y}$ refers to the linear correlation coefficient and ranges from -1 to 1, Cov(xy) refers to the covariance shared by the two variables x and y, δ_x and δ_y refer to the standard deviations of array x and array y respectively.

The sample covariance was estimated from Equation 3.2.

$$s^2{}_{x,y} = \sum \frac{(x_i - \bar{x})(y_i - \bar{y})}{n-1} \text{--------------------} 3.2.$$

Where n refers to the number of data points in the sample. The results of ($r_{x,y}$) for the 18 rainfall stations are presented in Table 5.1 in Section 5.

From this table it was possible to identify stations that had significantly linearly related data to those of the station with missing data. From the linearly related stations reference base stations were identified to fill in the missing records. The base stations were taken to be those that had data corresponding to the missing data records. These are the stations that had values of $r_{x,y} \geq 0.5$ in relation to the station with missing records.

The threshold value of $r_{x,y} \geq 0.5$ was taken on the basis that values beyond 0.5 are closer to 1, the value that signifies perfect relationship. Values below 0.5 are closer to zero, the no relation value. A value of $r_{x,y}$ near or equal to 0 implies little or no linear relationship between x and y. On the other hand the closer $r_{x,y}$ comes to 1 or -1, the stronger the linear relationship between x and y. Positive values of $r_{x,y}$ imply a positive linear direct relationship between y and x; that is y increases as x increases. Negative values of $r_{x,y}$ imply a indirect linear relationship between y and x; that is y decreases as x increases (McClave and

Sincich, 2000). Values beyond 0.5 may therefore be considered significant while those below 0.5 may be considered insignificant.

By associating stations that show significant correlation of $r_{x,y} \geq 0.5$, the missing data records were estimated through linear regression. We define a mathematical relation between data from the subject station and other correlated stations by Equation 3.3.

$$y = a + \sum_{i=1}^{n} a_i x_i \text{ -----------------------------------3.3.}$$

Where y refers to the data series from station with missing data, x_i refers to the data series from reference stations, a_i refers to the coefficient of correlation of respective stations and it determines the contribution of x_i to the value of y, and a refers to the y-intercept or the constant of the equation.

This is a multiple regression equation, which can be used to predict values of y using known values of x from more than one reference station. Details of this may be found in any standard statistics textbook such as (McClave and Sincich, 2000, Frank and Althoen, 1994, Thiessen, 1997 and Moore, 1974, among others). By choosing the station that is most correlated with the station with the missing data Equation 3.3 may be reduced to a simple linear regression one of the form:

$$\hat{y} = \hat{\alpha} + \hat{\beta}x \text{ --3.4.}$$

Where $\hat{y}$ refers to the expected value of the missing data, $\hat{\alpha}$ refers to the y-intercept of the regression equation, $\hat{\beta}$ refers to the slope of the regression equation and x refers to the corresponding data from the station with high correlation.

In this study, both multiple and linear regression methods were used. The linear regression was used where only one station showed a strong linear relationship compared with the other stations. Multiple regression was used where more than one station showed correlation coefficients of about the same value. The multiple regression method was limited to three variables in order to avoid over-fitting the model. Equations 3.3 and 3.4 were used to estimate the missing data records by regression analysis.

Alternatively where only one station was to be used, a double mass curve was evaluated for the reference station, the station with corresponding data sets, and the subject station, the station with missing data sets. It has been found that the use of double mass

curves gives very reliable correlation relationships between adjacent stations (LVEMP, 2002). When a trend line is fitted to the mass curve, its equation may be used to fill in as many gaps as possible in the subject station using corresponding data from the reference station. The major shortcoming to this method is that only one station is used to fill in the missing records. If the stations are "too far apart" the results may not be representative. Too far apart here refers to the case where the subject station is in a different rainfall zone from the reference station. The method is also based on inter-station correlation because the gradient of the trend line of the mass curve is a measure of the linear relationship between the two stations.

The double mass curve method was used extensively in this study especially where two adjacent stations showed a high degree of correlation. It is important to note that both the regression method and the double mass curve method rely heavily on the linear correlation analysis as a way of identifying the related stations. Equation 3.5 was used to estimate the missing data records by double mass curve method.

$$y = ax + b \text{--3.5.}$$

Where y refers to the missing record series, x refers to the corresponding record series in reference station, a refers to the gradient of the trend line and b refers to the intercept of trend line.

If the equation is adjusted so that $b = 0$ we get:

$$y = Px \text{--3.6.}$$

Where P refers to the gradient of the trend line through the origin.
Equation 3.6 may be interpreted to mean that data set y is P times the corresponding data set in array x. Therefore to fill in the missing data records in y, corresponding data records in x were multiplied by P before being used in y.

3.2.2 Evaporation Data

Methods used to estimate evaporation records include: the water budget, the energy budget, the mass transfer techniques, and use of evaporation pans. Of all these methods, the pan method is the least expensive and is also known to provide good estimates of evaporation records (Ward and Robinson, 1990).

The methods used to fill in missing evaporation records are in principle the same as those used to fill in the rainfall data records. However there are practical differences caused mainly by the unique nature of evaporation records as described below.

(i)   There are fewer evaporation stations since evaporation is only recorded at agro meteorological or synoptic stations. These are few and far apart.

(ii)   The use of correlation between adjacent stations is usually not relevant since the stations are very far apart. It would therefore be inappropriate to use correlation to extend the records.

(iii)   There is no well-defined method of choosing evaporation for wet, average and dry years. Although the potential evaporation may vary significantly from one day to the other, annual evaporation varies only a little.

In addition to the above, little is known about the local spatial variability of evaporation although it is known to be higher at the equator than at the poles. It is however less variable than either rainfall or runoff. In this case, a station network of much lesser density is required for a correct assessment of the evaporation. For general purpose and preliminary estimation of evaporation, a density of one station per 5000 square kilometres is considered sufficient (Ponce, 1989). Evaporation data from Kisumu Meteorological station were applied to the Nyando catchment since the station and the catchment are in the same climatic zone. Kericho records were only used for comparison purposes.

Missing evaporation records are normally estimated using the seasonal mean method. The seasonal mean is defined as the mean of the evaporation values on date d of the calibration period. Equation 3.7 gives the seasonal mean value.

$$\hat{y}_d = \frac{1}{L} \sum_{r=1}^{L} y_{d,r} \text{ ---3.7.}$$

In Equation 3.7 $y_{d,r}$ refers to the observed evaporation on the date d in year r, L refers to the number of years in the calibration period and $\hat{y}_d$ refers to the seasonal mean value of evaporation on date d.

The mean of all the measurements on a day such as 1st January were used to fill in all the gaps that appeared on 1st January of any year during the period under consideration.

3.3 Estimation of Areal Rainfall

Since the rainfall measurements obtained from rainfall stations are essentially point measurements, there is need to estimate the total rainfall over an area from the original point measurements. From the many point measurements over the catchment area, the total quantity of water falling over the catchment is evaluated. This is expressed as a mean depth (mm) over the catchment area. This value is called Areal Rainfall. It is a value representing the catchment rainfall and is usually referred to as lumped input where the entire catchment is considered as a single unit with uniform properties. The lumped input concept offers a pragmatic approach in which initiative, skill and experience can be used to assess model behaviour. Areal rainfall is a better representation of the catchment rainfall than point observations. It is therefore used in many hydrological studies.

There are many ways of deriving the areal rainfall over a catchment from rain gauge measurements. They include: Arithmetic mean method, thiessen polygon method, isohyetal method, isopercentile method, Chidley and Keys method, hypsometric method, among others (Shaw 1988). Details of how each method works may be obtained from Shaw (1988) or any of the engineering hydrology books available. The choice of a particular method solely depends on the characteristics of the catchment. In this study arithmetic mean method was used to derive daily average areal rainfall. The method was chosen because it is easy to apply and the results obtained are sufficiently accurate.

3.3.1 Arithmetic Mean Method

The arithmetic mean method was used to areally average the rainfall records in this research on account of the spatial homogeneity of the records. This is the simplest method of calculating areal rainfall. Though the method does not take care of irregular distribution of rainfall throughout the catchment, as does the isohyetal method, it gives sufficiently accurate results when estimating rainfall for long periods such as months, seasons and years (WMO, 1975). The simultaneous point measurements of daily rainfall for a selected duration at all the representative gauges are summed and the total divided by the number of gauges used. The rainfall stations used in the calculation are usually those inside the catchment area, but neighbouring gauges outside the boundary may be included if it is considered that the measurements are representative of the nearby parts of the catchment. In this study only gauges inside the catchment were used. Equation 3.8 gives the areal rainfall.

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$ --3.8.

Where n refers to the number of rain gauge stations in the catchment, x_i refers to the daily rainfall recorded by i^{th} rain gauge and $\bar{x}$ refers to the areal rainfall.

3.4 Data Quality Control

In this section methods used to check the quality of data used in this study are presented. Data quality is a crucial component in any research. It is therefore important to subject data to various quality control checks before carrying out any research in order to establish their quality. One of the basic quality control checks is visual examination of raw and plotted data in order to identify erroneous data. Other methods include basic statistical analysis and homogeneity tests.

3.4.1 Identification Of Spurious Data From The Time Series Plots

Due to the large amount of data involved, it was not easy to visually examine the raw data for the purpose of identifying erroneous ones. A time series for daily data was therefore plotted per year for each of the eighteen rainfall stations in order to identify outliers and erroneous data. From the time series it was suspected that some stations had some erroneous record values. Since not all outlier values are erroneous, values were declared erroneous if they were not comparable to corresponding values in the neighbouring stations. The erroneous values were deleted and then estimated using Equation 3.3 whose details are given in section 3.2.1.

The same was done for flow gauging station number 1GD03 as well as for the two evaporation stations.

3.4.2 Basic Statistical Analysis

Many statistical analyses assume that the sample of data set used for any study has a frequency distribution close to normal. It is therefore important to have a good knowledge of the nature of the distribution of the data set before carrying out any study.

Basic statistical analysis was thus carried out on annual rainfall, runoff and evaporation data in order to establish the frequency distribution and variability of data in the catchment. Measures of central tendency and variability were computed for rainfall, runoff

and evaporation records from all the selected stations. This gave a preliminary idea of the temporal and spatial variability of data from the selected stations

3.4.2.1 Rainfall Data

Basic statistical analysis was carried out for each of the selected stations in order to establish the mean, standard deviation, standard error, and coefficients of variation, skewness and kurtosis. This was done on annual rainfall data and therefore annual values had to be calculated. Results of annual data statistics are presented in Table 5.4 and in Figure 5.1 in Section 5.2.1.

3.4.2.2 Runoff Data

Basic statistical analysis was carried out on data from the 8 flow gauging stations that were chosen. The results are presented in Table 5.5 and in Figure 5.2 in Section 5.2.2.

3.4.2.3 Evaporation Data

Two evaporation stations were selected. Basic statistical analysis was carried out for each of them; the results are presented in Table 5.6 and in Figure 5.3 in Section 5.2.3.

3.4.3 Test For The Consistency Of Data

Many statistical analyses assume that data being used is homogeneous. This means that the data is a sample from a single population. Data from different populations are considered inconsistent and may not produce satisfactory results when used in a research. Establishment of the consistency of data is therefore a vital part of any research if the research is to be meaningful. This section presents the methods used to test data consistency and where inconsistency was found, the methods that were applied for the correction.

3.4.3.1 Rainfall Data

One of the main sources of error in rainfall records is the location of the gauge in relation to obstructing objects such as trees and buildings. In an attempt to avoid such errors, gauges are normally relocated. The movement of the gauges affects the consistency of the records from the gauge, where records before and after the gauge relocation are statistically different. The errors may also result from a marked change in the environment of

the neighbourhood of the gauge location due to calamities such as earthquakes, landslides, fires, etc.

Another source of inconsistency in records is usually a change in observational procedure such as a change from an old gauge to a modern one. Data measured using the modern gauges are usually statistically different from those measured using the old gauge. Statistically the two sets of data come from two different populations. Inconsistent data are not homogeneous and as such they cannot be subjected to any study as a single data set. There is need, therefore to ensure that data are homogeneous before carrying out any study using the data as the input.

Mass and double mass curve analyses are the two most popular methods in hydrology used for testing and correcting data for any inconsistency (Ogallo, 1981). Other methods that have been used to test the inconsistency of data include: the Von Neuman ratio test, cumulative deviations test, the likelihood ratio test and the runs test (Ward, 1990). The mass curve analysis method was used in this study for the purpose of testing data inconsistency. The method is easy to use and gives satisfactory results. The method involves plotting cumulative values of actual records (y-axis) in a station at each time step (x-axis). In this study daily values of rainfall were plotted at time steps of one day.

A single straight-line graph is always obtained for homogeneous records suggesting that each of the recorded data values comes from the same parent population. This means that there has not been any change in measurement technique, gauge location or the environment of the location of the gauge. If data is heterogeneous, there will always be a marked deviation from the original straight line leading to the appearance of two or more distinct straight lines. A break in the slope of the original line signifies inconsistency. This is an indication of heterogeneity, which is normally detected from the significant and persistent deviations of some points from the straight line beginning at a particular instant. These deviations start from a particular year when a change in the observational procedure occurred, leading to the appearance of two obvious straight lines in the plotted data starting from two different positions. Using the ratio of the gradients of the two straight lines obtained, the inconsistency can be corrected so that the data now appears to have come from the same parent population. Adjusting the records prior to the break in slope to reflect the new situation performs the correction. To do this, rainfall records prior to the break are multiplied by the ratio of the slopes after and before the break. Details on how the method is applied can be obtained from Ogallo (1981, 1987) and WMO (1970, 1980).

34

In this study daily rainfall data from each of the eighteen selected stations was subjected to homogeneity test through mass curve method before areal average was calculated.

First all the erroneous data were cleaned out and replaced with their corresponding estimated data at each station. A daily mass curve was then plotted for each station for the period 1985 to 1993. The procedure involved plotting cumulative daily rainfall values (y-axis) at time steps of one day (x-axis). A trend line was then fitted for each mass curve and its R^2 value evaluated. This is a value that measures the extent to which the plotted data sets are related.

The results of double mass curve plots, from some selected stations, are presented in Table 5.7 and in Figure 5.4 in Section 5.3.1. A mass curve for areal average daily rainfall data was also plotted. The aim was to test whether the daily average areal rainfall data was also homogeneous. The results are presented in Figure 5.5.

3.4.3.2 Runoff Data

Unlike rainfall, runoff data are continuous. Determinants of data quality are therefore slightly different from those of rainfall data. The main determinants of the quality of stream flow data are accuracy of measurement, continuity of data and the length of record at the gauging station. Quality control of discharge data is done in order to remove outliers and erroneous data. Visual inspection of a time series is deemed sufficient for this process. Station 1GD03 had data from 1969 to 1997. Though it has also had the problem of missing data, it has nearly continuous data for a period of eight years, 1983 to 1990 and only 118 days with missing data for a period of thirteen years. This is only about 1.8 % of missing data between 1983 and 1995. Basic statistical analysis also showed that station number 1GD03 had better quality data than all the other flow gauging stations.

The data from station 1GD03 were subjected to the requisite quality control checks in order to remove outliers and erroneous data. They included plotting a time series of raw data. Where outliers and erroneous data were found, they were deleted and then estimated using Equation 3.6 whose details are given in Section 3.2.1.

A mass curve of daily discharge from gauging station 1GD03 was then plotted for the purpose of testing the consistency of daily runoff data. The results are presented in Figure 5.6 in Section 5.3.1. A trend line was then fitted to the mass curve and was observed to agree with the daily average discharge data.

3.4.3.3 Evaporation Data

One of the sources of errors in evaporation records is the location of the evaporation pan in relation to obstructing objects such as trees and building structures. To avoid the errors introduced by the shielding effects of these obstacles, pans are normally relocated. This brings about inconsistency in data records before and after relocation.

Another source of error is the day-to-day weather variations, such as sunny, windy and cloudy weathers, which greatly affect the rate of evaporation. Evaporation records are also affected by the surface over which evaporation takes place. It has been observed that evaporation from a particular surface is directly related to the opportunity for evaporation (availability of water for evaporation) provided by that surface (Viessman and Lewis, 1994). For an open surface of water, the evaporation opportunity is 100 percent, while for soils it varies from a high of 100 percent when the soil is highly saturated to almost zero percent at stages of very low moisture content. Other types of surface provide diverse degrees of evaporation opportunity and these normally vary widely with time. It should therefore be expected that changes in land cover brought about by diverse land use methods are likely to bring about inconsistency in evaporation records where evaporation records before and after the changes are statistically different.

To establish consistency of evaporation data, a single mass curve was plotted for both Kisumu and Kericho stations. Cumulative daily mean values of evaporation were plotted. The results of the Kisumu station are presented in Figures 5.7 in Section 5.3.1.

The quality controlled daily areal average rainfall; daily average discharge and evaporation data were then split into two independent sets. It is common practice in rainfall-runoff modeling to split the available data into two independent periods, namely the calibration and the verification periods. The calibration period is devoted to estimating the appropriate values of the model parameters and weighting series. The verification period is used to obtain an unbiased indication of the consistency in performance, which can be expected from the model when it is applied to data other than those used for calibration.

4.0 METHODS OF ANALYSIS

In this section a brief discussion of hypothesis, test of hypothesis, and model efficiency criteria are presented.

4.1 Hypothesis

Nyando catchment can be modelled using input-output linear systems models. In particular the Simple Linear Model (SLM), the Linearly Varying Gain Factor Model (LVGFM) and the Linear Perturbation Model (LPM) are capable of modelling the catchment at different levels of accuracy. A conceptual model and in particular the Soil Moisture Accounting and Routing (SMAR) model would be expected to perform even better than the linear systems models given that it includes evaporation as an extra model input in addition to the fact that it is more physically based than the linear systems models.

This study looks at the performance of each of the three linear systems models and the conceptual model on data from Nyando catchment under the above stated hypothesis. The rainfall-runoff models, which were used, are the SLM, the LVGFM, the LPM and the SMAR model from among many others. Details of these models are discussed in the following subsections.

4.1.1 Linear Systems Models

Linear systems models are those models that are formulated in terms of linear equations and processes. In nature however, physical processes are generally non-linear. But linear models are often substituted in the interest of mathematical expediency in order to simplify the otherwise complicated physical processes. The simplicity of the linear systems models is a definite advantage although it is usually achieved at the cost of a certain loss of detail in the real process.

These models presume that rainfall is the cause of runoff. The models make no attempt to account for the physical nature of transformation of rainfall into runoff. The data requirements for calibration are only rainfall as input and runoff as output. Thereafter rainfall is used to forecast runoff. Though simple in nature, these models have been shown to perform equally well as the higher order physically based models. They have the advantage that their data requirements are readily available.

Linearity of these models is based on the fact that output depends only on input and not the time at which the input is applied. For a linear system in which the cause precedes the effect and in which no input or output occurs prior to time t = 0, the relationship of input-output and impulse response is expressed by the convolution integral:

$$y(t) = \int_0^t x(t)h(t-\tau)d\tau \text{ --------------------------4.1.}$$

Where $h(t)$ refers to the unit impulse response function or instantaneous unit hydrograph (IUH) ordinate at time t, $x(t)$ refers to the input at time t, $y(t)$ refers to the output after time t, τ refers to the time into the past so that $t - \tau$ occurs before t. It varies between the past and present time so that: $\tau = 0$ is in the past and $t = \tau$ ($t - \tau = 0$) is in the present.

The impulse response function $h(t)$ represents the relative effect of antecedent rainfall intensity on the discharge rate at any instant. It describes the characteristic of the catchment in terms of how it responds to rainfall. It is expected that $h(t)$ values should decrease monotonically towards zero as we approach the memory length. The integral gives a continuous weighting of previous rainfall intensities by the ordinates of the instantaneous unit hydrograph (IUH).

When the input function is expressed as a series of pulses or mean values over successive short time intervals T, the response to a unit pulse of duration T may be a more convenient expression of the operation of the system than the impulse response. Equation 4.2 gives the input-output relationship expression in terms of the pulse response

$$y_i = \sum_{j=1}^{m} x_{i-j+1}h_j \text{ -----------------------------------4.2.}$$

Where m refers to the Memory length of the system; the interval between the occurrence of rainfall and the time when its effect on the stream flow finally ceases or simply the number of pulse response ordinates, y_i refers to the ith output series and h_j refers to the jth ordinate of pulse response.

Equation 4.3 can use the impulse response function to obtain the pulse response function for any specified duration T.

$$h(T,t) = \frac{1}{T} \int_{t-T}^{t} h(t)dt \text{ -----------------------------4.3.}$$

Where $h(t)$ refers to the impulse response function, $h(T,t)$ refers to the pulse response function for duration T.

In this study the performance of three linear systems models and one conceptual model were assessed for the data collected from Nyando river catchment. These models are the Simple Linear Model (SLM), the Linear Perturbation Model (LPM), the Linearly Varying Gain Factor Model (LVGFM) and the Soil Moisture Accounting and Routing (SMAR) model from among many others.

4.1.1.1 The Simple Linear Model (SLM)

The intrinsic hypothesis of the primitive SLM, introduced by Nash and Foley (1982), is the assumption of a simple linear time-invariant relationship between the total input x_i, the areal mean rainfall, and the total output y_i, the discharge recorded at the gauging station.

The SLM is the simplest of all the linear systems models and usually serves as a convenient starting point in rainfall-runoff modeling. Nash and Sutcliffe (1982) introduced the model as a basis of efficiency comparison with other more elaborate models. It was never meant to be a substantive model in its own right. However, although the model is considered primitive, it performs quite satisfactorily in catchments that exhibit near perfect linear relationship between total rainfall and total discharge. These are catchments where a plot of cumulative rainfall and discharge data series shows that there is near perfect linear relationship between the two series. SLM would therefore be expected to yield results that are as good as those from models that are considered more substantive in such catchments.

From the model's postulates of a linear, time invariant relationship between the total rainfall and the total discharge, it is considered as a quick means of assessing the extent of linearity existing in the rainfall-runoff relationship in a catchment. By analyzing the deficiencies of SLM on a particular catchment, it is possible to apply a more realistic and elaborate model that is more responsive to the catchment's hydrologic characteristics. This is usually done following the principle of progressive modification from the simple to the more complex as suggested by Nash and Sutcliffe (1970). This study aims at identifying a suitable model for the Nyando catchment following this principle.

39

The convolution summation relation of the form shown below expresses the SLM in its discrete form (Kachroo and Liang, 1991):

$$y_i = \sum_{i=1}^{m} x_{i-j+1} h_j + e_i \text{-----------------------------------4.4.}$$

Where h_j refers to the jth ordinate of the pulse response function, e_i refers to the ith forecast error term, x_i refers to the ith rainfall input series, y_i refers to the ith runoff output series and m refers to the memory length of the system which implies that the effect of any input x lasts only through m intervals of duration T.

Daily areal averaged rainfall and daily runoff data were used to calibrate the model. The runoff data are the daily averaged river flow depths recorded at the outlet of the catchment.

The SLM is normally considered a convenient starting point in rainfall runoff modeling. It is therefore included in the Galway Flow Forecasting System (GFFS) for use as a naïve model for the purpose of efficiency comparison with other models. This is a windows based software package for flow forecasting developed at the University College Galway. Any model that does not perform better than SLM may hardly be considered as a serious rainfall-runoff model.

The model was applied, after calibration, on quality-controlled data from Nyando river catchment. The efficiency of the model was tested using the Nash and Sutcliffe criterion given in Equation 4.44 whose details are given in Section 4.4.1.

4.1.1.2 The Linear Perturbation Model (LPM)

This model lays more emphasis on the observed seasonal behaviour inherent in the observed rainfall and discharge data series. It is based on the assumption that there is normally a linear relationship between departures from the seasonal expectations of input data series (X- X_d) and departures from the seasonal expectations of output series (Y- Y_d). It actually attempts to accommodate the fact that rainfall-runoff relationship is time dependent in the long term due to the seasonal behaviour of both rainfall and runoff.

Nash and Barsi (1983) originally proposed the LPM as a 'Hybrid Model' for flow forecasting on large seasonal catchments. The model attempts to incorporate and utilise the seasonal information of the actual observed rainfall and discharge time series in input-output models.

For the discrete system with recorded data sampled at one day interval or averaged over one day interval, the discrete LPM works on the assumption that, during a year in which the rainfall is identical to its seasonal expectation, X_d, the corresponding discharge is also identical to its seasonal expectation Y_d. However, in all the other years, when the rainfall and the discharge values depart from their respective seasonal expectations, these departure series, $(X - X_d)$ and $(Y - Y_d)$, are assumed to be related by a linear time-invariant system.

The model input consists of two parts namely: the series of the seasonal expectations of daily rainfall ($_sX_i$), which is transformed to the series of the seasonal expectations of daily discharge ($_sQ_i$) through an undefined relation. The second part is the remainder of the input ($X'_i = X_i - _sX_i$), which is routed through a linear system. The total output of the LPM is then the arithmetic sum of the seasonal expectations of the discharge ($_sQ_i$) and the output ($X'_i = X_i - _sX_i$) of the linear system. This is illustrated in Figure 4.1 (Kachroo et al, 1988).
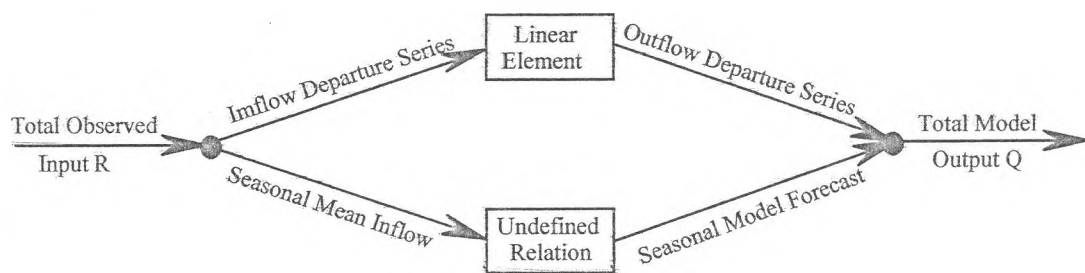
Figure 4.1: Schematic representation of the Linear Perturbation Model (LPM).
Source: (Kachroo et al, 1988)

This model is considered suitable for catchments that exhibit marked seasonal variations as it involves the assumption of a linear relationship between departures from seasonal expectations of both input and output series. The relation between the departure series of the LPM has the convolution summation of the form:

$$Q_i = \sum_{j=1}^{m} R_{i-j+1} h_j + e_i \text{ -------------------------4.5.}$$

In Equation 4.5, Q_i refers to $Y_i - Y_d$, the departure series of outflow, Y_i, from its own seasonal mean Y_d, R_i refers to $X_i - X_d$, the departure series of inflow (or rainfall) X_i, from its seasonal mean X_d, m refers to the memory length with $j = 1, 2, 3... m$, h_j refers to the ith ordinate of

the discrete pulse response relating the departure series of input and output, e_i refers to the ith output error term, d refers to a particular day of the year so that d = 1,2...365 and n refers to the number of output ordinates so that l = 1, 2, 3, ..., n.

On a catchment with highly seasonal variation of discharge, subtraction of seasonal means from the original series removes much of the dependence on linearity. The emphasis is now shifted to the seasonal variation through the assumption of a linear relationship between departures from the seasonal means. This assumption is less restrictive than the same assumption about the total actual input and output values (Kachroo and Liang, 1991). For a catchment that exhibits less seasonal behaviour, the performance of LPM would be of the same order of efficiency as that of the SLM.

The model was applied, after calibration, to quality-controlled data from Nyando river catchment. The efficiency of the model was tested using the Nash and Sutcliffe criterion given in Equation 4.44. Equation 4.47 was used to test for the efficiency of this model against the SLM. Details of these equations are given in Section 4.4.1.

4.1.1.3   Linearly Varying Gain Factor Model (LVGFM)

This model was first proposed by Ahsan and O'Connor (1994) for the single input to single output case. It is based on the variation of the gain factor with the selected index of *the prevailing catchment wetness index. It may also be regarded as a wetness-index-based* model.

Surface runoff is directly related to effective rainfall, which is inversely related to the hydrologic abstractions. During rainy periods, infiltration plays a major role in abstracting total rainfall. Actual infiltration rates and amounts vary widely both in time and in space, being highly dependent on the initial level of the soil moisture. For any given storm, the initial level of the soil moisture is referred to as the antecedent moisture.

Catchments with low initial moisture levels are said to be dry and are not conducive for high runoff response. Conversely, catchments with high initial moisture levels are said to be wet and are likely to produce large quantities of runoff. Direct runoff is therefore a function of antecedent moisture. The average moisture level in a catchment varies daily, being replenished by rainfall and depleted by evaporation and percolation. The main assumption of the LVGFM model is that the amount of rainfall that transforms to runoff is a function of the state of the soil moisture.

Depending on the soil moisture index, all the rainfall or part of it may be transformed to runoff. This model is an improvement of the SLM. It is reasonable to expect that the wetter the catchment, the higher the proportion of rainfall that transforms into runoff. Since SLM does not accommodate this concept, it should be expected that the model would over estimate the low magnitude flows and under estimate the flows of relatively high magnitudes.

The gain factor is defined as the ratio of the total output flow volume to the total input rainfall volume, during the calibration period, when both are expressed in the same units. The gain factor accounts for the change in the proportion of rainfall that transforms into runoff with the change in the soil moisture content. It can be more than or less than unity depending on whether the output is more or less than unity. The gain factor is a measure of the proportion of rainfall that on average transforms into runoff. This value is in reality not constant. It varies with the magnitude of the soil moisture content, which may not be a simple function of time and which may explicitly depend on the input and the output of the system. The proportion of rainfall that transforms to runoff therefore varies with the wetness index. The higher the wetness index the more the rainfall that transforms to runoff. The SLM does not account for the variation of the gain factor. This is a major deficiency of the SLM and a model that accommodates the variation of the gain factor would therefore be preferred. One such model is the linearly varying gain factor model, LVGFM, which seeks to address the constant gain factor deficiency of the SLM by incorporating a response function which varies gradually with the catchment wetness, both in scale G, the gain factor, and in shape. The SLM is used as the auxiliary model of the LVGFM to generate the catchment wetness index, z_t. It is prudent to expect this model to give better results than the SLM.

Using a time varying gain factor G_i the model output has the structure of the form:

$$y_i = G_i \sum_{j=1}^{m} w_j x_{i-j+1} + e_i \text{------------------4.6.}$$

With $\sum_{j=1}^{m} w_i = 1$

Where x_i refers to the magnitude of the pulsed input during the ith time interval, y_i refers to the magnitude of the pulsed output at the end of the ith time interval; G_i refers to the magnitude of the gain factor at the end of the ith time interval. This is the time varying coefficient of runoff that accounts for the short-term time dependence of runoff, which is due

to the effect of variable soil moisture conditions. w_j refers to the jth ordinate of the non-parametric discrete weighting function, m refers to the memory length of the system and e_i refers to the ith output error term.

In its simplest form, G_i is linearly related to an index of the soil moisture state z_i by the equation:

$$G_i = a + bz_i \text{---} 4.7a.$$

Where a and b are constants with the upper limit of G_i restricted to unity. With parameter a in Equation 4.7a above set to a = 0, we get:

$$G_i = bz_i \text{--} 4.7b.$$

Where b is a parameter of the LVGFM, to be determined. The value of z_i is obtained from the outputs of the primitive SLM, operating as the auxiliary model, by the equation:

$$z_i = \frac{\hat{y}_{a,i}}{\bar{y}} = \frac{G_a}{\bar{y}} \sum_{j=1}^{m_a} \hat{h}_j x_{i-j+1} \qquad \text{------------} 4.8.$$

Where $\hat{y}_{a,i}$ refers to the simulated output of an optimally calibrated discrete SLM, $\hat{h}_j$ refers to the estimate of the standardized pulse response of the auxiliary SLM, G_a refers to the estimate of the standardized gain factor of the auxiliary SLM, $\bar{y}$ refers to the mean discharge in the calibration period and m_a refers to the assumed memory length of the auxiliary SLM.
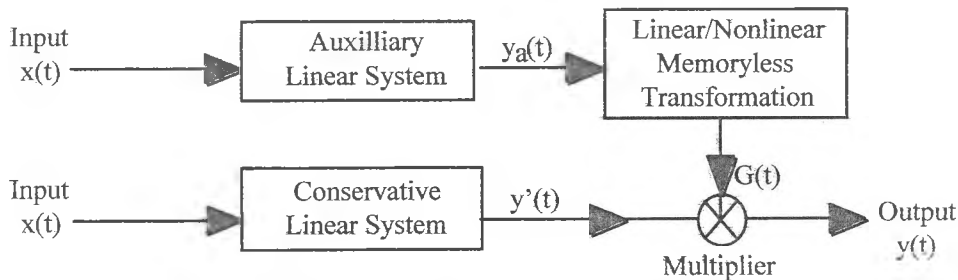
Figure 4.2: Schematic representation of the structure of the Linearly Varying Gain Factor Model, LVGFM. Source: (Ahsan and O'Connor, 1994).

The parameters of the LVGFM can now be estimated by the direct linear least squares regression method as described elsewhere in this chapter. Figure 4.2 shows a representation of LVGFM

The efficiency of the model was tested using the Nash and Sutcliffe criterion given in equation 4.44. Equation 4.47 was used to test the efficiency of this model against that of the SLM.

4.3 Conceptual Soil Moisture Accounting and Routing (SMAR) Model

Conceptual models are based on simplified descriptions of water movement in a catchment both in time and space, on its surface and through its sub-surface, accounting for the water storage and movement in discrete but relatively short time intervals. This involves complex systems of equations based on theoretical or physical concepts governing the hydrologic processes of a particular catchment (Opere, 1991).

The SMAR model is a lumped quasi-physical conceptual rainfall-evaporation-runoff model, with quite distinct water-balance and routing components that influence discharge in a catchment. This is achieved through a series of steps, which represent some of the known physical processes in a hydrological system.

The Soil Moisture Accounting and Routing (SMAR) model is a development of the 'Layers' conceptual rainfall-runoff model introduced by O'Connell et al. (1970), its water-balance component being based on the 'Layers Water Balance Model' proposed in 1969 by Nash and Sutcliffe (1970). Using a number of empirical and assumed relations which are considered to be at least physically plausible, the non-linear water balance component, the soil moisture accounting, ensures satisfaction of the continuity equation over each time-step by preserving the balance between the rainfall, the evaporation, the generated runoff and the changes in the various elements (layers) of soil moisture storage.

The routing component, on the other hand, simulates the attenuation and the diffusive effects of the catchment by routing the various generated runoff components, the outputs from the water balance component, through linear time-invariant storage elements. For each time-step, the combined output of the two routing elements adopted here, the one for generated 'surface runoff' and the other for generated 'groundwater runoff', becomes the simulated discharge forecast produced by the SMAR model (Liang, 1992). This linear operation is expressed by the equation of the form:

$$y_t = \sum_{j=1}^{m} x_{i-j+1} hj + e_t \quad \text{-------------------------------------- 4.9.}$$

Where h_j refers to the pulse response ordinates, which uniquely express the transformation of the system. This is to be solved for Nyando. y_i refers to the estimated discharge, x_i refers to the generated runoff, e_i refers to the output error term and m refers to the memory length of the system.

The optimum response series h_i can be determined directly by the method of ordinary least squares for the given series of x and y. Equation 4.9 may then be reformulated as:

$$y_i = G \sum_{j=1}^{m} x_{i-j+1} u_j + e_t \quad \text{--------------------------------4.10.}$$

Where u_i refers to the standardized pulse response series whose sum is 1 and G refers to the gain factor and is equal to the sum of the ordinates h_j.

The version of the SMAR used in the present study has nine parameters, five of which control the overall operation of the water-budget component, while the remaining four parameters control the operation of the routing component. The SMAR is calibrated to the observed data using the user's choice optimisation procedure to minimise the selected measure of error between the observed and the model estimated discharges. In the context of the SMAR model, the selected measure of model error used for this study was a weighted combination of the sum of squares of the discharge forecast errors and the corresponding index of volumetric fit, the ratio of the total volume of the estimated discharge hydrograph to that of the corresponding observed hydrograph. As the Nash-Sutcliffe (1970) model efficiency criterion R^2 is based on the sum of squares of model forecast errors only, it follows that the R^2 value obtained for the SMAR model calibrated using the weighted measure of error will generally be less than that for which the index of volumetric fit is neglected in the calibration process, as is the case for the linear models used in this study.

This should be borne in mind when comparing the R^2 model efficiency values of various other models considered in this study with that of the SMAR model. The SMAR model is considered suitable for catchments where the number of rainy days is small and the rate of potential evaporation is high. Semi-arid catchments, where linear models are known to perform poorly, are good examples.

Figure4.3: Schematic representation of the Liang (1992) version of the SMAR Model and a summary description of its parameters; after Shamseldin et al. (1999).
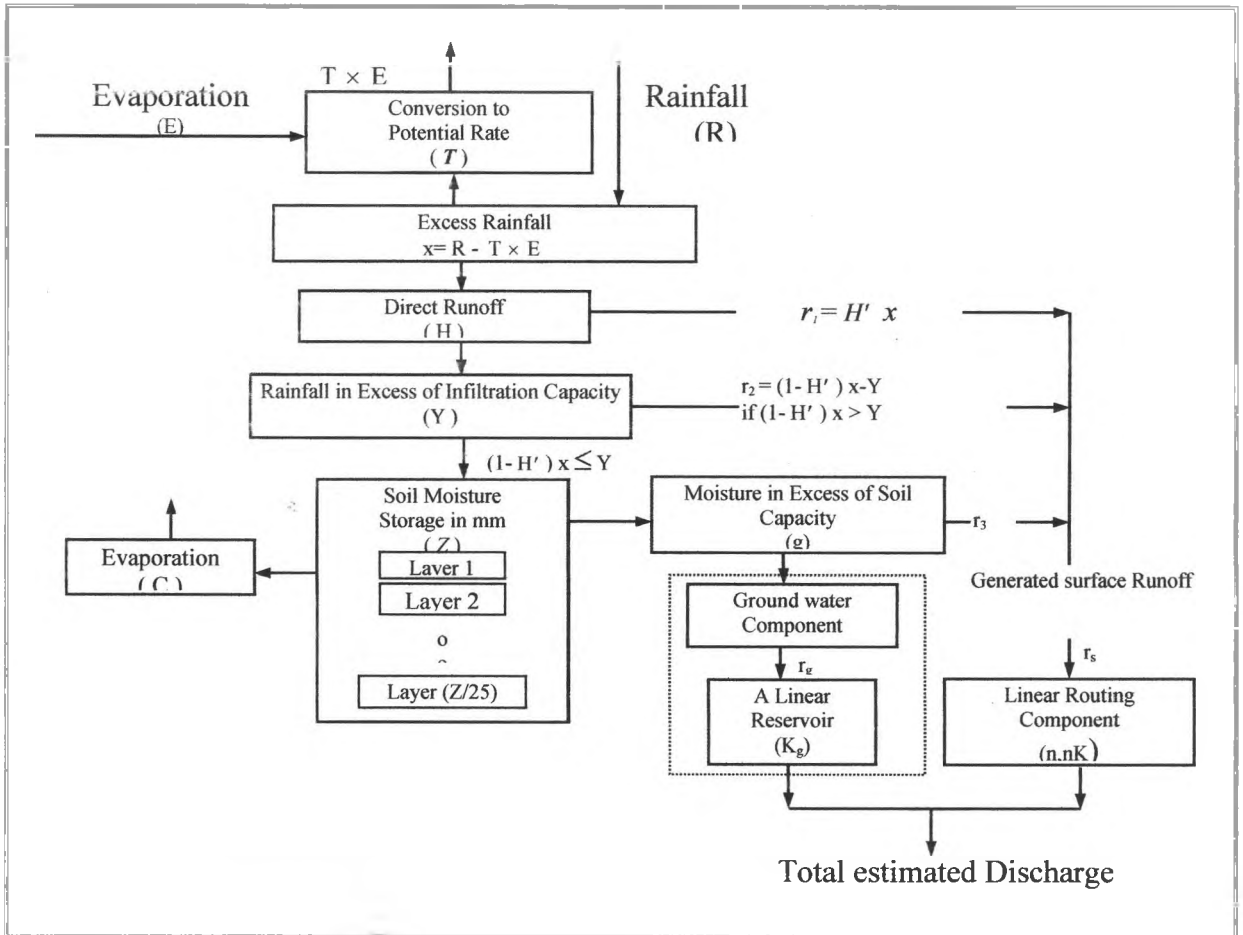
Parameter	Description
Z	The combined water storage depth capacity of the layers (mm)
T	A parameter (less than unity) that converts the given evaporation series to the model-estimated potential evaporation series.
C	The evaporation decay parameter, facilitating lower evaporation rates from the deeper soil moisture storage layers
H	The generated *direct runoff* coefficient
Y	The maximum infiltration capacity depth (mm)
n	The shape parameter of the Nash gamma function *surface runoff* routing element; a routing parameter
nK	The scale (lag) parameter of the Nash gamma function *surface runoff* routing element; a routing parameter
g	The weighting parameter, determining the amount of generated *groundwater* used as input to the *groundwater* routing element.
K_g	The storage coefficient of the *groundwater* (linear reservoir) routing element; a routing parameter

For a humid catchment with a high number of rainy days, a low rate of evaporation and soil moisture almost at field capacity, the SMAR model does not have a particular advantage over the linear systems models (Opere, 1998).

4.4 Test of Hypothesis

Linear systems models are many and varied. It would therefore be difficult to test and compare the performance of all of them. Only the three models described above were selected for this study. Their hypothesis was tested as follows:

To test the linearity of the models, input data was applied at different time intervals and their respective outputs compared. If the models are truly linear the outputs of different models were expected to be of the same order.

The three models selected were calibrated using data from the Nyando catchment. The efficiency of each model was judged on two levels, namely:

(i)   The level of its accuracy where simulated outputs were compared with observed output. This is the calibration phase.

(ii)   The consistency of the level of the model accuracy when subjected to different samples of data. This is the verification phase.

The efficiency of each model was judged through the sum of squares of differences between the observed and simulated discharges. This is the Nash and Sutcliffe criterion. The model with the least sum was judged most efficient. It would therefore be used to forecast the floods in the catchment. The forecasts would therefore be used to plan for evacuation and supply of medicine ahead of the actual flooding as well as to plan for a more rational use of the water from river Nyando. This is expected to reduce the number of casualties in the area besides increasing productivity through rational use of the water.

4.4.1 Models' Calibration

The essential ingredients of any model are variables and parameters. Variables are the physical quantities such as rainfall, discharge or evaporation. Parameters are the quantities that control the behaviour of the variables.

Model calibration is the process by which the values of model parameters are identified for use in a particular application. It involves the solution of the impulse response, *h(t),* series of the hydrologic system. This is a value that gives a relation between the characteristics of a catchment and the response in the storm runoff to a predetermined input

rainfall. Values of pulse response functions are unique to a particular catchment. Two or more catchments would have same values of pulse response functions only if they share similar hydrometeorologic characteristics. Hence it is important to calibrate a given model afresh for use in a new catchment. The process consists of the use of rainfall-runoff data and a procedure to identify the model parameters that provide the best agreement between simulated and recorded discharge values.

Calibration of a model is the demonstration that the model is capable of producing field observations, which are then taken to be the calibration values. Calibration thus involves solving the model equation for the unknown values of pulse response functions using the known values of the input and output from past records in order to see how well the model reproduces some known observations from a catchment. This is like tuning the model to fit the conditions of a particular catchment. Calibration is done using data from the catchment to be modelled. The model calibration is normally carried out for a period when there is simultaneous and continuous data for rainfall, discharge, and evaporation in the case of a conceptual model, for at least six years. The success of model calibration depends upon good quality data.

Since all models are approximations to reality, there is need to calibrate the models for a particular catchment using data from within the catchment. This is done by empirically searching for the pulse response function that best relates the rainfall to the runoff. The process of searching for the optimum pulse response function may be referred to as solving the model. The pulse response converts the rainfall into runoff through the convolution equations: 4.4, 4.5, 4.6 and 4.10 whose details are given in section 4.2.1, 4.2.2, 4.2.3 and 4.3 respectively. The time interval is taken to be one day rather than continuous since in almost all forecasting applications data are available at daily intervals.

$$
\begin{bmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ y_m \\ y_{m+1} \\ \cdot \\ \cdot \\ y_n \end{bmatrix} = \begin{bmatrix} x_1 & 0 & . & . & . & 0 \\ x_2 & x_1 & . & . & . & 0 \\ \cdot & \cdot & . & . & . & \cdot \\ \cdot & \cdot & . & . & . & \cdot \\ x_m & x_{m-1} & . & . & . & \cdot \\ x_{m+1} & x_m & . & . & . & \cdot \\ \cdot & \cdot & . & . & . & \cdot \\ \cdot & \cdot & . & . & . & \cdot \\ x_n & x_{n-1} & . & . & . & x_{n-m+1} \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \\ \cdot \\ \cdot \\ h_m \\ \cdot \\ \cdot \\ \cdot \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \cdot \\ \cdot \\ e_m \\ e_{m+1} \\ \cdot \\ \cdot \\ e_n \end{bmatrix} \quad \text{-----------4.11.}
$$

Where *m* refers to the memory length and n refers to the number of observations of y.

During model calibration, values are adjusted within a predetermined range of uncertainties until the model produces results that best approximate the set of observed values. This is achieved through split sampling where about two thirds of the data are used for calibration and then testing the model consistency by its ability to reproduce the remaining data.

In calibrating single-input linear models, equations 4.4, 4.5, and 4.6 are written out for each output ordinates y_i to y_n to yield *n* linear equations in matrix form as shown in Equation 4.11. Same calibration and verification periods were used for each model.

The above formulation assumes that x values over the memory length prior to x_1 are all zero. The useful values of y begin at y_m, after the end of the memory length; y_m is the transition value from zero values to nonzero values. Equation 4.11 can be written in vector form as:

$$Y = X\hat{H} + E \text{ ---4.12.}$$

Where Y refers to the column vector of the output series (n, 1), X refers to the matrix of the input series (n, m), $\hat{H}$ refers to the column vector of pulse response ordinates or the parameter (m, 1), and E refers to the column vector of the model errors (n, 1).

In identifying the operation of the system, the known values of input and output from past records are used to determine $\hat{H}$ where

$$\hat{H} = (h_1 h_2 h_3 ... h_m)^T \text{ --4.13.}$$

Only after $\hat{H}$ has been determined, will the operation of the system be assessed. Values used to determine $\hat{H}$ are not part of the values used to determine the operation of the system up to the mth value.

The objective here is to minimise the sum of squares of model output errors. In this case the optimum value of $\hat{H}$, is determined directly by the method of ordinary least squares as:

$$\hat{H} = (X^T X)^{-1} X^T Y \text{ ---4.14.}$$

The ordinary least squares method provides the only solution that satisfies the condition of the minimum sum of squares of model residuals.

The model calibration and verification periods are as indicated in Table 4.1.

Table 4.1: Calibration and verification periods of the models for the Nyando catchment.

Calibration period		Verification period	
Full years	Starting date	Full years	Starting date
4	1 January 1985	2	1 January 1988

4.4.1.1 Least Squares Calibration of The Discrete Form of The SLM

The model was calibrated using daily data by ordinary least squares method as described in the preceding section. The memory length (m), i.e., the number of pulse response ordinates, was chosen by trial and error.

In practice it is simpler to calibrate the SLM using discrete formulation. This is necessitated by the fact that real life rainfall and discharge data are obtained in discrete form. The relation between discrete rainfall input x_i and runoff output y_i is given in terms of a discrete pulse response series u_j, by the following discrete convolution summation equation

$$y_t = \sum_{i=1}^{m} u_j x_{t-j+1} + e_i \text{ --4.15.}$$

Where m refers to the memory length of the system and e_i refers to the model error term.

If we have N number of rainfall-runoff data pairs for calibrations, Equation 4.11 can be written out for each of the output ordinates y_i from the data point, m, onwards giving rise to a vector matrix equation of the form:

$$Y = Xu + e \text{---4.16.}$$

Where u refers to the column vector (m,1) having the pulse response ordinates as its elements i.e., $u = (u_1 u_2 ... u_m)^T$, Y refers to the matrix $(y_m y_{m+1}...y_{n-m+1})^T$, e refers to the matrix $(e_m e_{m+1}...e_{n-m+1})^T$, and X refers to the matrix (N+m, N-m+1).

When the data for the calibration period are considerably larger than the memory length of a system, Equation 4.11 represents an over-determined system of linear equations. Thus the vector, u, of the pulse response ordinates can be estimated by the linear least-squares regression technique by minimising an objective function, J, of the form:

$$J = \sum_{i=1}^{N} \left(y_t - \sum_{i=1}^{m} u_{i-j+1} \right)^2 \text{ ---4.17.}$$

Here the first term in the brackets represents the observed flow while the second one represents the estimated flow. This is thus an expression of the error between the observed and the estimated flows. The objective is to minimise this error so that the estimated flow is as close to the observed flow as possible. This is done by intuitive adjustment of the values of u_j. The smaller the value of Equation 4.17, the more confident we are that the estimates are good. The values of u_j for which Equation 4.17 assumes its minimum value, are the least squares estimates of the catchment parameters responsible for the conversion of rainfall to runoff. The process of determining these values is the model calibration. In Equation 4.17, u_j are the pulse response ordinates. Minimising of J with respect to u_j results in the following least squares estimate of u

$$\hat{u} = \left(X^T X\right)^{-1} X^T y \text{ ---4.18.}$$

Where $\hat{u}$ refers to the estimated vector of pulse response ordinates given by:

$$\hat{u} = (h_1 h_2 ... h_m)^T$$

The estimated gain factor corresponding to the SLM pulse response ordinates calibrated by the above described least squares method is given by:

$$G_a = \sum_{j=1}^{m} \hat{u}_j \text{ ---4.19.}$$

The standardised pulse response ordinates of the calibrated SLM can subsequently be obtained from

$$\hat{h} = \frac{\hat{u}}{G_a} \text{ --4.20.}$$

Using this calibrated SLM, the simulation mode flow forecasts, $y_{a,i}$, for a time instant i, can be obtained from

$$\hat{y}_{a,i} = \sum_{j=1}^{m} \hat{u}_j x_{i-j+1} = G_a \sum_{j=1}^{m} \hat{h}_j x_{i-j+1} \text{ --------------------4.21.}$$

With the wetness index series defined by Equation 4.8

Where $\hat{y}_{a,i}$ refers to the simulated outflow of the SLM at the ith series obtained from Equation 4.21 and $\bar{y}$ refers to the Mean flow during the calibration period.

4.4.1.2 Calibration of The LPM

To calibrate the LPM, at least six years of continuous daily data is required for the process. From the model's basic assumption that perturbations from the expected input values are linearly related to the perturbations from the expected output values, the model was calibrated as follows (Kachroo et al, 1991):

The first step was to calculate the seasonal expectation of rainfall X_d and the corresponding seasonal expectation of runoff Y_d for the period of calibration. The following equations were used to obtain the seasonal mean values:

$$\hat{x}_d = \frac{1}{L}\sum_{r=1}^{L} x_{d,r} \text{---4.22.}$$

$$\hat{y}_d = \frac{1}{L}\sum_{r=1}^{L} y_{d,r} \text{---4.23.}$$

Where $\hat{x}_d$ refers to the expected rainfall values on date d, $\hat{y}_d$ refers to the expected discharge values on date d, $x_{d,r}$ refers to the observed rainfall values on date d, year r, $y_{d,r}$ refers to the observed discharge values on date d, year r and L refers to the number of years in the calibration period.

It was observed that as L becomes large, the estimates of y_d and x_d become smoothly varying functions of date. However since L is limited to the calibration period, the values exhibit fluctuation and there is therefore need to smooth out these fluctuations. The smoothing may be smoothed by the method of Fourier analysis (Bird and May, 1979). Equation 4.24 does this through obtaining the Fourier series representations of y_d and x_d.

$$y_d = a_o + \sum_{j=1}^{p} A_j \cos\left(\frac{2\pi j d}{n}\right) + B_j \sin\left(\frac{2\pi j d}{n}\right) \text{---------4.24.}$$

The seasonal mean values were smoothed by the method of Fourier analysis through the removal of small sample fluctuations. This was done through writing a Fourier series representation (Equation 4.24) of the seasonal mean values obtained from Equations 4.22 and 4.23.

Where a_o is the mean of y_d given by Equation 4.25.

$$a_o = \frac{1}{n}\sum_{d=1}^{n} y_d \text{---4.25.}$$

A_j and B_j are the Fourier coefficients given by Equations 4.26 and 4.27.

$$A_j = \frac{2}{n}\sum_{d=1}^{n} y_d \cos\left(\frac{2\pi j d}{n}\right) \quad j = 1, 2, 3, \ldots, p \text{ ---------4.26.}$$

$$B_j = \frac{2}{n}\sum_{d=1}^{n} y_d \sin\left(\frac{2\pi j d}{n}\right) \quad j = 1, 2, 3, \ldots, p \text{ -----------4.27.}$$

j is the order of the harmonic while p is the maximum number of harmonics and n is equal to 365 for daily series.

The relative values of these coefficients represent the location of the harmonic on the time scale. Equation 4.28 gives the variance of unsmoothed seasonal values accounted for by each harmonic.

$$C_j = \frac{1}{2}\left(A_j^2 + B_j^2\right) \text{ --4.28.}$$

This value may be expressed as a percentage, P_j, of the total variance of y_d series about its mean a_o. C_j represents the relative importance of each component harmonic. A total of 182 harmonics would be required to represent y_d at all points. However adequate smoothing is normally achieved by using only the first four harmonics. The estimated Fourier coefficients are presented in Tables 5.8 and 5.9 in Section 5.3.3.

In the second step the smoothed seasonal values were subtracted from the corresponding rainfall and runoff series for the period of calibration to yield the time series of perturbations.

The third step was to estimate the pulse response function for the Nyando catchment using the method of least squares as described earlier in this section.
Equation 4.29 is the vector matrix used for the purpose of calibration.

$$Q = R\hat{H} + E \text{ --4.29.}$$

Where Q refers to the column vector, $(n, 1)$, of output series $(Q_i = Y_i - X_d)$, R refers to the, matrix, (n, m), of the input series $(R_i = X_i - X_d)$, $\hat{H}$ refers to the matrix, $(m, 1)$, of pulse response ordinates. This is the one to be solved where $\hat{H} = (h_1 h_2 h_3 \ldots h_m)^T$ and E refers to the column vector, $(n, 1)$, of model error terms.

$\hat{H}$ is determined directly by the method of least squares through Equation 4.30.

$$\hat{H} = \left(R^T R\right)^{-1} R^T Q \text{ ------------------------------------4.30.}$$

4.4.1.3 Calibration of The LVGFM

Calibration of LVGFM requires substituting for G_i from Equation 4.7b into Equation 4.6 to obtain the following relation

$$y_i = \sum_{j=1}^{m} w_j' r_{i-j+1} + e_i \quad \text{---4.31.}$$

Where $w' = bw_j$ ---4.32a.

$r_{i-j+1} = z_i x_{i-j+1}$ ---4.32b.

z_i refers to the catchment wetness index and is obtained from Equation 4.8.

Equations 4.32a can be solved directly for w_j' elements, where j = 1,2,3...m by the least squares regression method the same way pulse response functions were determined for SLM. In vector matrix form the equation may be written as

$$\mathbf{y} = \mathbf{R}\mathbf{w'} + \mathbf{e} \quad \text{---4.33.}$$

Where **y** refers to the matrix $(y_m y_{m+1}...y_N)^T$, **w'** refers to the matrix $(w_1' w_2'...w_m')^T$, **e** refers to the matrix $(e_m e_{m+1}...e_N)^T$ and **R** refers to the matrix (m, N-m+1).

A least squares estimate of w' is given by

$$\hat{w}' = (R^T R)^{-1} R^T y \quad \text{---4.34.}$$

An estimate of $\hat{b}$ of the parameter b may be obtained as follows:

Writing $\hat{b}\hat{w} = \hat{w}'$ ---4.35

implies that $\hat{b}\sum_{i=1}^{m} \hat{w}_i = \sum_{i=1}^{m} \hat{w}_i'$ and therefore $\hat{b} = \sum_{i=1}^{m} \hat{w}_i'$ since $\sum_{i=1}^{m} \hat{w}_i = 1$.

Thus $\hat{b}$ is simply equal to the sum of all elements of w'. Using this estimate, the corresponding estimated weighting function ordinates are given by Equation 4.36.

$$\hat{w}_i = \frac{\hat{w}_i'}{\hat{b}} \quad \text{---4.36.}$$

Where i = 1,2,...m

Simulated flow y_i at the time instant i may thus be obtained using equations 4.37.

$$y_i = \sum_{j=1}^{m} \hat{w}_j' z_{i-j+1} \quad \text{---------------------------------------4.37a or}$$

$$y_t = \hat{b} \left[\sum_{j=1}^{m} \hat{w} z_i x_{i-j+1} \right]$$ --4.37b.

4.4.1.4 Calibration of The SMAR Model

The calibration procedure of the SMAR model involves successive choices of combinations of the five water balance parameters H, T, Y, C, or Z in order to determine the dominant parameters for the Nyando catchment. These parameters are to be optimised, where H refers to the generated 'direct runoff' coefficient, T refers to a parameter (less than unity) that converts the given evaporation series to the model-estimated potential evaporation series, Y refers to the maximum infiltration capacity depth (mm), which is obtained by integrating the instantaneous infiltration rate over the storm period, C refers to the evaporation decay parameter, facilitating lower evaporation rates from the deeper soil moisture storage layers and Z refers to the combined water storage depth capacity of the layers (mm).

These parameters cannot be evaluated effectively together because the operations of several of them are similar in their determination of the volumes of effective rainfall. For instance H and Y, C and Z, C and T, and T and Z are similar in their effects on the volume of effective rainfall. Therefore in any application the model consisting of a subset of these elements must be chosen.

Consider a case where we set H = 0. The proportional runoff component will be removed from the model. By setting C = 1.0, the layers are replaced by a single storage capacity Z, all of which is available for evaporation at the potential rate. The effect of T is removed if it is set at T = 1.0. Similarly the effect of Y is removed if it is set at a value exceeding the observed daily maximum rainfall during the period of record. Z becomes inoperative if it is set to a sufficiently large value.

Further, not all these parameters would be important in any one application as was found out in previous studies by O'Connell et al (1970), Mandeville et al (1970) and Gallagher (1986). The parameters to be retained depend on the characteristics of the catchment being studied.

After estimating the above parameters, the series of generated runoff (x) was then calculated. The pulse response function was then estimated by the method of ordinary least

squares, i.e. G and u_i. Computed discharge $(\hat{y})$ was then computed before finally determining the value of the objective function, F, where

F = f(H, T, Y, C, Z and G, u_i) or

$$F = \sum_{j=1}^{n} (y_t - \hat{y})^2 \text{ --4.38.}$$

The search was continued to find the values of H, T, C, Y and Z , which together with the best value of u would minimise F. Since the calculation of y involves the calculation of a pulse response function by ordinary least squares, this operation has to be repeated each time F is calculated (Kachroo, 1992).

To guarantee that the total volume of generated runoff is equal to the total volume of observed discharge for the period of calibration, the constraint:

$$\sum_{j=1}^{n} \hat{x}_j = \sum_{j=1}^{n} y_j \text{ --4.39}$$

was imposed where $\hat{x}_j$ refers to the effective rainfall or generated runoff, y_j refers to the observed discharge and n refers to the length of calibration period.

Optimisation of F was carried out with G constrained to unity using the method of constrained least squares. This method of calibration eliminates the possibility of interference between the operations of the water balance and the routing components of the model.

For a sub-model consisting of two-water-balance parameters, the volumetric constraint can be imposed by drawing a contour diagram of the objective function surface F as a function of the water balance parameter values. On this diagram the ratio of the estimated to the observed discharge volumes can also be plotted. The overall optimum parameter values correspond to the point of minimum value of the objective function surface on the line of the unit volumetric ratio (Kachroo, 1992).

The model calibration and verification periods are the same as those of the linear systems models. The Rosenbrock optimisation technique was used to estimate the parameters of the water balance components and the pulse response function of the routing component was estimated by the method of constrained least squares.

First each parameter was optimised individually in order to determine the percentage of the initial variance accounted for by each parameter. While one parameter was being

optimised the rest were set to their inoperative values. This was repeated for each of the parameters H, T, Y, C, and Z in trials 1 to 5 as shown in Table 5.15. In trials 6 to 11 two water-balance parameters were taken together while the rest were set to their inoperative values. In trials 12 to 16 three water-balance parameters were taken together while the rest were set to their inoperative values.

The results of the sub model forms used and the parameter values obtained over the calibration and verification periods for the Nyando catchment are given in Chapter 5, Table 5.15.

4.4.1.4.1 General Linear Routing Component

After selecting a suitable model configuration for the water balance component, together with the optimum values of the parameters, it was then necessary to establish a linear relationship between the perturbations of the simulated runoff from its mean and the perturbations of the observed discharge from its corresponding mean. To distinguish this from the LPM, we refer to it as the general linear model.

If x and y are linearly related, then the corresponding perturbations are also linearly related because the mean of the generated runoff is equal to the mean of the observed discharge during the calibration period (Kachroo, 1992). The advantage of seeking a relationship between perturbations instead of the original series is that the method of ordinary least squares can be used instead of the constrained least squares without significantly relaxing the volumetric constraint.

The linear routing component was applied under the shape of the gamma function impulse response given by Equation 4.40.

$$h(t) = \frac{1}{K\Gamma(n)} e^{-\left(\frac{t}{K}\right)} \left(\frac{t}{K}\right)^{n-i} \quad \text{---4.40.}$$

This relates an input in blocks of duration T, to an output expressed as a series of ordinates at the same interval. Equation 4.41 gives the corresponding pulse response.

$$h(T,t) = [S(t) - S(t-T)]/T \text{ ---4.41.}$$

Where $\quad S(t) = \int_0^t h(\tau)d\tau$

The corresponding impulse response was obtained from $h(T, t)$ by Equation 4.42.

$$h_j = \frac{1}{T} \int_{(j-1)T}^{jT} h(T,t)dt \text{ ---4.42.}$$

The results are presented in Table 5.17 in the next chapter.

4.5 Model Verification

This refers to the process of measuring the accuracy of the calibrated model during calibration. Once the model has been calibrated, a different set of data is normally used to test the model's accuracy in predicting runoff using rainfall data. This is done using concurrent rainfall and runoff data where simulated runoff values are compared with the corresponding observed values. It is actually an attempt to demonstrate the accuracy of the calibrated model. If the results lie within acceptable limits, the data and the model are accepted and the model is verified. If the simulated results lie outside the acceptable limits, the basic assumptions and the data used should be thoroughly checked to establish the source of the problem.

In order to determine the range of uncertainties, it is advisable to present observed data against simulated data on a scatter plot. The observed data is presented on the horizontal axis and the simulated data on the vertical axis. The mean absolute error should be calculated and taken as the mean error within which the uncertainty should lie.

4.5.1 Model Efficiency Criteria

The criteria by which the performance of each model is judged are presented in this section. The efficiency criteria that express the accuracy of each model are generally linked with the objective function used in calibration for optimizing its parameters.

A commonly used objective function is the sum of squares of differences, F, between the observed and estimated discharges, where the summation is taken over the whole calibration period (Kachroo, 1992).

$$F = \sum (y - \hat{y})^2 \text{ ---4.43.}$$

Where F represents the sum of squares of differences or the index of disagreement, y represents the measured or observed output and $\hat{y}$ represents the model output estimates.

The quantity F is an index of residual error that reflects the extent to which a model is successful in reproducing the observed discharges. It is thus useful as an appropriate preliminary criterion for expressing model accuracy. It can also be used to compare various hydrological forecasting models on the same catchment provided they have the same length of data record. The lower the values of F, the better the performance of the model as this indicates small differences between estimated and observed series.

Nash and Sutcliffe (1970) corrected the shortcomings of Equation 4.43 by defining model efficiency R^2 that is analogous to the coefficient of determination in linear regression. By so doing different models could be compared even in different catchments. They defined the efficiency R^2 as the proportion of the initial variance accounted for by F.

$$R^2 = \frac{F_0 - F}{F_0} \text{---4.44.}$$

Where initial variance, F_0, was defined as

$$F_0 = \sum (y - \bar{y})^2 \text{-------------------------------------4.45.}$$

Where $\bar{y} = \frac{1}{N} \sum y_i$, the mean of y in the calibration period, N refers to the number of data points, F_0 refers to the no model value of F or initial sum of squares of error from a primitive model such as long-term mean values.

The Efficiency criterion, R^2, is the proportionate reduction of the initial variance by means of substantive input–output transformation model. This is a measure of the percentage improvement of the applied substantive model from the naïve or no model situation. In this case it is the applied model that is considered substantive. In application to the calibration period, these quantities, i.e., F_0 and F, are all obtained with reference to the calibration period so that R^2 is identical to the coefficient of determination and varies between zero and one.

In the verification period, F_0 is calculated as the sum of squares of deviations within the verification period from the mean of calibration period. In this case R^2 compares the sum of squares of model errors with the sum of squares of errors that would occur when the only forecast available for the verification period were the mean value of the discharge during the calibration period. Garrick et al (1978) suggested that a model's performance should be judged against the best estimates that could be obtained in the absence of any model

or in the event of using a naïve model.

The sum of squares of residual errors associated with the forecasts of a particular model, expressed as a proportion of the corresponding sum of squares of errors associated with a chosen naïve model, is an indication of the efficiency of that model relative to the particular naïve model. This is what is measured by the R^2 value.

Equation 4.44 may take on negative values in the verification period especially when the model under test produces output forecasts which are worse estimators than is the mean of the recorded output over the calibration period (Kacharoo, 1992).
To compare the relative accuracies of different models using the same data, R^2 criterion provides a convenient index of comparison of the corresponding sums of squares of model residual errors. A similar criterion may be used to express the proportion of the initial variance unaccounted for by model 1, the less substantive one, which is subsequently accounted for by model 2 the more substantive one. Such a criterion may be expressed in the form.

$$r^2 = \frac{R_2^2 - R_1^2}{1 - R_1^2} \text{---4.46.}$$

Which simplifies to

$$r^2 = \frac{F_1 - F_2}{F_1} \text{---4.47.}$$

where the subscripts refer to models 1 and 2 respectively. Thus r^2 is the measure of the proportion of improvement of the more substantive model over the less substantive one.

Another measure of the degree of model performance is the volumetric agreement between observed flows and the model estimated flows in a given period, say calibration or verification periods. This is calculated as the ratio of the observed mean discharge to the estimated mean discharge series. This ratio is called the index of volumetric fit (IVF). It serves as a quick guide as to whether the model overestimates or underestimates the flow. It is defined by the following equation:

$$IVF = \frac{\sum_{i=1}^{T} \hat{y}_i}{\sum_{t=1}^{T} y_t} \text{---4.48.}$$

Where $\hat{y}_i$ refers to the estimated discharge series, y_i refers to the observed discharge series and T refers to the period under consideration.

A value of IVF = 1 represents a perfect volumetric agreement between the estimated, $\hat{y}_i$ and the observed y_i, discharges, where observed and estimated discharge volumes are equal. A value of IVF < 1 represents underestimation of the total volume of discharge in the period T. A value of IVF > 1 represents overestimation of the total volume of discharge in the period T.

For the cases where IVF ≠ 1, an improved forecast may be achieved through the use of an appropriate updating procedure. The updating procedure used in this study is described in detail in Section 4.3.2.

The mean square error (MSE) between the observed and the estimated flows is also considered a viable measure of the performance of a model. Equation 4.49 defines the MSE.

$$MSE = \frac{1}{D}\sum_{i \in D}(y_i - \hat{y})^2 \text{------------------------------}4.49.$$

Where $\hat{y}$ refers to the simulated model flow, y_i refers to the observed flow and D refers to the duration of interest.

The relative performance of a model in fitting the data of two different samples may be evaluated through the comparison of the MSE calculated for each split sample. It is considered a good measure of model performance especially when comparing different models on the same catchment. The smaller the value of MSE, the closer the model estimated values are to the observed values.

In order to achieve the specific objectives, daily data analysed to establish their quality before applying them to the various models. Once the data quality control was carried out, various models were ran to establish the one that best fits the Nyando catchment.

4.5.2 Updating Model Forecasts

Model updating refers to the process of modifying the output from a model just before issuing a forecast. This is done by making corrections to the output in accordance with the behaviour of errors occurring in the previous forecasts up to the time at which the present forecast is being made. The persistence of forecast residuals is an indication of a possible model inadequacy. It may be difficult to see how to improve the model structure in order to

remove the persistence. In such cases, there is need to resort to updating procedures, where errors are used to improve the forecasts.

The validity of any updating procedure depends on the existence of some degree of persistence in the substantive model residuals. Analysis of this persistence is used as a means of suggesting the appropriate updating procedure.

When a model is used for real-time forecasting, there is need to incorporate an appropriate updating procedure. The model is then said to operate in updated or operational mode. In this mode the model transforms the inputs to outputs after updating (Kachroo, 1992). The performance of models in updated mode may then be compared. The comparison procedure is the same as that used when the models are in simulation mode.

In an ideal situation the model should resemble the actual system so that the residuals in the calibration period should be a series of unrelated quantities whose expectation is zero with a small variance. In reality, however, persistence in residuals is normally observed. This results from either inadequacies from the model structure, incorrect estimation of the model parameters, errors in the data or absence of any consistent relationship in the data.

The nature of the persistence provides the basis for setting up an updating procedure whereby the output from the model can be modified prior to issuing the forecasts by making corresponding allowances for the persistence of the observed errors in the recently computed output values. Analysis of the model residuals obtained in the calibration for evidence of persistence provides the basis for the updating procedure. One of the commonly used updating procedures is the auto regressive updating component.

Though the results in simulation mode were satisfactory for all the models during calibration, there was need to update the models in order to improve the output accuracies during verification period.

4.5.2.1 Autoregressive Updating Component

Let the series of the residuals e_t follow an autoregressive (AR) scheme of order p, such that

$$e_t = \phi_1 e_{t-1} + \phi_2 e_{t-2} + \ldots + \phi_p e_{t-p} + a_t \quad \text{------------------------4.50.}$$

Where a_t is white noise of mean zero and variance σ_a^2.

Multiplying the equation by the general term e_{t-j} and taking expectations results in an equation in autocovariances. Division by the variance of e_t series results in an equation in the autocorrelation coefficients of the form:

$$\rho_j = \phi_1 \rho_{j-1} + \phi_2 \rho_{j-2} + ... + \phi_p \rho_{j-p} \quad j = 1,2,3...,p \text{----------4.51.}$$

For each value of j, varying from 1 to a maximum of p, there is one such equation. Hence the maximum number of equations will be p. These are collectively known as Yule-Walker equations (Box and Jenkins, 1976).

If the population quantities $\{\rho_k\}$ are replaced by the sample autocorrelation coefficients, r_k, in the ordinary way, then these equations can be solved for p unknowns $\{\phi_p\}$. These values are then inserted into Equation 4.50. If Q_{kj} denotes the jth coefficient in an autoregressive process of order k then Q_{kk} is the last coefficient in the equation:

$$\rho_j = \phi_{k1} \rho_{j-1} + \phi_{k2} \rho_{j-2} + ..., + \phi_{k(k-1)} \rho_{j-k+1} + \phi_{kk} \rho_{j-k} \quad j = 1,2,3,...,k \text{--------4.52.}$$

The quantity Q_{kk} regarded as a function of lag k, is called the partial autocorrelation function. It has a cut off after p unlike the autoregressive process, which has an infinite autocorrelation function. Hence Q_{kk} provides an indication of the appropriate order of the autoregressive models.

Having fitted the autoregressive model to the residuals, the correction of the next forecast, based on the observed error of the most recent past forecast, is given by:

$$\hat{e}_t = \hat{\phi}_1 e_{t-1} + \hat{\phi}_2 e_{t-2} + ..., + \hat{\phi}_p e_{t-p} \text{------------------------4.53.}$$

The updated forecast then becomes:

$$\hat{Q} = \hat{y}_t + \hat{e}_t \text{--4.54.}$$

Where $\hat{y}_t$ refers to the non-updated model output and $\hat{e}_t$ refers to the estimated error correction based on the previous error observations.

In practice forecasts are required over a lead-time L of several days defined by:

$$\hat{Q}_t(L) = \hat{y}_t(L) + \hat{e}_t(L) \text{---} 4.55.$$

where $\hat{y}_t(L)$ refers to the un-updated model output of lead-time L and $\hat{e}_t(L)$ the lead-time error correction.

The lead-time L error correction $\hat{e}_t(L)$ can be expressed as a weighting of previous observations. It may be written as

$$\hat{e}_t(L) = \hat{\phi}_1(L)e_{t-L} + \hat{\phi}_2(L)e_{t-L-1} + \ldots + \hat{\phi}_p(L)e_{t-L-p+1} \text{-----------} 4.56.$$

Where the coefficients $\Phi_i(L)$ are known as minimum mean square error forecast weights. Equation 5.57 can calculate them from the lead-one forecasts weights.

$$\Phi_i(L) = \hat{\phi}_1(L-1)\hat{\phi}_i(1) + \hat{\phi}_{i+1}(L-1) \text{-----------------------------} 4.57.$$

Most models imply some dependence of the output forecast value on input quantities between the time of making the forecast and the time for which the forecast is made. This is the lead-time of the forecast. Model updating enables the model to be used as a real-time forecasting tool where the updating is actually an empirical correction to each of the simulation mode forecasts. Therefore if the simulation mode forecast is not good, it is expected that even the updated forecast may not be all that good. This is so because updating does not actually improve the structure of the model from the scientific point of view. It only improves the output through the empirical correction.

CHAPTER 5

5.0 RESULTS AND DISCUSSIONS

In this chapter, the results obtained from the various methods used in this study, are presented.

5.1 Estimation of Missing Data

Estimation of the missing data records formed a vital part of this study. Results that were obtained from the estimation of missing rainfall, runoff and evaporation records are hereby presented.

5.1.1 Rainfall Data

Before attempting to fill in any missing data, it was necessary to first establish the linear relationship between the stations. This was done by calculation of the linear correlation coefficients, $r_{x\,y}$, between the rainfall stations that were selected for this study. The equation used to compute the inter-stations linear correlation coefficients; $r_{x\,y}$ was presented in Section 3.2.1. The purpose of generating these values was to find out whether data from a certain station were related to data from other stations. To do this, coding of the stations was necessary so that the data could fit the table. The results of the correlation coefficients are presented in Table 5.1.

The results indicate that most rainfall stations are reasonably linearly correlated to each other since $r_{x,y} \geq 0.5$ for most of the stations. This threshold was taken on the basis that values beyond 0.5 are closer to 1, the value that signifies perfect positive relationship. Values below 0.5 are closer to zero, the no relationship value. For the purpose of this study, values beyond 0.5 may therefore be considered significant while those below 0.5 may be considered insignificant.

Except for station G1, whose $r_{x,y}$ values range from −0.3 to 0.5, all other stations exhibited values of linear correlation with $r_{x,y} \geq 0.5$. The low values of $r_{x,y}$ for station G1 indicate lack of a reasonable linear relationship between data from this station and data from the rest of the stations. This station was also found to have abnormally high values of rainfall when compared to corresponding values at other stations. This was a clear indication that data from this station was inconsistent with data from the other stations. On the basis of the unusually high values of data and the very low linear correlation coefficient

values, it was decided that station G1 should not be included in the calculation of areal average rainfall over the Nyando catchment.

From the high values of linear correlation coefficients, it was possible to fill in the missing records at the subject station using data from related stations by multiple or linear regression method using the stations that showed a high degree of linear relationship. The subject station stands for the station with missing data records. The equations applied for the purpose of filling in the missing data records are presented in Section 3.2.1. For the linear regression method, the related station was taken to be the one that exhibited an unusually high value of $r_{x,y}$, compared to other stations, between itself and the subject station. Values of $r_{x,y}$ that are greater than 0.5 for each station are bolded as can be observed from the table. The bolded values indicate the stations that are highly correlated and could be used to fill in the missing records in the subject station. The choice of three stations for regression purposes was based on the fact that fitting too many stations does not normally improve the outcome in any significant way. As stated earlier the rain gauge coding used here is purely for the purpose of convenience.

Table 5.1: Linear inter-station correlation coefficients ($r_{x,y}$) for annual rainfall between rainfall gauging Stations within the Nyando catchment. Bolded value those that are equal or greater than 0.6. Coding of the stations was only for the purpose of calculating $r_{x,y}$.

G1	G2	G3	G4	G5	G6	G7	G8	G9	G10	G11	G12	G13	G14	G15	G16	G17	G18
1.0																	
0.3	1.0																
-.2	**0.6**	1.0															
0.2	0.3	**0.7**	1.0														
0.4	**0.7**	0.3	-.1	1.0													
0.3	0.0	-.1	-.3	**0.6**	1.0												
-.2	0.5	0.4	-.1	0.4	0.1	1.0											
-.3	**0.6**	**0.7**	0.1	0.5	0.3	0.4	1.0										
0.2	**0.7**	0.4	0.3	**0.9**	**0.6**	**0.6**	**0.6**	1.0									
-.2	**0.9**	**0.8**	-.1	**0.6**	0.0	**0.6**	**0.6**	**0.6**	1.0								
0.0	**0.6**	0.3	**0.6**	0.5	0.1	**0.6**	0.3	0.5	**0.7**	1.0							
0.1	**0.8**	**0.8**	0.3	0.4	-.2	0.4	0.5	0.4	**0.7**	0.5	1.0						
0.1	**0.9**	**0.6**	-.3	**0.7**	0.1	0.4	**0.6**	**0.8**	**0.7**	0.2	**0.6**	1.0					
0.3	0.5	0.0	0.3	**0.7**	0.5	-.1	0.5	**0.6**	0.3	0.2	0.0	0.5	1.0				
0.1	**0.9**	**0.7**	**0.6**	**0.7**	0.0	**0.7**	0.5	**0.7**	**0.9**	**0.8**	**0.8**	**0.7**	0.2	1.0			
-.1	**0.9**	**0.8**	0.5	0.4	-.3	0.5	**0.6**	0.5	**0.8**	0.4	**0.9**	**0.8**	0.2	**0.8**	1.0		
0.3	**0.8**	**0.7**	0.5	0.5	-.2	0.2	0.4	0.5	**0.7**	0.3	**0.9**	**0.8**	0.2	**0.7**	**0.9**	1.0	
0.5	**0.7**	0.4	0.3	**0.6**	0.2	0.4	0.3	0.5	**0.6**	**0.6**	0.4	0.5	0.3	**0.7**	0.4	0.4	1.0

5.1.2 Runoff Data

Just like for the rainfall data, the missing runoff records were estimated by the regression method. The results of the application of the correlation equation are presented in Tables 5.2 and 5.3.

From the results in Table 5.2, it is observed that station 1GD03, Nyando downstream, shows a highly reasonable degree of linear relationship with all the other stations except Kibos, station 1HA01. Data from this station seem not to a good linear relationship with data from the other stations as shown by the low values of $r_{x,y}$. Data from this station may therefore not be used to fill in missing records in any of the other stations. Any one of the rest of the stations could therefore be used to fill in missing data records in station 1GD03 using linear regression method, as presented in Section 3.2.1, provided the station has data corresponding to the missing records.

Table 5.3 shows stations that are significantly correlated with 1GD03. The first row shows the gauging stations whose $r_{x,y}$ values are equal or greater than 0.5. The second row shows values of $r_{x,y}$ between 1GD03 and the rest of the stations except 1HA01. The third row shows the percentage R^2 values of double mass curves between 1GD03 and the other correlated stations as given by the trend line.

Table 5.2: Linear inter-station correlation coefficients ($r_{x,y}$) for daily average discharge between gauging station 1GD03 and some selected stations within the Nyando catchment.

	1GD03	1GD04	1GB11	1GC04	1GC05	1GD07	1GG01	1HA01
1GD03	1.0							
1GD04	0.6	1.0						
1GB11	0.6	0.3	1.0					
1GC04	0.7	0.5	0.5	1.0				
1GC05	0.7	0.4	0.4	0.6	1.0			
1GD07	0.7	0.2	0.4	0.5	0.7	1.0		
1GG01	0.5	0.9	0.2	0.3	0.	0.1	1.0	
1HA01	0.2	0.4	0.0	0.1	0.1	0.1	0.4	1.0

Table 5.3: Significance of stations' relationship with station 1GD03 as given by $r_{x,y}$ and R^2 %.

	1GD03	1GD04	1GB11	1GC04	1GC05	1GD07	1GG01
$r_{x,y}$	1.0	0.6	0.6	0.7	0.7	0.7	0.5
R^2 %	100	94.11	99.78	99.91	99.44	98.72	79.80

The high values of R^2 indicate that the trend line is a good fit to the observations. The respective equations of the trend lines could therefore be used to fill in the missing records in 1GD03. The R^2 values indicate that station 1GC04 is the one most likely station to be used for the purpose of filling in the missing data records in station 1GD03 since it has the highest value of R^2 (99.91) as well as the highest value of $r_{x,y}$ (0.7). If corresponding data is not available, station 1GB11 may be used as it has the second highest value of R^2 (99.7 %). However the correlation coefficient value for 1GB11 is lower (0.6) than that of other stations except stations 1GG01 and 1HA01. Hence station 1GC05 may be preferred on account of the high correlation coefficient value (0.7) and the R^2 value (99.44 %) respectively. The ideal procedure would be to use a station that shows the highest values of R^2 and to move to a lower one only when the corresponding data is missing. The reality on the ground was that station 1GD07 was the preferred station because the recorded values of data from this station, with a mean of 9.62 cumecs, were closer in magnitude to those of station 1GD03, with a mean of 17.49 cumecs, than those from any other station. All the other stations had much lower values of the mean than 1GD07.

5.1.3 Evaporation Data

The missing evaporation data records were estimated by seasonal mean method. This was done using the seasonal daily mean evaporation. The mean of all the daily evaporation measurements on a day such as 1st January were used to fill in all the gaps that appeared on 1st January of any year. This was done using data from the same station. In this case it was done using daily evaporation data from Kisumu Meteorological station.

5.2 Data Descriptive Statistics

The quality of data in research occupies a central place in any scientific research. This is because the quality of data will always determine the degree of accuracy and therefore reliability of the results. It is therefore important to devote more time in establishing the quality of data before commencing any analysis. This section gives a brief overview of descriptive statistics of rainfall, runoff and evaporation data

5.2.1 Rainfall Data

The summary of the basic statistics of rainfall data is presented in Table 5.4. The basic statistics are intended to give a preliminary idea of the frequency distribution and

the variability of data from various rainfall-gauging stations within the Nyando catchment.

From these results, the mean annual rainfall was observed to vary from about 1100 mm to 1700 mm with a majority of stations recording between about 1150 mm and 1550 mm p.a. Thus the catchment may be considered as one that on average experiences moderately heavy rainfall.

Standard deviation is a measure of how data sets are distributed about their mean value. A small standard deviation means that the data sets are bunched closely about their mean value while a large standard deviation means that the data sets are spread over considerable distances away from their mean value. The standard deviation basically measures distances by which individual values in a frequency distribution depart from their mean value. A sample with a large standard deviation is more spread out from the mean value than that with a smaller one. Within a single sample, the standard deviation measures the number of data sets that lie within 1, 2 or 3 standard deviations (Freund, 1967).

From Figure 5.1, it is observed that there are moderately low values of standard deviations, except for station G1 which has an unusually large value of standard deviation, ranging from about 125 mm to about 340 mm. This is an equivalent of about ten to twenty percent of the mean; an indication that most of the observations are fairly well clustered around the mean annual rainfall values.

It is observed from Figure 5.1 that station number G1 has a standard deviation of about forty percent. This is a much higher value of standard deviation than those found at the other stations. It indicates that data from this station are more spread out from the mean value compared to the data from the other stations. Values from this station may indeed be considered as inconsistent with values from the other stations. It was also found out that data from this station showed very low correlation in relation to data from the other stations. This further justifies the exclusion of this station from the calculations of the mean areal rainfall over the Nyando catchment.

The coefficient of variation is a measure of relative dispersion of data sets. It measures dispersion relative to the size of data unlike the standard deviation. It measures the extent to which individual data points scatter around the mean value. A coefficient of variation value of about 10 to 20 % suggests a low degree of dispersion whereas a coefficient of variation value of about 70 to 80 % means that individual values scatter fairly widely around the mean value of the frequency distribution (Thiessen, 1997).

The coefficients of variation in most stations, as observed in Table 5.4, are below 0.2 or twenty percent except for station G1. These fairly low values of coefficients of variation indicate low variability of rainfall data. They may be interpreted to mean that the annual mean can be relied on as a measure of central tendency. Most observations may therefore be considered as clustered around the mean.

Table 5.4: Summary of basic statistics of annual rainfall data over the Nyando catchment. The coding of stations was only for the purpose of convenience.

Station No.	Station Name	Code	Mean	Standard error	Standard deviation	Coefficient of variation	Skewness	Kurtosis
8935001	Songhor	G1	1958.4	403.7	1211.2	0.61	0.8	-0.7
8935033	Savani	G2	1542.5	114.5	343.5	0.22	-0.2	0.5
8935148	Kipkurere	G3	1169.6	92.9	278.8	0.24	-0.1	-0.9
8935159	Ainabkoi	G4	1349.5	79.7	239.1	0.18	-0.1	-1.1
8935161	Nandi Hills	G5	1350.3	85.4	256.3	0.19	-0.2	0.2
9034086	Ahero	G6	1171.3	42.5	127.4	0.11	0.4	-0.7
9035020	Kipkelion	G7	1076.3	49.1	147.2	0.14	0.8	0.1
9035075	Kaisugu	G8	1712.9	85.0	254.9	0.15	-0.2	-0.1
9035148	Koru	G9	1295.7	63.4	190.2	0.15	-0.1	-0.2
9035151	Entomology	G10	1086.2	76.1	228.4	0.21	0.8	-0.7
9035155	Londiani	G11	1094.6	68.5	205.5	0.19	1.3	1.7
9035188	Tinga	G12	1274.9	47.7	143.2	0.11	-1.0	0.9
9035201	Kipkorech	G13	1672.0	94.7	284.1	0.17	-0.4	0.1
9035220	Homaline co.	G14	1344.6	82.8	248.4	0.18	-0.1	-0.4
9035226	Londiani Forest	G15	1096.1	58.2	174.7	0.16	0.6	-0.2
9035240	Keresoi	G16	1425.8	94.6	283.7	0.20	-0.5	1.0
9035256	Malagat	G17	1152.2	64.3	192.8	0.17	-1.1	0.3
9035263	Tinderet	G18	1589.3	75.7	227.1	0.14	0.8	1.1

Skewness is a measure of the degree of departure of a frequency distribution from symmetry. A frequency distribution that is not symmetrical is said to be skewed. The extent of departure from symmetry is normally measured using Pearson's coefficient of skewness given by:

$$Sk_p = \frac{3(mean - median)}{S \tan darddeviation} \quad \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots 5.1$$

For a symmetrical distribution, Pearson's coefficient of skewness value is zero. The greater the departure from symmetry of a frequency distribution, the higher the absolute values of Sk_p (Moore, 1974).

Except for station G1, with a skewness coefficient value of 1.3, all other stations have a skewness coefficient values within twice their standard error. The standard error of

skewness coefficient is approximated using the relation $\left(\dfrac{6}{N}\right)^{\frac{1}{2}}$, where N is the number of observations. Values of skewness coefficients are considered statistically significant if they are greater than twice their standard errors (Ogallo, 1977).
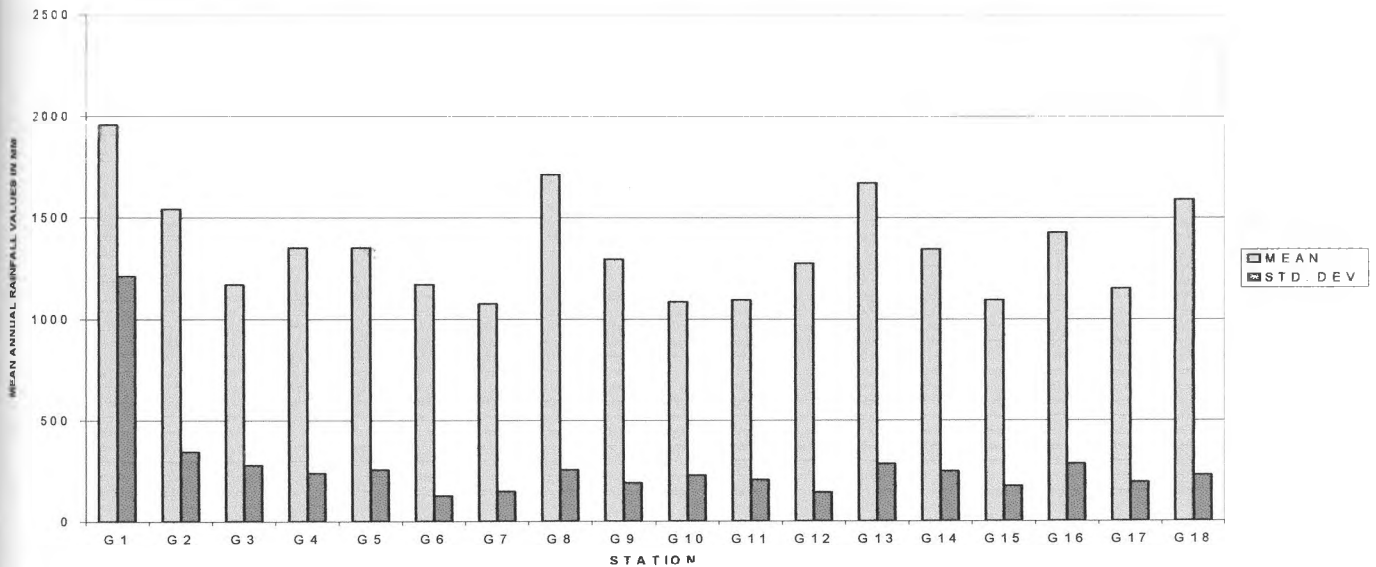

Figure 5.1: Bar graph of mean annual rainfall depth and standard deviation (mm) over the Nyando catchment

From Table 5.4, the values of skewness coefficients vary from -1.0 to 1.3. These are low values and very close to zero, which is the value for perfect symmetry. The values lie within twice their standard errors and they can therefore be considered insignificant. The low values may therefore be considered an indication that rainfall data is approximately symmetrically distributed about the mean value for all the rain gauge stations.

Kurtosis is a measure of the general peakedness or flatness of a frequency distribution where a kurtosis coefficient value of about 1 is considered normal (Frank, 1994). Values of kurtosis coefficients are considered statistically significant if they are greater than twice their standard errors. The standard error of kurtosis coefficient is approximated using the relation $\left(\dfrac{24}{N}\right)^{\frac{1}{2}}$, where N is the number of observations (Ogallo, 1977).

From Table 5.4, it is observed that the values of kurtosis coefficients vary from -1.1 to 1.7. They all lie within twice their standard error and may therefore be considered

insignificant. They are fairly close to one, the value that signifies normal peakedness of the frequency distributions and are therefore an indication that the distribution curves for these data are close to normal.

Since skewness and kurtosis are both measures of normality of a frequency distribution, we may conclude that rainfall data over the Nyando catchment generally follows a normal distribution.

5.2.2 Runoff Data

The results of the descriptive statistics of runoff data are presented in Table 5.5. It is observed from the table that the mean discharge from the catchment as gauged from 1GD03 is about 17cumecs.

Unlike in the case of rainfall, there is more variability in the runoff data as seen from the unusually high values of the coefficient of variation. All the values are above 20 % except for station 1GD04. This suggests a higher degree of dispersion of the flow data than that of the rainfall data.

The skewness coefficient values are all positive. Except for stations 1GD03, 1GD04 and 1GC04, with skewness coefficient values within twice their standard errors (1.36), the other stations show skewness coefficient values beyond twice their standard errors. Values of skewness coefficient within twice their standard errors may be considered insignificant while those beyond twice their standard errors are significant. The flow data over the Nyando catchment may therefore be considered as positively skewed. This means that there are more flow values that are less than the mean flow value.

Except for stations 1GD03, 1GD04 and 1GC04 with kurtosis coefficient values within twice their standard errors (2.72), the other stations show kurtosis coefficient values beyond twice their standard errors. This shows that the frequency distribution of the flow data is more peaked than that of the rainfall data.

Most outstanding is station 1HA01 where the coefficient of variation, skewness and kurtosis are unusually high compared to the rest of the stations. This shows that data sets from this station have more variability than those from the other stations as shown by the high value of the coefficient of variation (2.0). The frequency distribution is also not symmetrical. It is positively skewed as shown by the high value of skeweness coefficient (5.5). The distribution of this data is also highly peaked as shown by the unusually high

value of kurtosis (32.1). Hence data from this station may not be compatible with data from the other stations.

Table 5.5: Basic statistics for daily average discharge data for some selected flow-gauging stations within the Nyando catchment.

Station number	Mean	Standard error	Standard deviation	Coefficient of variation	Skewness	Kurtosis
1GD03	17.49	0.34	8.71	0.50	0.79	0.92
1GD04	4.64	0.15	3.90	.084	1.32	1.23
1GB11	7.83	0.07	1.83	0.23	1.99	5.35
1GC04	0.39	0.01	0.20	0.51	0.73	1.52
1GC05	1.03	0.04	0.91	0.88	1.75	3.88
1GD07	9.62	0.28	7.16	0.74	1.89	4.89
1GG01	2.05	0.11	2.89	1.41	2.11	3.32
1HA0I	1.21	0.10	2.42	2.00	5.5	32.09
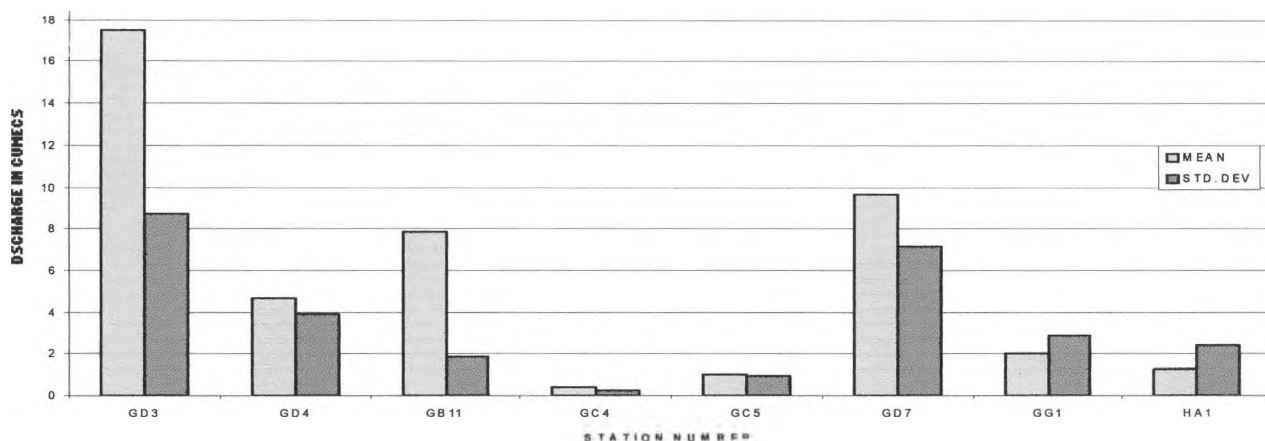
Figure 5.2: Bar graph of daily average discharge and standard deviation (cumecs) for some selected flow-gauging stations within the Nyando catchment.

Thus the majority of the runoff data are less than the mean discharge value. This is unlike the case of rainfall data where the skewness values range from slightly negative to slightly positive values indicating a possible symmetrical distribution of the annual rainfall on average.

Figure 5.2 shows that there are high values of standard deviations ranging from about 0.2 to about 8.7 cumecs. This is equivalent to about twenty to sixty percent of the mean. This is an indication that discharge observations are scattered far apart from the mean discharge value. This is unlike the case for rainfall data. However station 1GD03 has the lowest value of the coefficient of variation; a possible indication that data from this

station may be considered to be more clustered around the mean value than is the case for the rest of the stations.

5.2.3 Evaporation Data

The results of the descriptive statistics of evaporation data are presented in Table 5.6. From the results we observe that the mean daily evaporation is about 6.3 mm for Kisumu and 3.8 mm for Kericho. There are low standard deviations as observed in Figure 5.3. These are 0.88 mm for Kisumu and 0.74 mm for Kericho. These values are equivalent to about ten percent and fifteen percent of the mean value for Kisumu and Kericho respectively. The low values are an indication that most of the evaporation data are clustered around the mean value. Coefficients of variation are below 0.2 i.e., 20 % for both stations indicating low variability of evaporation data.
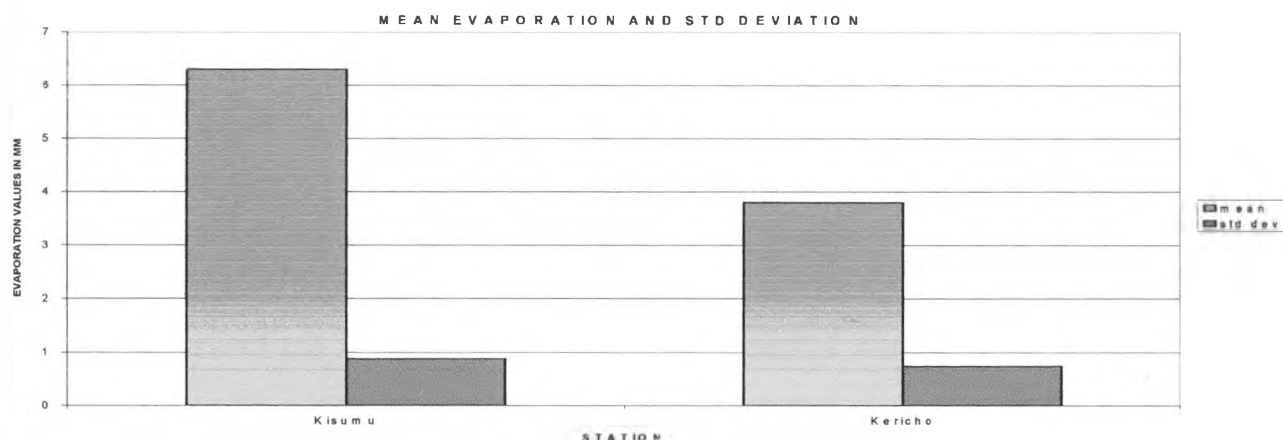
Figure 5.3: Bar graph of mean daily evaporation and standard deviation (mm) for Kisumu and Kericho stations

Table 5.6: Basic statistics for daily average evaporation data for Kericho and Kisumu meteorological stations.

Station Name	Maximum	Minimum	Mean	Standard deviation	Coefficient of variation	skewness	Kurtosis
Kisumu	10.4	4.7	6.3	0.88	0.14	1.02	1.87
Kericho	5.9	1.7	3.8	0.74	0.19	-0.51	0.36

5.3 Test For Consistency of Data

Consistency of data is important in any research for the results to be meaningful. It was therefore important to test the consistency of data and make corrections where

necessary. Mass curve method was used to test for the consistency of daily rainfall, runoff and evaporation data respectively, over the Nyando catchment. The method was described in detail in Section 3.4.2.

5.3.1 Rainfall And Runoff Data

Results of rainfall data consistency testing are presented in Table 5.7 and in Figures 5.4 and 5.5.

The results from this method show that in general only a single line could be fitted in all the rainfall and runoff records. When trend lines were fitted to each of the mass curves it was observed that the percentage R^2 values were quite high, indicating that the trend line was a very good fit to the observed data. This means that observed data were quite close to the single straight line and therefore data were quite consistent.

From table 5.7, it is observed that out of the eighteen rain gauge stations available, seventeen stations exhibited percentage of R^2 values greater than 99 %. These are very high values and on this basis we can conclude that daily rainfall data over the Nyando catchment were generally homogeneous.

Figure 5.4 presents mass curves of some selected rain gauge stations in the catchment. The selection criterion for these stations was such that the mass curves of the station with the highest, the station with the medium as well as the station with the lowest R^2 percentage values of the trend lines should be represented. These mass curve plots, as observed in the figure, show that the plotted values are clustered along a single straight line; in this case the trend line. There were no marked deviations of any of the plotted values from the straight line. This shows that the stations had highly homogeneous data since plotted data are clustered close to the trend line.

From visual inspection of the mass curves, the data appears not to have any inconsistency. This is corroborated by the high values of R^2 as presented in Table 5.7. The rainfall data over the Nyando catchment may therefore be declared homogeneous on account of the very good fit of the trend line to the observations in all cases.

Figure 5.5 shows the mass curve for daily average areal rainfall over the Nyando catchment and Figure 5.6 shows the mass curve for daily average discharge at gauging station 1GD03 respectively.

Visual inspection of the plotted daily values gives the impression that the values are fairly well clustered along a single straight line, as there are no obvious deviations from the

single straight line. This is a possible indication that the daily average areal rainfall and the daily average discharge data have no inconsistency and could therefore be considered of acceptable quality.

Table 5.7: Results of test of consistency for rainfall data over Nyando catchment as given by R^2 %.

Station No	Station Name	Consistency (R^2 (%))
8935001	Songhor, Kaabirir	97.06
8935033	Nandi Hills Savani	99.68
8935148	Kipkurere Forest Stn.	99.52
8935159	Ainabkoi Forest Stn	99.36
8935161	Nandi Hills Tea Est.	99.64
9034086	Ahero Irrigation Stn.	99.76
9035020	Kipkelion Railway Stn.	99.71
9035075	Kaisugu House	99.68
9035148	Koru Bible School	99.61
9035151	Londiani Entomology	99.72
9035155	Londiani Forest Stn	99.72
9035188	Tinga Monastry	99.78
9035201	Kipkorech Estate	99.79
9035220	Koru Homaline co.	99.49
9035226	Londiani Forest	99.78
9035240	Keresoi forest	99.65
9035256	Malagat Forest Stn	99.69
9035263	Tinderet Tea Est.	99.78

Trend lines were then fitted for each mass curve and the R^2 values evaluated. The high values of R^2, 99.8 % for daily areal average rainfall and 97.7 % for daily average discharge, corroborate the visual impression that the observations are usually clustered along a single straight line. Both sets of data may, therefore, be declared homogeneous on the basis of the high values of R^2, which indicates a very good fit of the trend line to the daily data observations. Closer examination of the figure shows that there are alternate high and low flow cycles that coincide with wet and dry cycles respectively. The sections of the plotted curve above and below the trend lines for both the daily average areal rainfall and daily average discharge data curves indicate these cycles. The cycles seem to repeat after every two to three years.
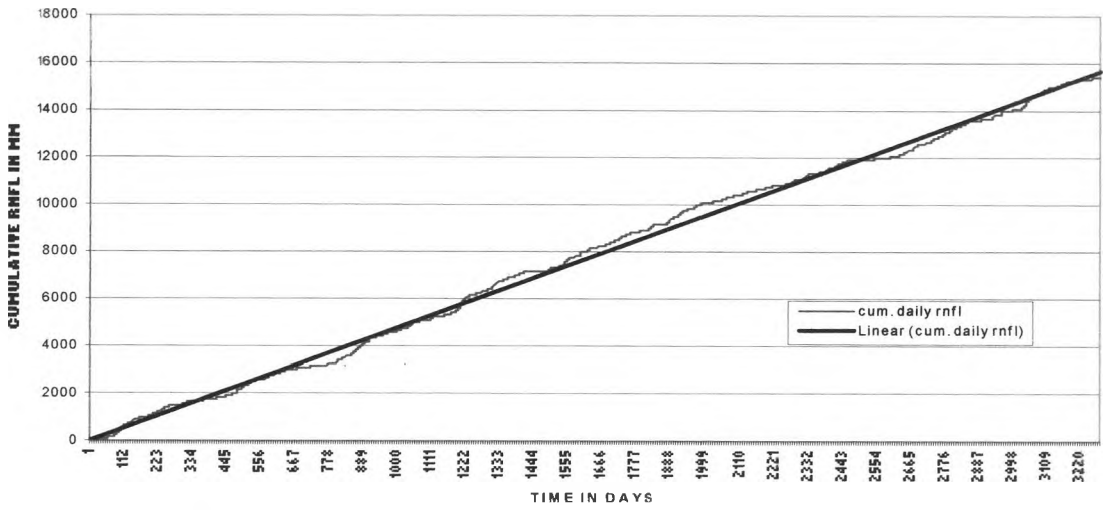
(a)
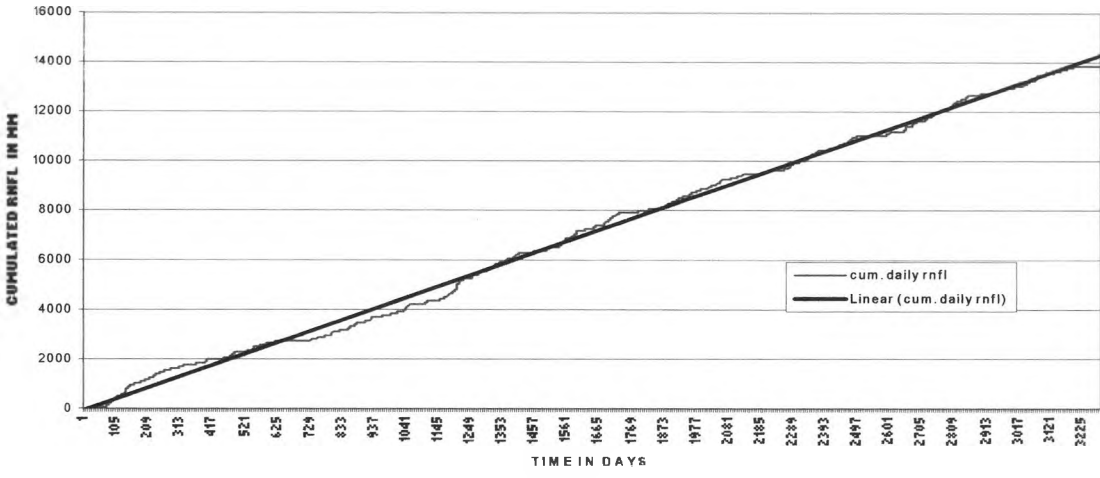
(b)
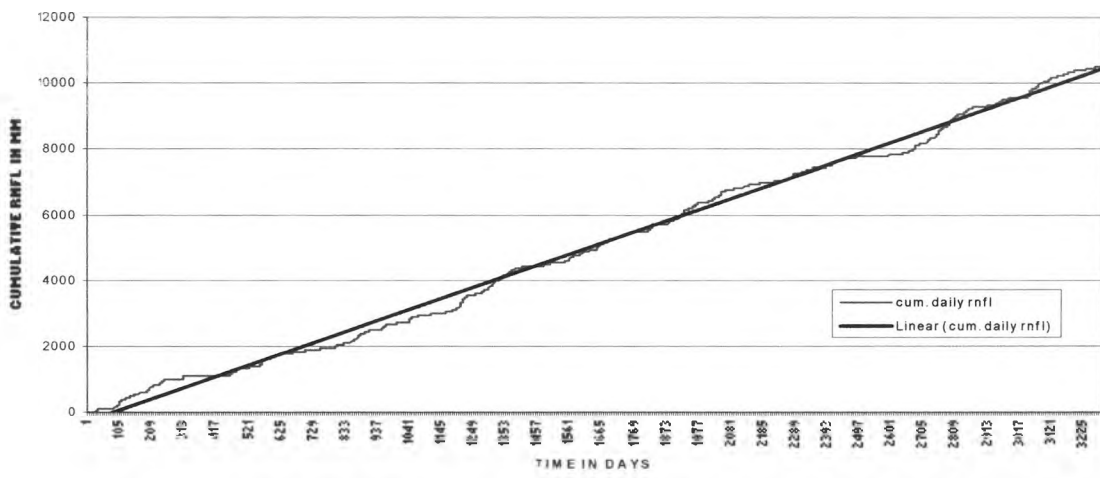
(c)

Figures 5.4 (a – c): Examples of mass curves from some selected rain gauge stations within the Nyando catchment.
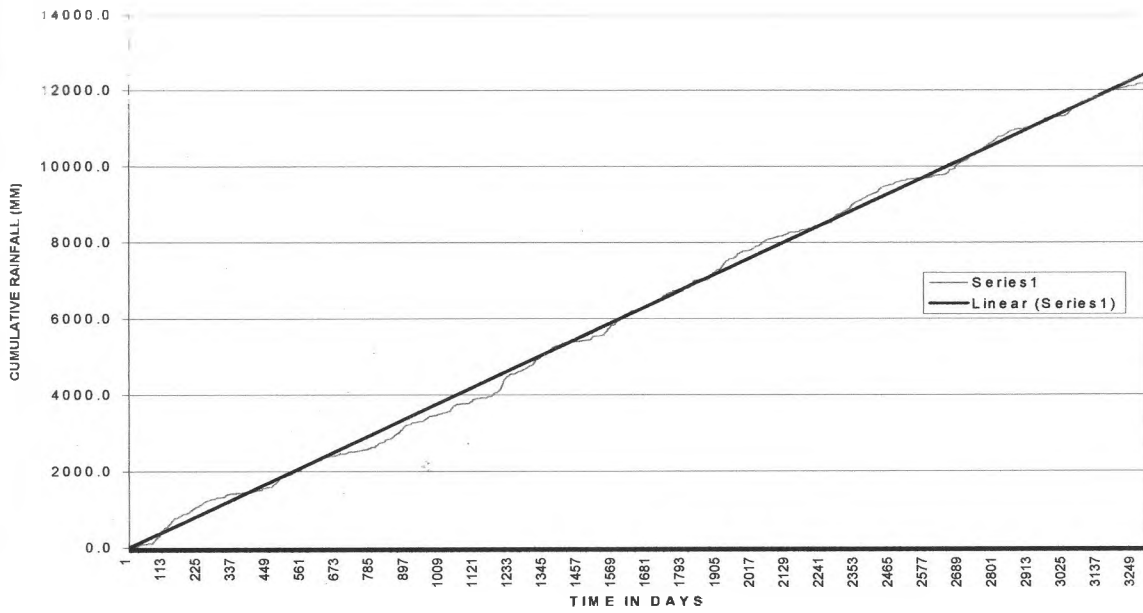
Figure 5.5: Mass curve for daily average areal rainfall to establish consistency of input data over the Nyando catchment.
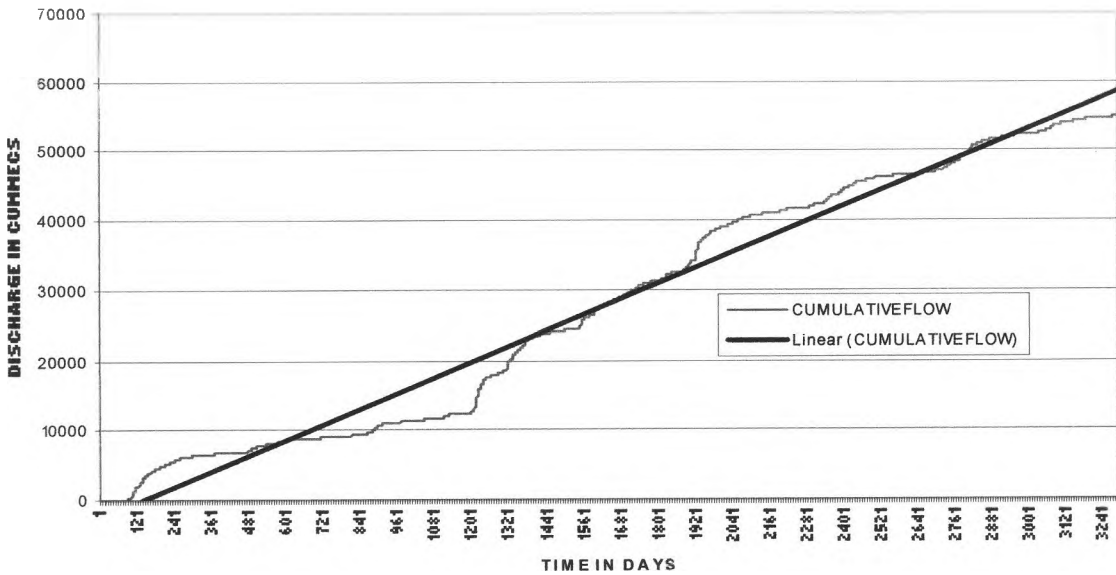
Figure 5.6: Mass curve for daily average discharge to establish consistency of runoff data over the Nyando catchment for discharge values measured at flow gauging station 1GD03.

5.3.2 Evaporation Data

Figure 5.6 presents the results of evaporation mass curve for Kisumu, which was taken as the representative station for the Nyando catchment. The selection of Kisumu was influenced by the fact that it is in the same rainfall zone as the Nyando. Further, evaporation is not known to exhibit significant spatial variation and one station can adequately serve 5000 square kilometres (Ponce, 1989). This is much more than the size of the Nyando catchment, which is about 3580 square kilometers. Hence a single evaporation station would be adequate for the catchment.
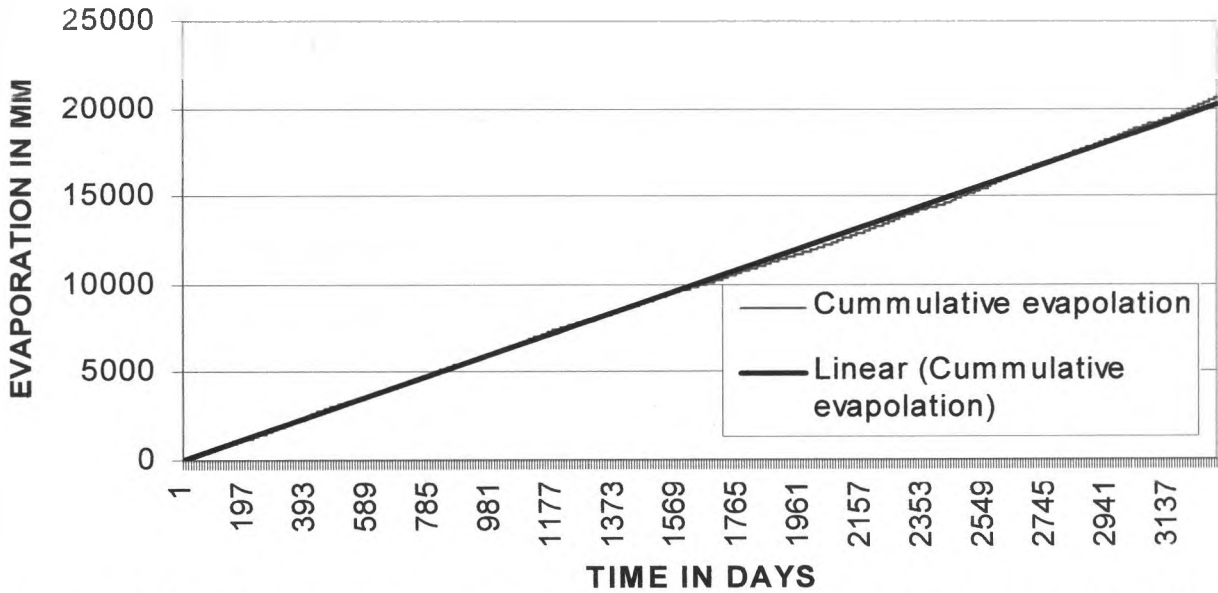
Figure 5.7: Daily Evaporation mass curve to establish consistency of evaporation data for Kisumu station.

The plotted daily values are clustered very close to the single straight line. Indeed it is difficult to visually detect any deviation from the straight line. The trend line was then fitted and R^2 evaluated. It was observed that the curve fits so well that it is difficulty to distinguish the trend line from the plotted daily values. This impression is corroborated by the very high value of R^2 (99.92 %). Evaporation data may therefore be declared homogeneous on the basis of these results.

5.3.3 Estimation of Fourier Series Coefficients

The Fourier series representation given in Equation 4.24 is used to remove any small sample fluctuations from seasonal mean values obtained from Equations 4.22 and 4.23.

Tables 5.8 and 5.9 show the results obtained from the estimation of Fourier series for the Linear Perturbation Model (LPM).

From these tables we observe that the first four harmonics have accounted for about 51% of the rainfall and about 68 % of the stream flow series as given by the respective values of P_j. The P_j values indicate that there is a higher seasonal variation in runoff than in rainfall data series.

Table 5.8: Fourier coefficients for smoothing the seasonal mean input and output series for the Nyando catchment. A_j and B_j are the Fourier coefficients; P_j is the jth percentage of the total variance of the y_d and x_d series about its mean a_o.

Rainfall series				Stream flow series			
Number of harmonic	Fourier coefficient			Number of harmonic	Fourier coefficient		
	A_j	B_j	P_i (%)		A_j	B_j	P_i (%)
1	-1.80	0.99	28.7	1	-13.53	4.13	32.8
2	-1.08	-0.52	10.3	2	-4.23	-7.30	11.6
3	0.74	-0.62	6.6	3	9.75	0.89	15.6
4	-0.22	0.83	5.6	4	-5.34	4.76	8.4
5	-0.43	0.51	3.1	5	-0.90	-2.27	1.0
6	-0.06	0.06	0.1	6	1.58	-1.93	1.0
7	0.36	0.38	1.9	7	-0.79	5.10	4.4
8	-0.06	-0.16	0.2	8	-0.86	-1.68	0.6

Table 5.9: Proportion of the variance of the unsmoothed seasonal input and output series accounted for by each of the first four harmonics and by their sum for the Nyando catchment.

Series	Harmonics				Sum of 1st 4 harmonics
	1st	2nd	3rd	4th	
Input % variance	28.7	10.3	6.6	5.6	51.2
Output % variance	32.8	11.6	15.7	8.4	68.5

Figures 5.8 (a and b) shows the smoothed and unsmoothed daily seasonal expectations of output (stream flow) series and input (rainfall) series over the Nyando catchment. The smoothing has been done by the first four harmonics. The graphs reveal that Nyando has a trimodal pattern of rainfall with the major peak, the long rains, centred in April followed by another one in august. A minor peak centred in October-November is also visible. The

stream flow follows the same pattern but lags behind rainfall by a few days as seen from the graphs.

It is also observed that there is a definite descending pattern in the proportion of variance accounted for by successive harmonics. In the case of stream flow series, harmonics after the fourth one could only account for about 32 % of the total variance of the stream flow series while in the rainfall series it would account for about 49 % which is obviously less than what is accounted for by the first four harmonics. This justifies the use of the first four harmonics for the purpose of smoothing.

5.4 Model Applications

After the models were calibrated and verified, they were then fitted on daily data from the Nyando catchment. Each of the four models was applied to the Nyando catchment using split record evaluation. The models that were used in this study are the Simple Linear Model (SLM), the Linear Perturbation Model (LPM), the Linearly Varying Gain Factor Model (LVGFM) and the Soil Moisture Accounting and Routing (SMAR) model. Details of these models and the methodologies were presented in Chapter 4. The model performance criteria used in each of the linear systems models case include the coefficient of efficiency (R^2), the index of volumetric fit (IVF) and the mean square error (MSE) while (R^2) and volume ratio were used for the conceptual SMAR model.

In terms of complexity, the SLM is the simplest model followed by the LPM, LVGFM and the SMAR model respectively. For SLM, LPM and the LVGFM, which are all linear systems models, the ordinary least squares solution was used for the estimation of pulse response function. In the case of the SMAR model, which is a conceptual model, the parameters were estimated using the simplex search method as embodied in the Galway Flow Forecasting Software (GFFS) package. This is a software package developed at the Department of Engineering Hydrology, National University of Ireland, Galway (O'Connor et al, 2001). It comprises a suit of models for simulation, updating and real-time forecasting applications. The software program uses hydrological data of daily interval only, stored in files written in standard UCG format.

The relative performance of each of the model was evaluated against the model efficiency criteria set out in section 4.5.1. The individual model performance was assessed against the performance of other models as a way of establishing the model that is most suited for the Nyando catchment. The GFFS software package was used in the application

of each model. The results obtained with each individual model and their comparisons are presented in this section.
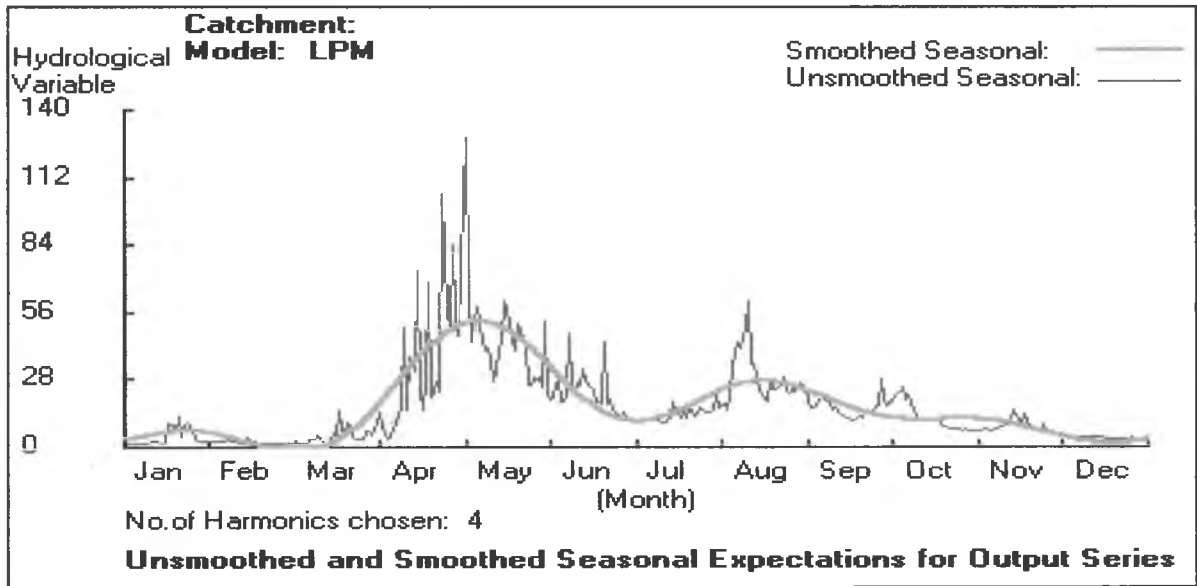
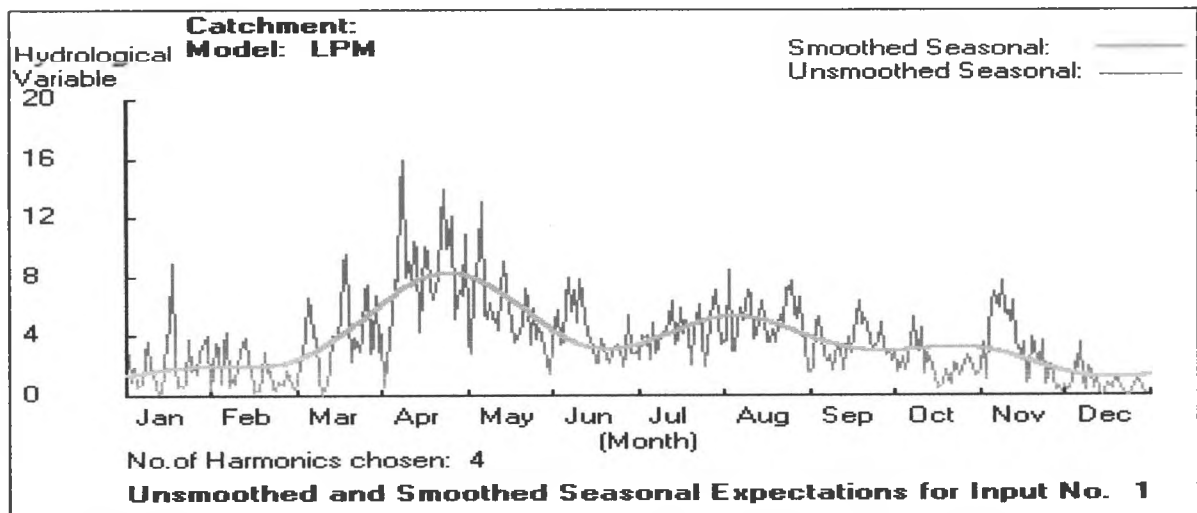Figure 5.8a: Graph of seasonal daily discharge (output) showing the smoothing by the first four harmonics.

Figure 5.8b: Graph of seasonal daily rainfall (input) showing the smoothing by the first four harmonics.

5.4.1 The No Model Situation

As a basis for comparison, it was found necessary to start with a situation where no model is available. This is the seasonal model situation where, in the absence of any model,
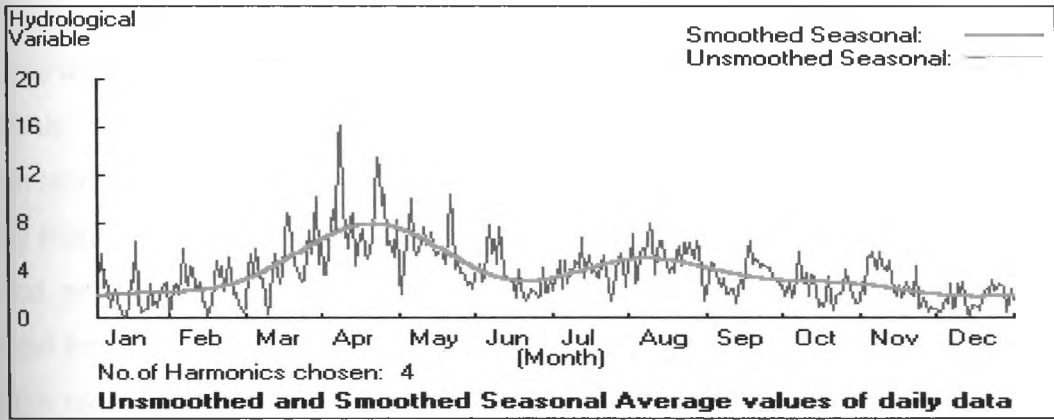
Figure 5.9a: Unsmoothed and smoothed seasonal mean daily rainfall for the Nyando catchment.
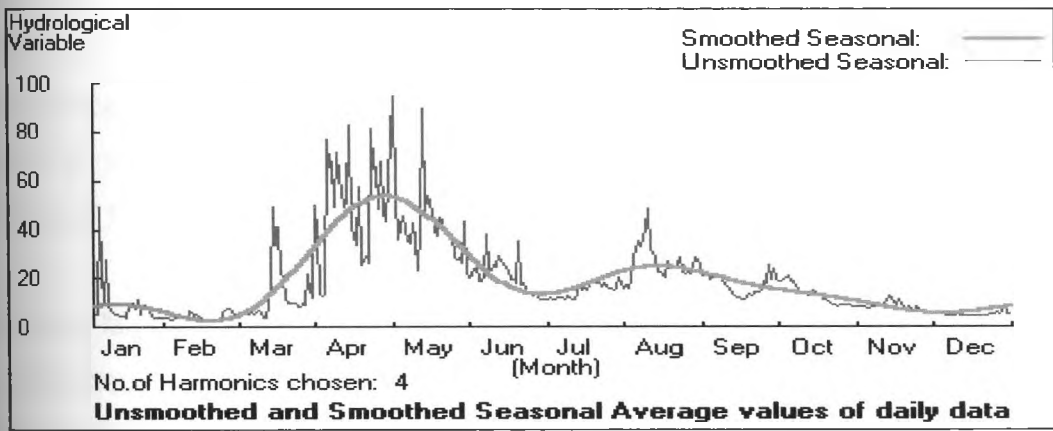
Figure 5.9b: Unsmoothed and smoothed seasonal mean daily discharge for the Nyando catchment.
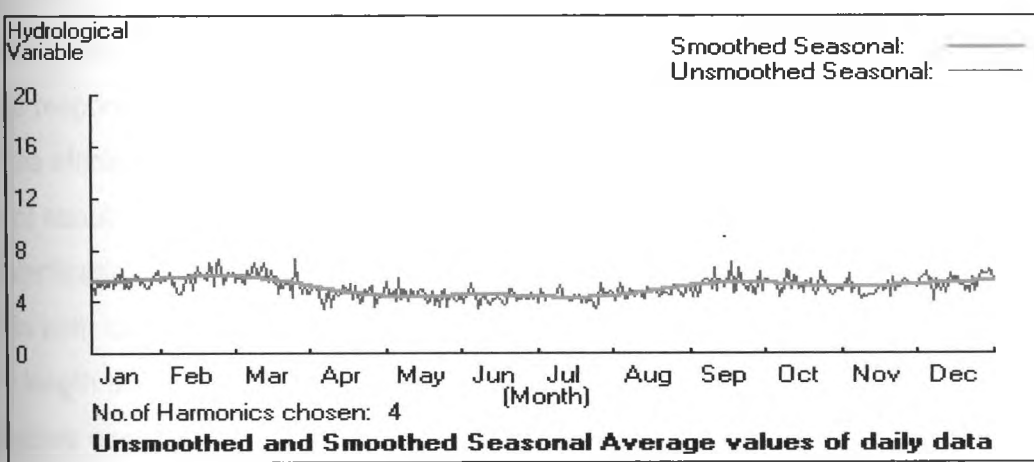
Figure 5.9c: Unsmoothed and smoothed seasonal mean daily evaporation for the Nyando catchment.

This may be taken as a preliminary indication that linear systems models may not perform as well as the conceptual model since they do not include evaporation as one of the model components; yet evaporation seems to have a more significant influence over the flow regime than rainfall in this catchment.

Generally there is lack of strong seasonality in the catchment as there is no clearly distinct dry period, perhaps due to the trimodal pattern of rainfall over the region. There is at least some rainfall throughout the year possibly due to the influence of Lake Victoria to the rainfall around this region. It is observed from Figure 5.9a that except for December through February, the smoothed seasonal daily rainfall is always above 2 mm. Nyando may therefore be considered as a moderately wet catchment. The effect of the land and sea breeze, brought about by the presence of Lake Victoria to the proximity of this catchment, brings about rainfall all the year round over most parts of the catchment. This most probably explains the absence of a strong seasonal variation in the catchment. It is therefore expected that LPM, which is normally expected to perform well in seasonal catchments, may not perform well for the Nyando catchment.

5.4.2 Simple Linear Model (SLM)

The synchronous rainfall-runoff daily data series of the calibration period were processed by the ordinary least squares method using Equation 4.20, which was described in detail in Section 4.1.1, to obtain the optimum SLM pulse response ordinates. The resulting pulse response function was fitted to the corresponding rainfall series to obtain the estimates of runoff series for both calibration and verification periods. The model was calibrated on daily data by ordinary least squares method. The memory length (m), the number of pulse response ordinates, was chosen by trial and error. After several trials, it was found that the efficiency of the model increased significantly with memory length up to a memory length of about 20 days that gave the best results in terms of R^2 efficiency for both calibration and verification periods. Beyond this value there was no marked improvement in the efficiencies in both calibration and verification periods that would warrant increasing the chosen memory length beyond 20 days.

The procedure was to try out memory lengths in steps of 5 days in order to determine how the efficiency changes with increasing memory length. The results of fitting the SLM on daily data from the Nyando catchment are presented in three formats.

Table 5.10: Summary of results obtained with the SLM for different memory lengths in simulation mode. R^2 is the measure of model efficiency; IVF is the index of volumetric ratio and MSE is the mean square error.

Memory length (days)	Model Performance					
	Calibration Period			Verification Period		
	R^2 (%)	IVF	MSE (10^3)	R^2 (%)	IVF	MSE (10^3)
1	10.6	0.75	6.66	-6.3	0.53	9.74
4	32.2	1.02	4.24	20.7	0.72	7.11
5	35.2	1.07	4.00	24.0	0.76	7.02
6	37.6	1.10	4.05	25.7	0.78	7.05
9	42.8	1.19	3.74	28.1	0.84	6.99
10	44.7	1.21	3.39	27.5	0.86	7.05
11	45.1	1.22	3.35	28.1	0.87	7.02
15	46.7	1.26	3.26	28.9	0.89	7.08
19	46.9	1.26	3.20	28.8	0.89	7.07
20	47.0	1.26	3.21	28.8	0.89	7.08
21	47.0	1.26	3.21	29.0	0.89	7.05
25	47.6	1.28	3.13	29.7	0.91	7.07

Table 5.11: Summary of results obtained with the SLM in simulation mode at the chosen memory length of 20 days.

Test period	Model Performance			
	Non parametric mode			Parametric mode
	MSE	IVR	R^2 (%)	R^2 (%)
Calibration	3.49	1.26	47.0	42.9
Verification	7.71	0.89	28.8	28.8

These are: a table of summary of result in terms of R^2, MSE and IVF as presented in Tables 5.10 and 5.11, which give model performance results in simulation mode for various memory lengths, pulse response ordinates as presented in Figure 5.10, and graphs of estimated and observed discharges as given in Figure 5.11.

From Table 5.10, it is observed that of the various values of chosen memory length, 9 days comes out as a possible memory length over the Nyando catchment, barring any instabilities. For this value of memory length, the efficiency during verification is 28.1 %,

which is higher than that at 10 days before starting to rise again. The mean square error is also less on this day than on the eighth and tenth days respectively. This was also confirmed by the comparison of standardized pulse response ordinates with the standard error. The 9th ordinate was found to be the first ordinate where the standard error exceeded the estimated value of the pulse response ordinate. This effectively means that the effect of rainfall on the discharge ends on the 9th day. However, possibly due to some form of instability, another memory length is observed around the twentieth day. It is observed from Table 5.10 that increasing values of m beyond 20 days may not bring about any significant improvements in efficiency during calibration and verification periods, respectively. It was therefore decided that 20 days be used as the chosen memory length for the Nyando catchment on the basis of the high values of R^2 obtained at this memory length.

Table 5.11 gives a summary of results from the simple linear model, SLM, in both parametric and non-parametric form, for the chosen memory length of 20 days while Figures 5.10a and 5.10b give the results in graphical form. From the table, it is observed that the model efficiencies in non-parametric mode are about 47 % and 29 % for calibration and verification periods, respectively. In parametric mode, the model efficiencies are about 43 % and 29 % in the calibration and verification periods respectively. This indicates a decline in the performance during calibration in parametric mode while no change is observed in the verification period. Considering that the SLM is classified as a preliminary model, these efficiencies may be considered a fairly good performance in the absence of any other model. These results indicate that the SLM would serve as a fair flow-forecasting tool for the Nyando Basin in the absence of any other model or in very exceptional cases. This conclusion is taken on the basis that performance values of SLM are comparatively better than those of the no model situation.

There is a slight discrepancy in the volume match as given by the index of volumetric fit (IVF) values. The values differ slightly from the desired value of unity in both calibration and verification periods. The value of 1.26 in the calibration period suggests that the model tends to overestimate the flow volume during the calibration period while underestimating the flow volume, as given by the value of 0.89, during the verification period.

The mean square error values are about 3.2 and 7.1 in calibration and verification periods respectively. This is about 7.4 % and 26.7 % of the mean values during the calibration and verification periods, respectively. Since this is basically a measure of

standard deviation, we may conclude that the estimated values do not deviate much from the observed values.

The efficiency of the model is below 50 % for both calibration and verification periods. It is therefore possible that the linear hypothesis alone may not be adequate for modeling rainfall-runoff transformation over the Nyando Basin. There may be need, therefore, to try out other models that attempt to capture what the SLM fails to capture. This is the temporal variation of the gain factor as well as the seasonal variation of rainfall and runoff. The temporal variation of the gain factor, which is determined by the temporal variation of rainfall, may be accounted for by incorporating a variable gain factor term through the LVGFM while seasonal variation may be accounted for by a seasonal component through the LPM. The gain factor, the ratio that measures the volume of rainfall that transforms to runoff, increases with time as rainfall continues.

Figure 5.10 presents the pulse response ordinates in both the constrained and unconstrained form. It is observed from the figure that the unconstrained pulse response ordinate estimates generally decrease in value from left to right. However they are not especially satisfactory in shape. It is observed that the first ordinate is negative and there is lack of a clear recession as one moves to the right.

The figure also shows that there is a tendency to oscillate every four days or so. This is a contradiction of a hydrologic system that implies high damping, stability and absence of feedback. From the shape of the pulse response function curve, it may be concluded that there is some form of storage in the catchment, which is responsible for the feedback.
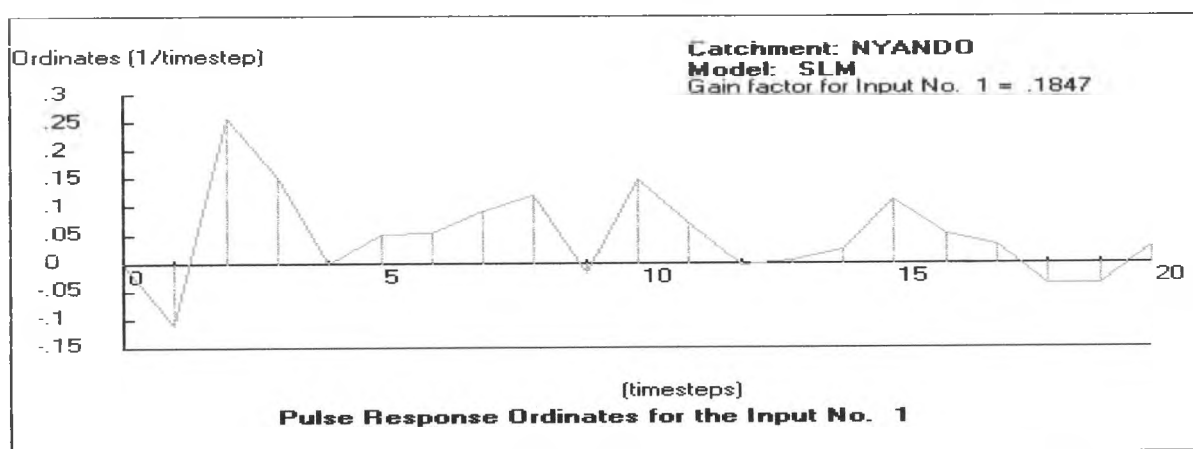
Figure 5.10a: Graphical representation of the pulse response ordinates over the Nyando catchment for Non-parametric form of SLM.
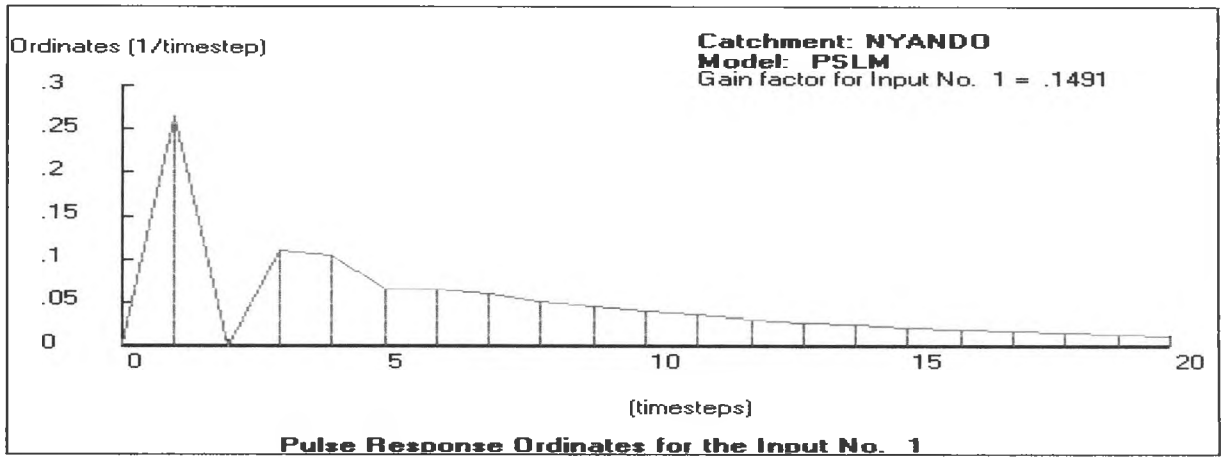
Figure 5.10b: Graphical representation of the pulse response ordinates over the Nyando catchment for parametric form of SLM.

This form of storage may be associated with the availability of soil moisture brought about by either the availability of rainfall throughout the year, the low gradient that slows down the flow velocity or vegetation cover that slows down loss of soil moisture through evaporation. All these factors combine to bring in a form of instability as is evident from the lack of a clear recession in the unconstrained pulse response ordinates.

The undesirable features of this unit hydrograph could possibly be eliminated through imposition of constraints on the shape of the response functions in order to eliminate the fluctuations in the estimated pulse response through smoothing. There is a clear recession when the shape is constrained as shown by the parametric simple linear model (PSLM) hydrograph. Imposing the constraint on the shape of the derived pulse response ordinates normally produces little loss in model efficiency as shown in Table 5.11.

Figures 5.11 (a –f) show rainfall together with observed and estimated discharges for the years 1985 to 1990. It is observed from these figures that the SLM captures the extreme flows very well, as displayed by the peaks and troughs in these graphs. It is evident from these graphs that the peak observed flow coincides with the estimated peak flow. Troughs of the observed flow also coincide with the troughs of the estimated flow.

However, it is also observed from the figure that the model overestimates discharge during low flow while underestimating it during high flow seasons. For instance the maximum observed discharge is about 358 cumecs. This corresponds to the estimated discharge of about 125 cumecs in May 1988 as seen in Figure 5.11d.
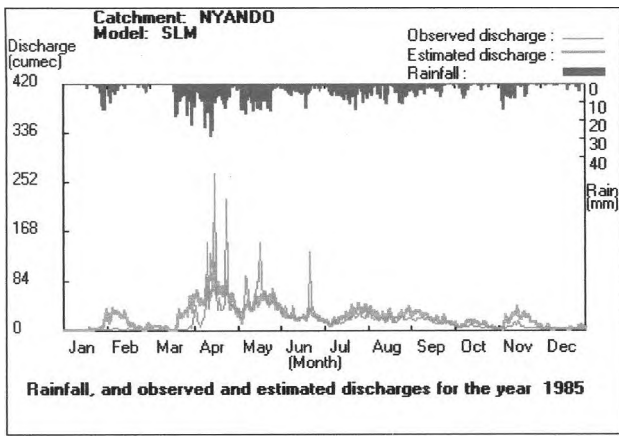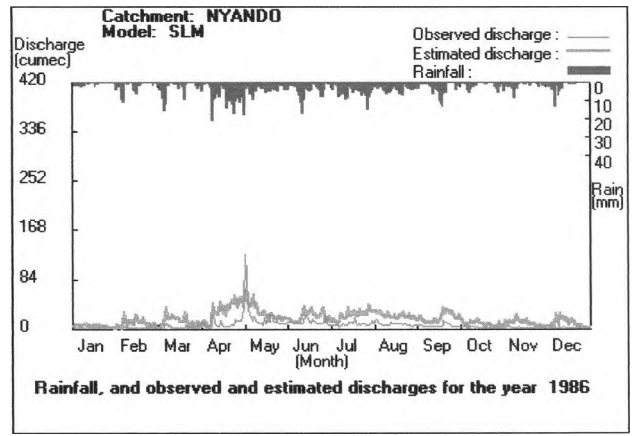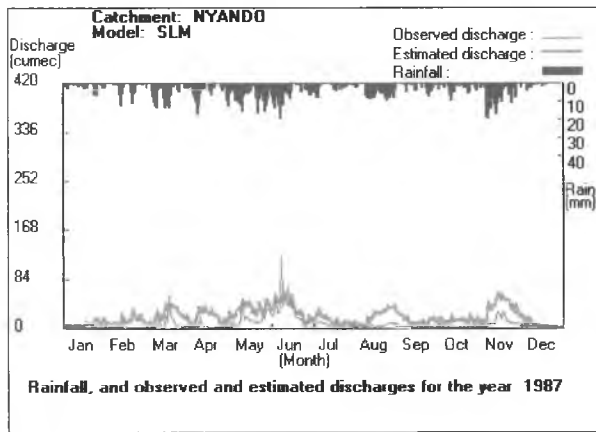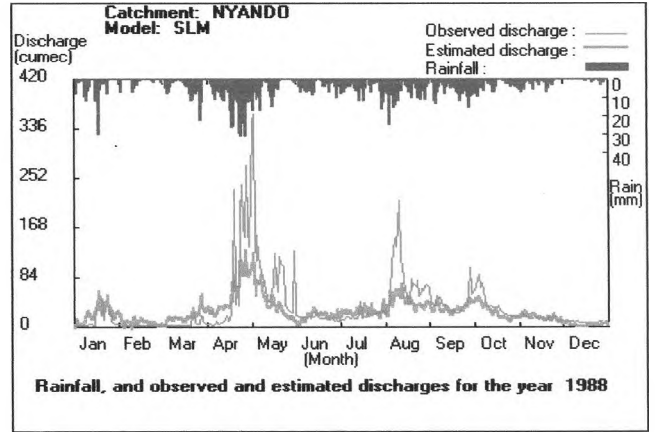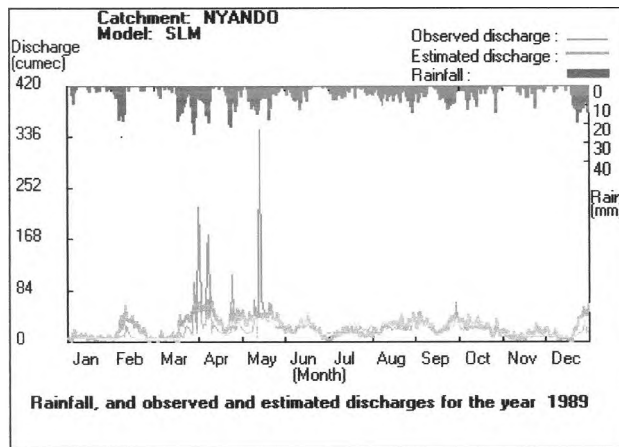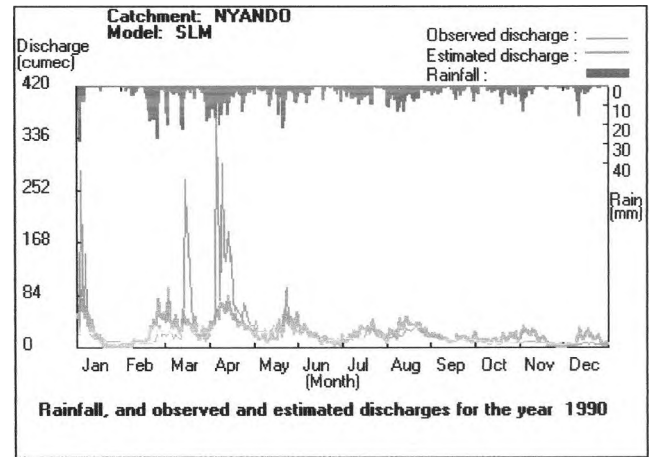
Figures 5.11(a – f): Graphical representation of rainfall; observed and estimated discharges over the Nyando catchment from SLM for the years 1985 to 1990.

The maximum estimated discharge is observed from the figure to be about 127 cumecs corresponding to the observed discharge value of about 274 cumecs. These values occur at the beginning of May 1988 as shown in Figure 5.11d.

The minimum observed discharge is about 1 cummec and corresponds to the estimated discharge of about 5 cumecs. These values occur at the beginning of March 1985 as shown in Figure 5.11a.

The minimum estimated discharge as observed from the figure is about –5 cumecs corresponding to the estimated discharge of about 3 cumecs. These values are found at the beginning of February 1988 as shown in Figure 5.11d during extreme flows.
The difference between the observed and the estimated discharge values is generally not notable except during the high and low flow seasons as shown in Figures 5.11 (a – b).

It is also observed from the figure that except for the years 1986 and 1987, Figures 5.11b and 5.11c, which are documented as El Nino years, where the estimated discharge is, on average, more than the observed discharge, the estimated discharge in the high flow season of March-April-May is less than the observed discharge. The observed discharge graph overlies the estimated discharge graph during the high flow season of March-April-May as observed in the figure. This is an indication that the model tends to under estimate the flow in this season. The graphs of the estimated and observed discharges have a good match for rest of the seasons except for 1988 as seen in Figure 5.11d, where observed discharges are notably higher than the estimated discharges during the high flow seasons.

There is also a general overestimate of flow during low flow seasons except for 1989 and 1990 as observed in Figures 5.11e and 5.11f where there is no marked difference between observed and estimated discharges. This is a definite model shortcoming and may be attributed to the model's inability to capture changes brought about by changes in soil moisture content due to seasonal and temporal variation of rainfall. This suggests a possible need to incorporate a seasonal factor into the model. It could also be attributed to the model's rigidity in the treatment of the soil moisture state of the catchment. The soil moisture state increases during the storm and also during the wet season. This affects the amount of rainfall transformed to runoff. It is expected that the higher the soil moisture content, the lower the rate of infiltration. This is expected to result in a higher proportion of rainfall being transformed to runoff especially during wet seasons than is predicted by the model. The SLM does not account for this temporal variation of soil moisture content leading to the

observed discrepancies. The optimized gain factor was found to be about 0.2, which suggests that only a small fraction of the average rainfall is transformed to runoff.

Other models such as LPM and LVGFM attempt to address the discrepancies observed in the SLM.

5.4.3 Linear Perturbation Model (LPM)

The synchronous rainfall-runoff daily data series of the calibration period were processed by the ordinary least squares method using Equation 4.30, whose details were given in Section 4.4.1.2 to obtain the optimum LPM pulse response ordinates. These were subsequently fitted to the daily rainfall departure series using Equation 4.5, described in Section 4.2.2, to simulate flows for both calibration and verification periods. 20 days were taken as the chosen memory length for the same reasons described earlier as well as the need to have a uniform memory length.

The final estimated runoff series of LPM was calculated by adding the expected seasonal mean runoff to the estimated runoff perturbation series. The difference between observed and simulated runoff were squared and then summed so that R^2 could be calculated using Equation 4.44.

The results for the model's performance in simulation and parametric modes are presented in three formats. These are: a table of summary of results in terms of R^2, MSE, IVF and r^2 as presented in Table 5.12, pulse response ordinates as presented in Figure 5.12 (a and b), and hydrographs of estimated and observed discharges as given in Figure 5.13 (a – f).

The r^2 is a value that is used to assess a given model's improvement over that of the SLM. It gives the value of the proportion of the variance accounted for by the more substantive model, which is not accounted for by the SLM. From the table of the summary of results, it is observed that for a memory length of 20 days, the model efficiencies in non-parametric mode are about 55 % and 25 % during the calibration and verification periods respectively. In parametric mode the efficiencies are about 59 % and 20 % in calibration and verification periods respectively. This indicates a fairly notable improvement during calibration and a fairly notable decline during verification periods.

Comparison of the results of the non-parametric LPM with those of the SLM shows that in the calibration period there is some marked improvement in efficiency, of about 15 % in non-parametric and about 29 % in parametric modes respectively, over the SLM during

the calibration period as shown by the r^2 values which indicate the proportion of the remaining variance unaccounted for by the SLM but subsequently accounted for by the LPM. There is, however a marked decline in efficiency, of about 5 % in non-parametric and about 12 % in parametric modes respectively during the verification period as shown by the negative r^2 values.

Table 5.12: Summary of the results obtained with the LPM in simulation and parametric modes. R^2 is the measure of model efficiency; IVF is the index of volumetric ratio and MSE is the mean square error.

Test period	Model Performance					
	Non parametric mode				Parametric mode	
	MSE	R^2 (%)	IVF	r^2 (%)	R^2 (%)	r^2 (%)
Calibration	2.95	55.0	1.00	15.1	59.2	28.5
Verification	7.71	24.7	0.72	-5.4	20.3	-11.9

When compared with the results of SLM, the LPM is observed to be fairly better than the SLM for the Nyando catchment during the calibration period as shown by the positive r^2 value. However, the LPM performs worse than the SLM during the verification period as indicated by the negative r^2 value, which stands at −5.4 % in non-parametric and about −11.9 % in parametric modes respectively.

The seasonal component of LPM seems to account for a larger percentage of the initial variance during the calibration period than the SLM. The seasonal component accounts for about 15 % of the initial variance during calibration but fails to account for 5 % of initial variance during verification, which was unaccounted for by the SLM. Even though the SLM performed fairly well for the catchment, the combination of the seasonal and the linear component of LPM appear to have improved the results significantly during the calibration period. During the verification period however, there is virtually no difference between the results of SLM and LPM both of which may be considered poor. The poor performance of LPM during verification may be due to the fact that Nyando catchmet does not exhibit highly seasonal variations. LPM is known to perform well in catchments that exhibit highly seasonal variations in rainfall. This is possibly why the results of LPM and those of the SLM are of the same order of efficiency.

The IVF is 1.0 during the calibration period. This is the ideal value and is an indication that the volume of the computed output is about the same as the volume of the observed output.

During verification the IVF is 0.72. Unlike in the calibration period, the IVF is almost always expected to be less than unity during verification period. This is because the seasonal discharge series obtained in calibration is used to obtain the perturbation series in the verification period.
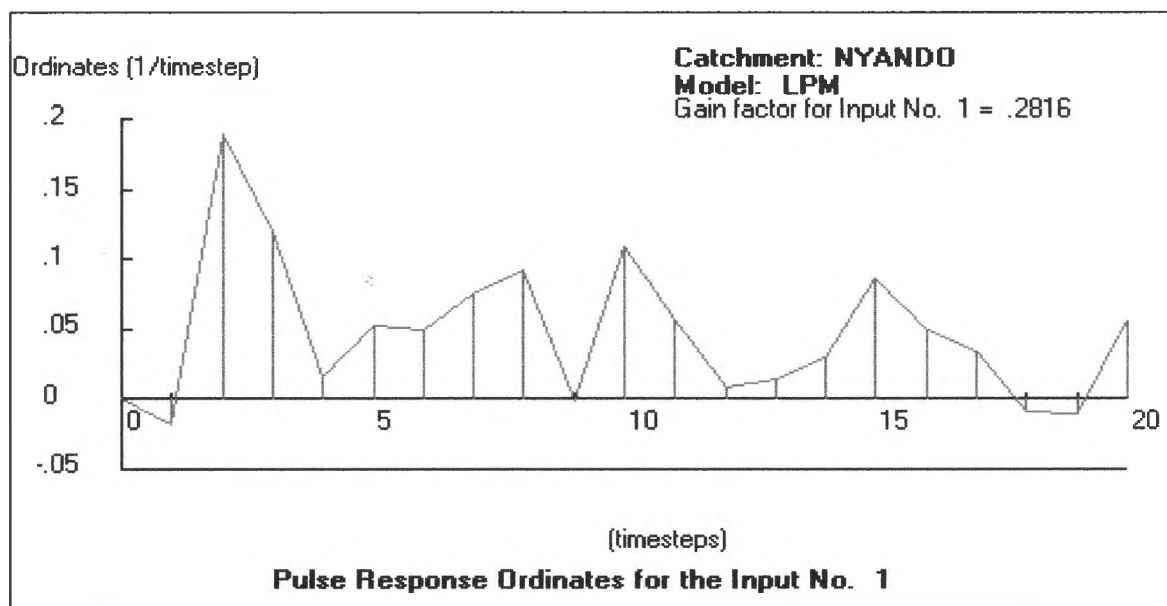

Figure 5.12a: Graphical representation of non-parametric form of the pulse response ordinates over the Nyando catchment for LPM
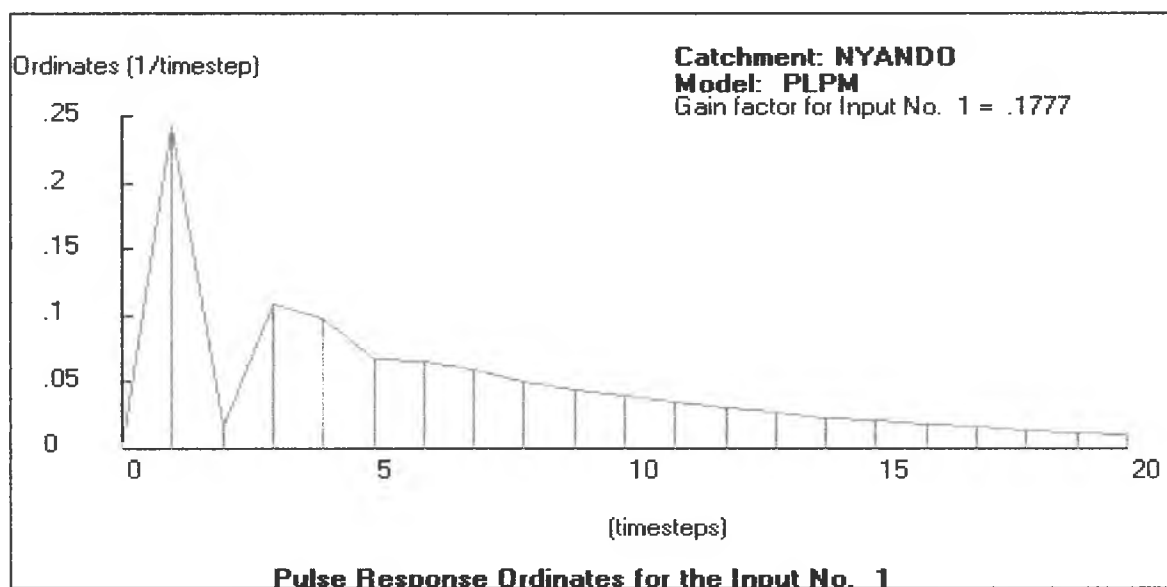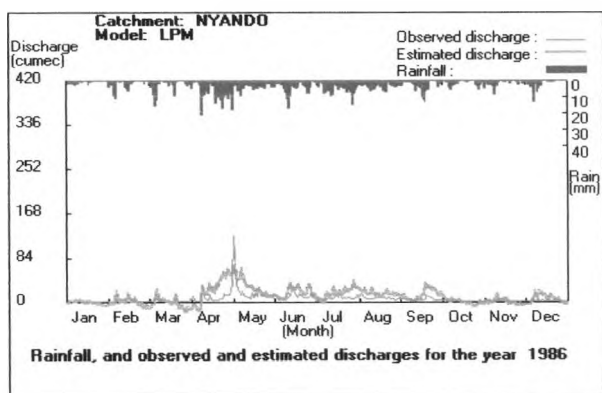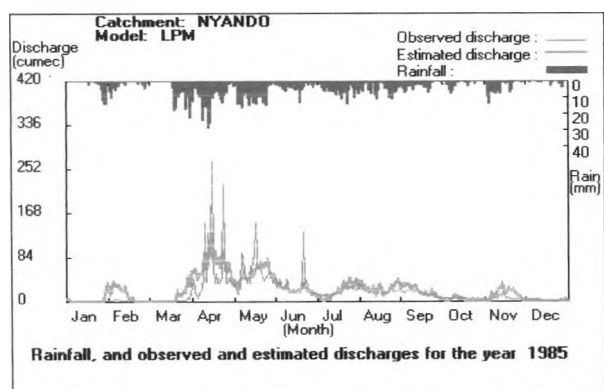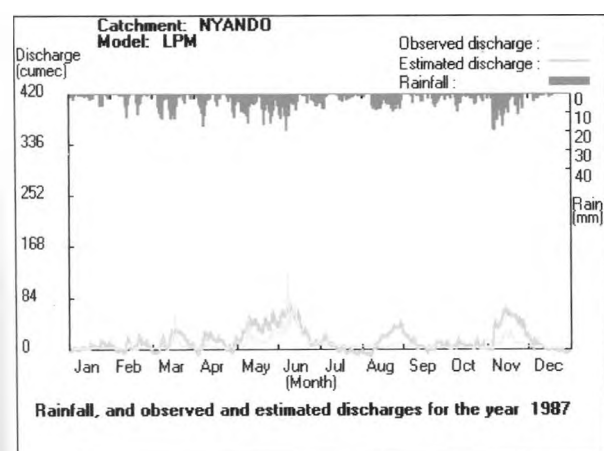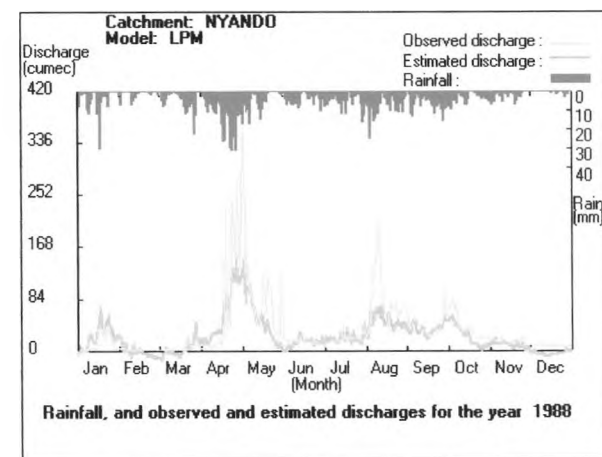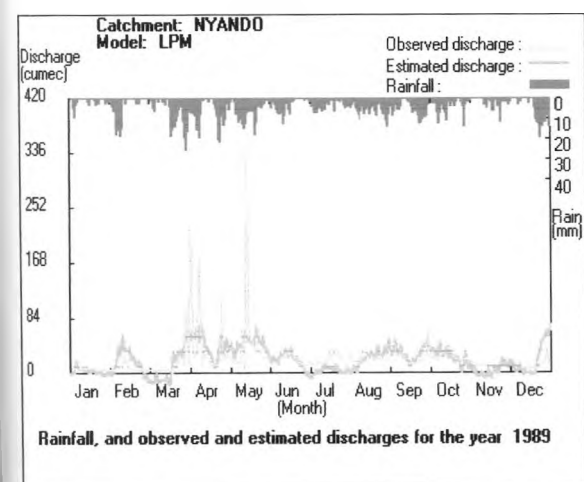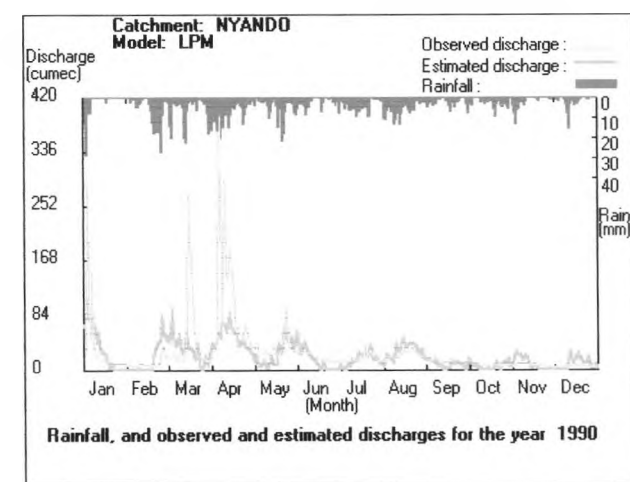

Figure 5.12b: Graphical representation of parametric form of the pulse response ordinates over the Nyando catchment for LPM.

(a)

(b)

(c)

(d)

(e)

(f)

Figures 5.13 (a – f): Graphical representation of rainfall, observed and estimated discharges over the Nyando catchment for the LPM for the years 1985 to 1990.

The resulting sum of perturbations is not necessarily equal to zero as is the case during the calibration period. The results here indicate that the IVF deviates significantly from unity during verification period. Such deviations from unity would possibly be brought about by inadequate estimation of the daily seasonal mean series.

Figures 5.12 (a – f) present the constrained and unconstrained pulse response ordinate estimates obtained with the LPM. On average the unconstrained ordinates, as shown in Figure 5.12a, exhibit a general decrease in size from left to right, which when smoothed take on an exponential shape as expected during the recessional phase.

Just like in the case of SLM, the pulse response ordinates lack a clear recession in non-parametric mode. It is difficult to visually decide which of the two is better than the other as both of them exhibit more or less the same shape. Lack of a clear recession is a demonstration that there could be some feedback resulting from some kind of storage as discussed in the previous section. The lack of a clear recession indicates the problem of instability in estimation is equally severe in both SLM and LPM. It was therefore found necessary to impulse constraints on the shape of pulse responses. There was a clear recession when the shape was constrained as shown by the parametric linear perturbation model (PLPM) hydrograph.

Figures 5.13 (a – f) shows the rainfall together with the observed and estimated discharges for the years 1985 to 1990. It is observed from these figures that LPM follows the same pattern as the one followed by the SLM where it overestimates discharge during low flow and underestimates it during high flow seasons. However the deviations between the observed and estimated discharge values are generally smaller than those of SLM.

It is observed from the figure that the highest observed discharge value is about 358 cumecs and corresponds to the estimated discharge value of about 150 cumecs. This is a slight improvement over the SLM. Both these values are somewhere at the beginning of May 1988 as shown in Figure 5.13d. The maximum estimated discharge is observed from the figure to be about 274 cumecs corresponding to the observed discharge of about 146 cumecs.

The minimum observed discharge is about 1 cummec and corresponds to the estimated discharge of about –22 cumecs. Though the deviations from the observed values are still large, they are observed to be smaller than those obtained with the SLM, an indication that simulated flows from the LPM match the observed flows much better than those from the SLM.

Optimised gain factor for the Nyando catchment was found to be about 0.28 for a memory length of 20 days. This is slightly bigger than that of the SLM, an indication that a slightly bigger proportion of rainfall would be converted to runoff when LPM is used than when SLM is used.

5.4.4 Linearly Varying Gain Factor Model (LVGFM)

Before the application of the LVGFM, it was necessary to first fit the auxiliary SLM to the rainfall-runoff data from the Nyando Catchment in order to determine the catchment wetness index z_i from Equation 4.8. Using the simulated runoff from the auxiliary SLM, the LVGFM was then applied to daily data from the Nyando catchment by following the calibration procedure explained earlier. The memory length of the LVGFM weighting function was taken to be the same as that used for fitting the auxiliary SLM.

The model performance results in simulation mode are presented in three formats. These are: a table of summary of results as presented in Table 5.13, the pulse response ordinates as presented in Figures 5.14 (a – f), and the hydrographs of estimated and observed discharges as given in Figure 5.15.

The results of R^2 percentage of all the linear systems models are displayed in Table 5.14 for the purpose of comparing the performance of these models amongst themselves.

Table 5.13: Summary of the results of application of the LVGFM in simulation mode. R^2 is the measure of model efficiency; IVF is the index of volumetric ratio and MSE is the mean square error.

Test period	Model Performance			
	MSE	R^2 (%)	IVF	r^2 (%)
Calibration	1.83	68.7	0.94	40.9
Verification	7.8	27.9	0.60	-1.3

From the table, it is observed that the R^2 efficiency is about 69 % and 28 % during calibration and verification periods respectively. There is a much better efficiency during calibration than those shown by either SLM or the LPM. This is clearly indicated by the r^2 value of about 41 %. This means that there is about 41 % improvement of the LVGFM over the SLM. This improvement could possibly be attributed to the model's ability to incorporate a variable gain factor as opposed to the SLM, which assumes a constant gain factor. This

suggests that LVGFM is a potential model for the Nyando catchment. In the verification period however there is a decline in performance over the SLM. This is clearly observed in the last column of the table. It appears that the SLM performs better than the other linear systems models during the verification period.

The Index of Volumetric Fit (IVF) values are 0.94 and 0.60 during calibration and verification periods respectively. The large deviation during the verification period reflects the model's tendency to estimate less volume than the observed volume during the verification period.

The mean square errors during calibration are less than those of the other two models. This suggests an improvement over the SLM and LPM respectively during the calibration period. However the mean square error during the verification is slightly higher than in the case of the other two models suggesting a decline in performance.

Figure 5.14 presents the pulse response ordinates as given by the LVGFM. It is observed from the figure that the pulse response ordinate estimates are not especially satisfactory in shape. They present a worse picture than any of the other two models as the decrease in magnitude of pulse response ordinates is not notable. In terms of lack of recession, these ordinates are worse than those of the other models. There is however no negative ordinate as was evident with the other models during translation time.

Figure 5.15 shows rainfall together with the observed and estimated discharges for the years 1985 to 1990. It is observed from these figures that LVGFM fits the data much better than either the SLM or the LPM.
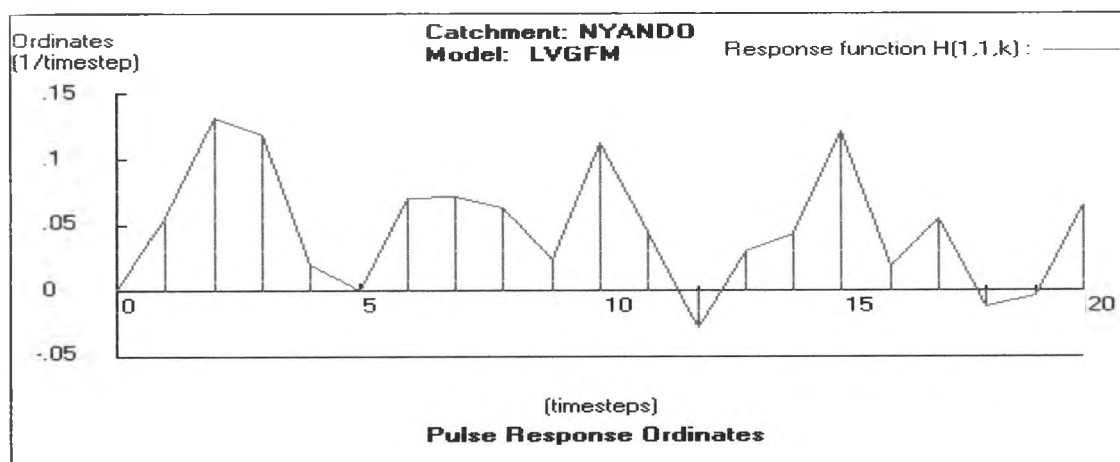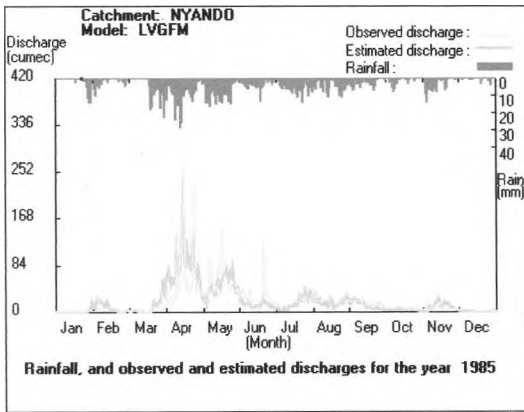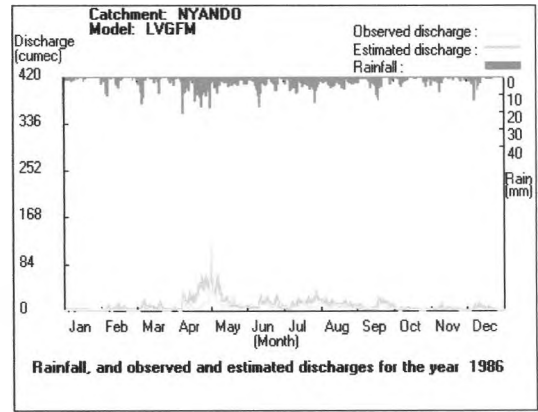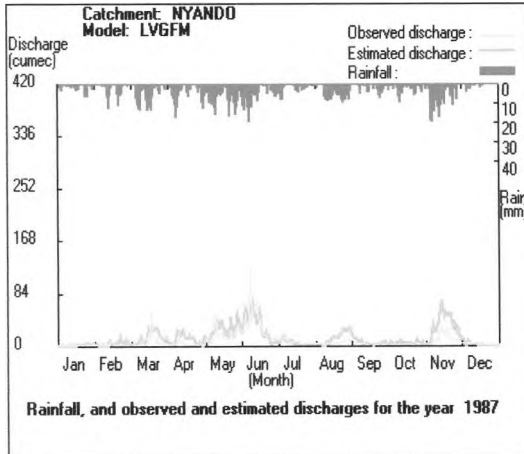

Figure 5.14: Graphical representation of the pulse response ordinates over the Nyando catchment for the LVGFM.
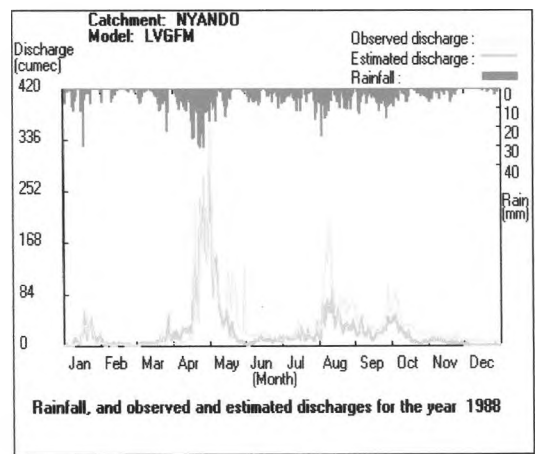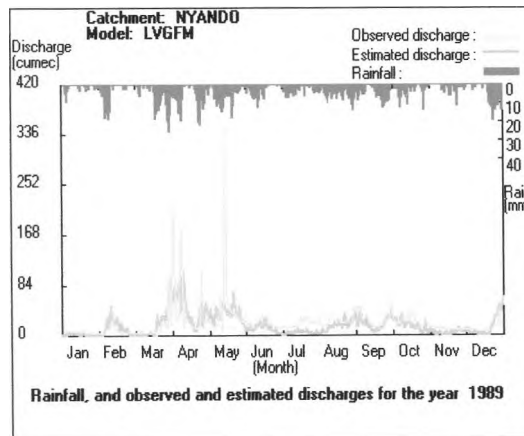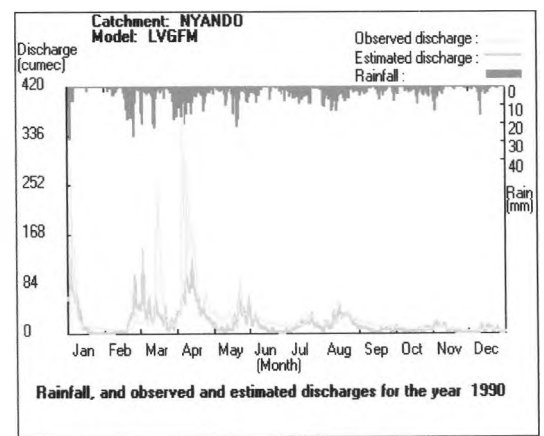
Figures 5.15 (a – f): Graphical representation of rainfall and, observed and estimated discharges over the Nyando catchment from LVGFM for the years 1985 to 1990.

It is observed from the figure that the maximum observed discharge value is about 358 cumecs and it corresponds to the estimated value of about 282 cumecs. The maximum estimated discharge as observed from the figure is about 282 cumecs corresponding to the maximum observed discharge value of about 358 cumecs. These values occur at the beginning of May 1988 as shown in Figure 5.15d.

Table 5.14: Comparison of results of the SLM model with the corresponding results of the LPM and the LVGFM

Calibration				Verification		
Model	SLM	LPM	LVGFM	SLM	LPM	VGFLM
R^2 (%)	47.0	55.0	68.7	28.8	24.7	27.9
r^2 (%)	-	15.1	40.9	-	-5.1	-1.3

The minimum observed discharge is about 1 cummec and corresponds to the estimated discharge of about 0.5 cumecs. The minimum estimated discharge as observed from the figure is about 0 cumecs corresponding to the observed discharge of about 1 cumec. These values are all found at the beginning of February 1988 as shown in Figure 5.15d. These results show that the LVGFM estimated values are much closer to the observed values than in any of the other two models; further suggesting that this model could be more suited for the Nyando catchment than the other two.

The results of SLM and their comparison with those of LPM and LVGFM as presented in Table 5.14 confirm that the linear hypothesis alone is not adequate for modeling the rainfall-runoff transformation for the Nyando catchment. It is observed from the table that the efficiencies of the other models are notably higher than that of SLM during calibration. The r^2 values show that there is an improvement of about 15 % and 41 % during calibration for the LPM and LVGFM respectively.

These results show that the LVGFM performs better than the other two models during calibration when applied to the Nyando catchment. This suggests that other factors besides the linear hypothesis should be considered during modeling. In this case the variable gain factor parameter introduced by the LVGFM and the seasonal component parameter introduced by the LPM. However the SLM seems to perform better than the other models during the verification periods as shown by the negative r^2 values.

5.4.5 Soil Moisture Accounting and Routing (SMAR) Model

After the calibration of the SMAR model, it was applied to the Nyando catchment. The calibration and verification periods are the same as those used for the linear systems models. The final estimated discharge series of SMAR model was then calculated. The difference between observed and simulated runoff were squared and then summed and then R^2 was calculated using Equation 4.44. Results of fitting the SMAR model are presented in Tables 5.15 and 5.16, which give the results of fitting the various model forms for both the calibration and verification periods for the Nyando catchment and in Figures 5.16 and 5.17 (a – f).

A comparison of the results of the SMAR model with the corresponding results of the SLM, LPM and LVGFM was carried out and the results are presented in Table 5.17.

In the applications of this model, the memory length was chosen as 20 days, same as that of the linear systems models, and the warm up period, the time allowed for the model to adapt to the new environment, as 60 days.

The application was done in various configurations depending on the water balance parameters considered. Each of these configurations was considered as a sub-model. These sub-models fall into three categories depending on the number of the water balance parameters they contain. These include those with one parameter only, two parameters only and those with three parameters only while the other parameters are set at their initial values.

The corresponding results are given in terms of the performance index (R^2), the unconstrained objective function value (F) and the ratio of the mean estimated to the mean observed flow (IVF) in both calibration and verification periods as shown in the table.

It may be observed from the table that in the first five trials of one parameter model, each of the five water balance parameters accounts for over 70 % of the initial variance (F_o) during calibration and over 41 % of the initial variance during verification periods respectively. The objective function value in all the five cases is about 0.31 during the calibration period and about 0.9 during the verification period. This value is close to the desired value of zero especially during calibration. It was further observed that the volume ratio (ratio of the estimated to the observed mean discharge or IVF) in all the five cases was about unity, ranging from about 1.00 to 1.03 during calibration and between about 0.6 and 0.7 during verification periods respectively. This ratio shows a close match between the observed and estimated discharge during calibration. However there is a discrepancy

between the observed and estimated discharge during the verification period. The estimated discharge is always less than the observed discharge. This is an indication that the model tends to underestimate flow during the verification period.

It is clear from these results that the water balance parameters C, Z, Y, H and T, when taken individually, are of almost equal importance in their contributions to the model performance.

Table 5.15: Test results of various SMAR model configurations during the calibration and verification periods over the Nyando catchment. C is the parameter that controls potential evaporation; Z is the total storage capacity in the soil layers; Y is the parameter hat represents the maximum infiltration capacity; H is the parameter that represents the available soil moisture content of the first five layers and T is the parameter that represents the estimated pan evaporation.

Trial	Model	Optimum parameter values					Calibration			Verification		
		C	Z	Y	H	T	IVF	F/day	R^2%	IVF	F/day	R^2%
1	C	0.8					1.00	0.31	71.1	0.67	0.87	43.3
2	Z		36				1.02	0.31	71.0	0.69	0.88	42.5
3	Y			50			1.02	0.31	70.8	0.63	0.90	41.2
4	H				0.6		1.03	0.32	70.7	0.63	0.90	41.2
5	T					0.8	1.03	0.31	71.6	0.71	0.88	42.7
6	CH	0.8			0.6		1.00	0.31	71.2	0.69	0.86	44.0
7	ZH		36		0.6		1.03	0.34	68.6	0.66	0.91	40.7
8	CZ	0.8	36				1.01	0.35	67.9	0.64	0.96	37.0
9	CY	0.8		50			1.00	0.34	68.6	0.58	0.94	38.4
10	ZT		36			0.8	1.02	0.31	71.5	0.62	0.94	38.6
11	ZY		36	50			1.03	0.34	68.6	0.67	0.91	40.6
12	CHT	0.8			0.6	0.8	0.98	0.32	70.44	0.56	0.93	39.3
13	CYT	0.8		50		0.8	1.01	0.31	71.0	0.70	0.87	43.1
14	ZYT		36	50		0.8	1.02	0.32	70.7	0.63	0.90	41.0
15	ZHT		36		0.6	0.8	1.09	0.33	69.1	0.70	0.90	41.2

From the table we observe that in the first five trials of one parameter each, the three most sensitive parameters for the Nyando catchment taken individually are C, T and Z which account for about 71.1 % during calibration and 43.3 % during verification periods,

71.6 % during calibration and 42.7 % during verification periods and 71.0 % during calibration and 42.5% during verification periods respectively, of the initial variance.

Considering that the error margins are small, the difference between the parameters is negligible. What this means on the ground is that the rate of evaporation (parameters C and T) may have a bigger influence on the flow regime over the Nyando catchment more than any other model component. The soil moisture content (parameter Z) is also an important factor in the catchment

In the next six trials, two parameters sub-models were tested. The CH combination gave a model efficiency of 71.2 % during calibration and 44.0 % during verification periods respectively compared with ZT with a model efficiency of 71.5 % during calibration and 38.6 % during verification periods respectively. Although ZT combination has a slightly better efficiency during the calibration period, the CH combination was preferred due to the higher efficiency during the verification period. The efficiency of CH combination during verification period was the highest as observed from the table.

Addition of a third parameter in trials 11 to 16 does not improve the results in any notable way. Indeed the introduction of a third term generally lowers the efficiency of the model in both the calibration and verification periods as observed from the table. The best three-parameter sub model combination is CTY with efficiency values of 71.0 % during calibration and 43.1 % during verification periods respectively. This is less than that of CH combination or that of the single parameter C sub-model. It seems therefore, that there are no further improvements in efficiency obtained by the addition of a third water balance parameter. Apparently the extra complexity introduced by the third term does not translate into an improvement in efficiency. It was therefore decided that the evaporation decay (C) and direct runoff (H) parameters of the water balance component be chosen for the Nyando catchment. This resulted in the adoption of the CH combination sub-model.

The results of the chosen model configuration and the optimum water balance parameter values are presented in Table 5.16. It was observed from the table that the unconstrained value of the objective function was about 0.31 mm^2day^{-1}. The model efficiency was about 71.2 % during the calibration and 44.0 % during the verification periods respectively. There is a marked improvement of the SMAR model over the SLM as shown by the r^2 values; about 46 % and 21 % in calibration and verification periods respectively. This improvement, which is obviously larger than that of either the LPM or the LVGFM,

could possibly be attributed to the model's ability to incorporate evaporation as an extra model component.

The volume ratio is about 1.00, which is a perfect fit during calibration and 0.69 during verification periods respectively. The corresponding water balance parameters are C = 0.8 and H = 0.6.

Figure 5.16 presents the unit hydrograph ordinates for the SMAR model. It is observed from the figure that the pulse response ordinate estimates form a smooth exponential decay curve. The shape is quite satisfactory as it shows a clearly stable recession. The clear recession suggests the absence of feed back unlike in the case of the linear systems models. This is what is expected in a hydrological system as it implies high damping and stability.

Table 5.16: The chosen model configuration optimum parameters under volumetric constraint and the results of the SMAR Model. R^2 is the measure of model efficiency; IVF is the index of volumetric ratio and MSE is the mean square error.

Model parameters			Calibration				Verification			
Configuration	C	H	IVF	F/day	R^2 (%)	r^2 (%)	IVF	F/day	R^2(%)	r^2 (%)
CH	0.8	0.6	1.00	0.31	71.2	45.7	0.69	0.86	44.0	21.3

Figures 5.18 (a – f) show rainfall together with the observed and the SMAR model's estimated discharges for the years 1985 to 1990. It is observed from these figures that the SMAR model fits the data much better than the linear systems models. The extremes are captured very well especially in 1988, as seen in Figure 5.16d, where the graph of the estimated discharge follows that of observed discharge very closely. There is no obvious overestimation or under estimation as was observed in the case of the linear systems models.

The maximum observed discharge value is about 358 cumecs. This corresponds to the estimated discharge value of about 210 cumecs. The maximum estimated discharge is observed to be about 274 cumecs and corresponds to the observed discharge of about 160 cumecs. These values are observed in May 1988 as seen in Figure 5.16d.

The minimum observed discharge is about 1 cumec and corresponds to the estimated discharge of about 0.35 cumecs. The minimum estimated discharge as observed

from the figure is about 0.2 cumecs corresponding to the observed discharge of about 2 cumecs.

By and large, the application of the SMAR model has improved the efficiency compared to the SLM and LPM. The model efficiencies increased from about 47 % to 71 % for the SLM and from about 55 % to 71 % for the LPM in calibration.

The improvement over the LVGFM is not notable in calibration. The improvement was from about 69 % to 71 %. In verification however, there is a marked improvement; from about 28 % to 44 %. The results of the LVGFM and those of the SMAR model do not differ much during calibration. This is an indication that the two models are equally good in simulating flow during calibration over the Nyando catchment. However the SMAR model may be considered superior on account of its better performance than the LVGFM during the verification period. In both calibration and verification, it shows a marked over the SLM.

Although the maximum estimated discharge corresponding to the maximum observed discharge is less than that of LVGFM, the SMAR model may still be considered to be superior to the LVGFM model since it has a higher efficiency than the LVGFM during the calibration and verification periods respectively. For example the observed discharge on the 3rd of May 1988 is about 135 cumecs while the corresponding SMAR model estimated discharge on the same day is about 136 cumecs, a close estimate. The deviations between the observed and the estimated discharge values are generally not quite serious with this model compared to those of linear systems models suggesting further that the model has a higher potential than the linear systems models to become the model of choice for the Nyando catchment.

Table 5.17: Comparison of the results of the SMAR model with the corresponding results of the SLM, LPM and LVGFM in simulation mode.

Calibration					Verification			
Model	SLM	LPM	LVGFM	SMAR	SLM	LPM	LVGFM	SMAR
R^2 (%)	47.0	55.0	68.7	71.2	28.8	24.7	27.9	44.0
r^2 (%)	-	15.1	40.9	45.7	-	-5.1	-1.3	21.3

It was observed from these results that the conceptual SMAR model performed consistently well for both calibration and verification periods. It performed much better than the linear systems models; showing a marked improvement in both calibration and

verification periods, respectively, unlike the linear systems models whose performance were lower than that of SLM in verification. It would seem therefore that this model is more suitable for the Nyando catchment than the linear systems models. In the absence of the SMAR model, LVGFM, which does almost as well as the SMAR model during calibration, may be applied. However it should be noted that the model does worse during verification than the SMAR model.
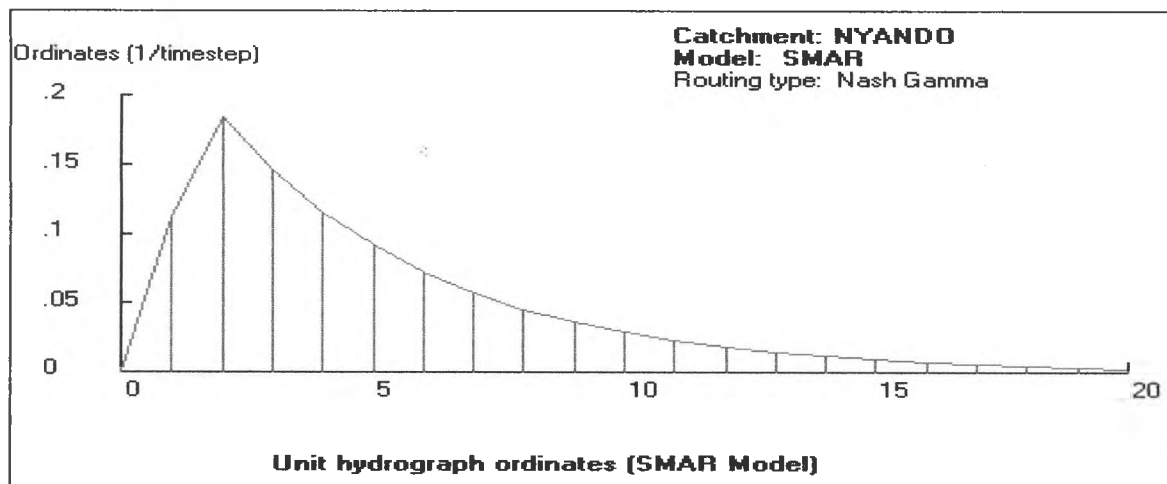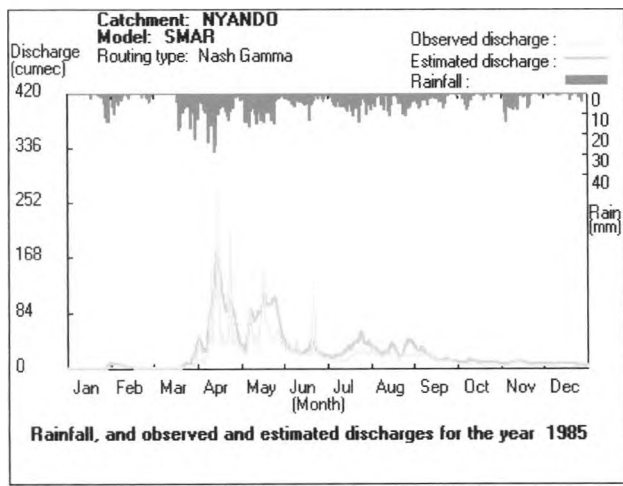

Figure 5.16: Graphical representation of the pulse response ordinates, estimated by the method of constrained least squares, over the Nyando catchment for the SMAR model.

A comparison of the results of the SMAR model with those of the linear systems models, as presented in Table 5.17, show a substantial improvement in the SMAR model performance over that of all the considered linear systems models in both calibration and verification periods.

This is shown by the r^2 values, which show an improvement of about 46 % and 21 % over the SLM in calibration and verification periods respectively. These values are higher than those observed with the linear models, especially in verification where the substantive linear models performed worse than the SLM as indicated by the negative values of r^2.
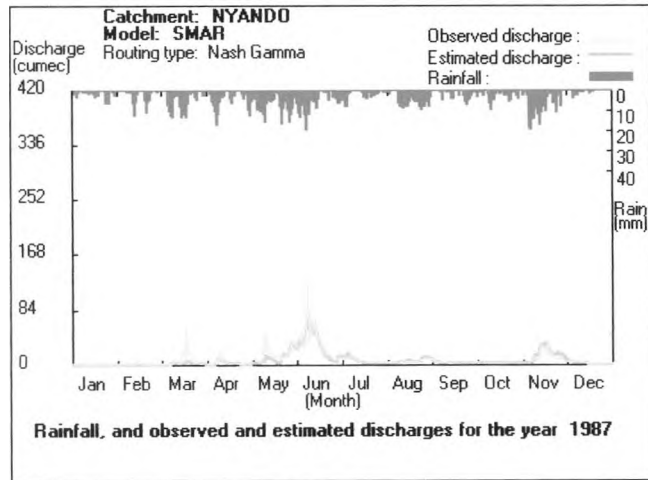
In terms of efficiency, the SMAR model is the best for the Nyando catchment followed closely by the LVGFM, then the LPM and the SLM respectively. It should also be noted that, whereas the efficiencies of the linear systems models are given in unconstrained mode, the efficiency of the conceptual SMAR model is given in constrained mode.
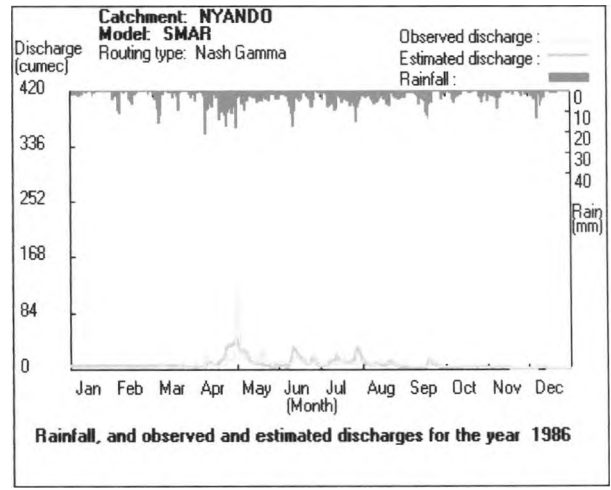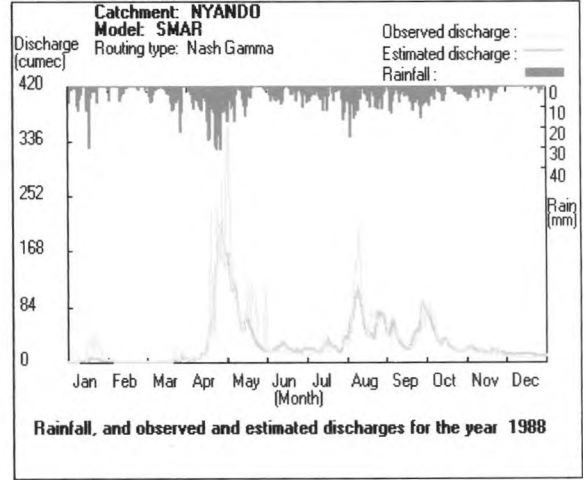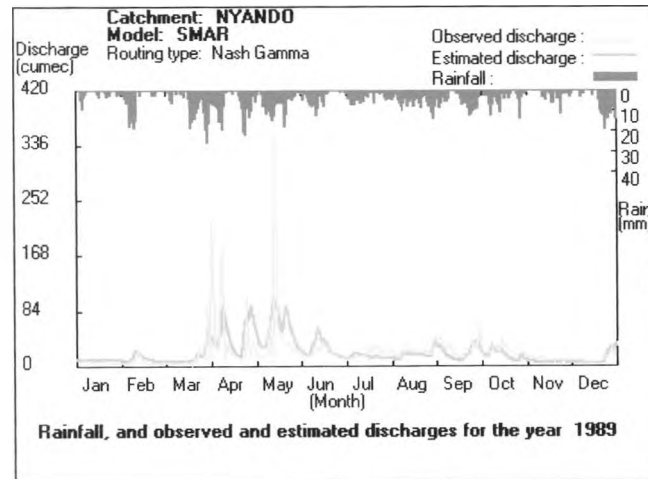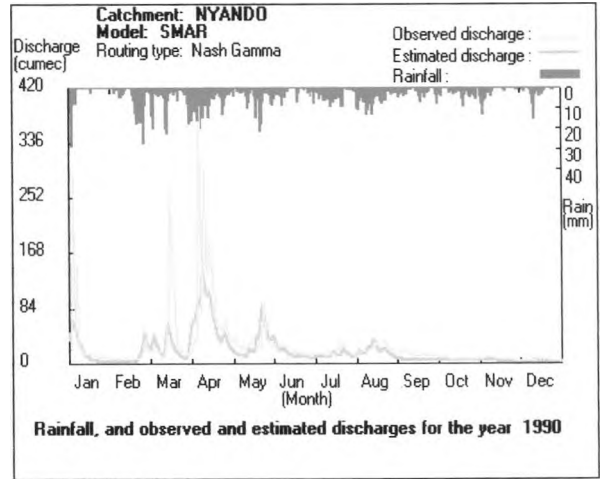
Figures 5.17 (a – f): Graphical representation of rainfall, observed and estimated discharges for the period 1985 to 1990 over the Nyando catchment for the SMAR model.

This means that the conceptual model would have performed even better if it were not constrained.

The ultimate success of any forecasting model depends on its ability to produce good results in simulation mode. So far the results have been given in simulation mode. The efficiencies of the models in this mode suggest that an improvement in the forecasts would be necessary in order to raise the level of forecasting efficiency before applying the models to real time forecasting. These improvements may possibly be achieved through updating of these forecasts. In the next section the performance of the linear systems models in updated mode is considered.

5.5 Autoregressive (AR) Updating Results

The autoregressive updating procedure was discussed in detail in Section 4.5.2.1. This procedure was used in an attempt to improve the forecasts using error corrections in AR. The efficiencies in simulation mode were far from being perfect since even for the best model for the catchment, the SMAR model, the efficiency was only about 70 % in the calibration and only about 40 % in the verification periods. There was need, therefore, to attempt to improve these efficiencies. One of the commonly used methods in forecasting improvements is the autoregressive (AR) updating where errors in the most recent forecast are used to correct the future model forecasts.

Table 5.18 presents the results of the third order autoregressive (AR) updating. This order was adopted since it gave the best efficiency values compared to the other orders. Orders lower than or higher than the third resulted in lower efficiencies than those of the third order.

Table 5.18: Summary of results of the AR updating. Cal stands for the calibration period and Ver stands for the verification period.

del	Non-updated				R^2 % updated						IVF					
	R^2 %		IVF		1day lead		2day lead		3day lead		1day lead		2day lead		3day lead	
	Cal	Ver	Cal	Ver	Cal	Ver	Cal	Ver	Cal	Ver	Cal	Ver	Cal	Ver	Cal	Ver
M	47.0	28.8	1.26	0.89	73.1	42.0	67.0	27.5	63.3	21.7	1.00	1.01	1.00	1.02	1.00	1.02
M	55.0	24.7	1.00	0.72	71.4	37.9	66.5	24.9	62.9	20.3	1.00	1.01	1.00	1.02	1.00	1.02
GFM	68.7	27.9	0.94	0.60	79.0	39.9	77.8	30.0	75.1	26.5	1.00	1.02	1.00	1.04	1.00	1.05

It is observed from the table that a lead-time of one day provides the highest efficiency in all the three models. The efficiency decreases progressively as one moves towards lead-times of more than one day both in calibration and verification periods. The longer the lead-time, the more the details that the model tends to "forget". Hence, the decrease in efficiency of the model performance. The same scenario is also observed in the case of the index of volumetric fit (IVF). The values show that the models tend to overestimate flow during verification as shown by values of IVF that are greater than unity. In this study up to three day lead-time was considered since subsequent lead times gave progressively decreasing efficiencies. Lead-times of more than three days resulted in efficiencies that are less than sixty percent for both SLM and LPM.

Comparing the simulation and updated mode efficiencies, it is observed that there is a marked improvement in updated mode efficiencies over the simulated ones both in calibration and verification periods. The most notable improvement is that of SLM where the efficiency improved from 47.0 % to 73.1 % during calibration and from 28.8 % to 42.0 % during verification periods respectively. This is about 55.5 % and 45.8 % improvement during calibration and verification periods respectively.

In the case of the LPM the efficiency improved from 55.0 % to 71.4 % during calibration and from 24.7 % to 37.9 % during verification periods respectively. This is about 29.8 % and 53.4 % improvement during calibration and verification periods respectively.

For the LVGFM, the efficiency improved from 68.7 % to 79.0 % during calibration and from 27.9 % to 39.9 % during verification periods respectively. This is equivalent to about 15.0 % and 43.0 % improvement during calibration and verification periods respectively. The improved forecasting efficiency is mainly due to the fact that the updating procedure largely eliminates the final forecasting errors.

Only the results of one-day lead-time are discussed in comparison to the simulation mode results since this is the lead-time that has the highest efficiency. Moving from one day lead time through to three days lead time, the efficiency decreases progressively from about 73 % to 63 % for SLM during calibration and from about 42 % to 22 % during verification. For the LPM, the efficiency decreases from about 71 % to 63 % during calibration and from about 38 % to 20 % during verification. In the case of LVGFM, the decrease in efficiency is from only about 79 % to 75 % during calibration and from about 40 % to 27 % during verification.

Figures 5.18 (a – c) present examples of results of estimated and observed discharge in updated mode for lead times of one and two days. The first row represents results of the simple linear model (SLM), the second row represents results of the linear perturbation model (LPM) and the third row represents results of the linearly varying gain factor model (LVGFM). It is observed from this figure that the maximum estimated discharge for SLM in updated mode is about 316 cumecs and 219 cumecs at one and two days lead times respectively. The maximum estimated discharge for LPM in this mode is about 334 cumecs and 232 cumecs at one and two days lead times respectively. The corresponding estimates for LVGFM are about 377 cumecs and 302 cumecs at one and two days lead times respectively.

It is clear from these observations that the deviations from the observed discharge are smaller compared to those observed in simulation mode. The smallest deviation occurs in one-day lead-time. Though the deviations increase proportionately with increase in the lead-time they are still smaller than those observed in simulation mode. Hence the updating seems to have improved the forecasts quite significantly, as they are closer to what is observed. These results show that the three models gave better forecasts in updated mode for all the lead times considered than the simulation mode forecasts.

In terms of the index of volumetric fit (IVF), it is observed that the value is 1.00 during the calibration and about 1.01 during the verification in all the time leads. These values show a significant improvement, a close match between the volumes of estimated and observed discharge, from the simulation mode values irrespective of the lead-time. These improvements are corroborated by the improved efficiency given by the R^2 values.

Overall, the updating resulted in improved forecasting efficiency over that of simulation mode in all the models both in calibration and verification periods. During calibration, SLM had the highest improvement of about 56 % while LVGFM had the least improvement of about 15 %. On the other hand, LPM had the highest improvement of about 53 % while LVGFM had the least improvement of about 43 % during verification. It is quite evident from these results that on average, the improvement during verification was high in all the models raging from about 43 % to 53 %. This contrasts with the improvement during calibration that ranges from about 15 % to 56 % with the only significant improvement being that of SLM.
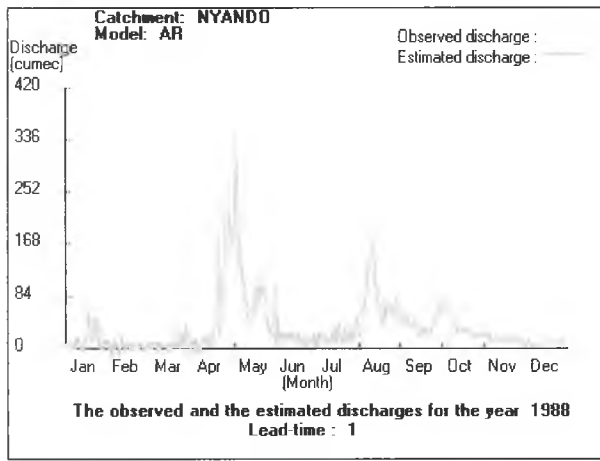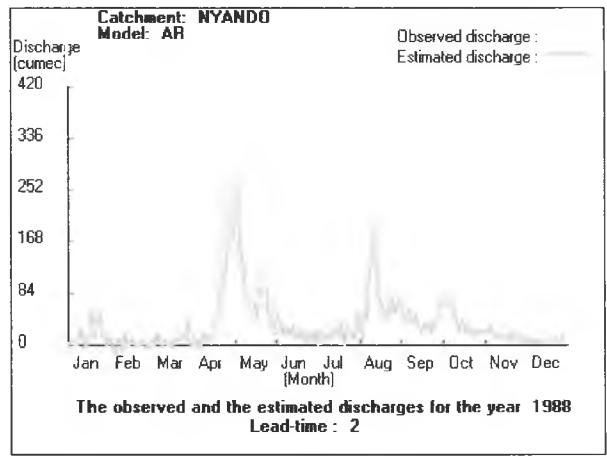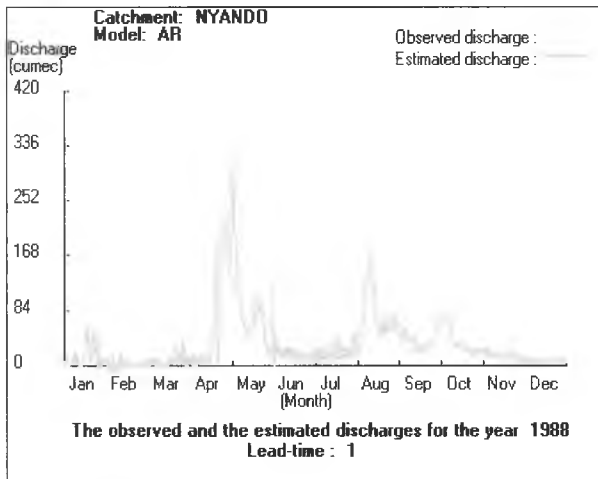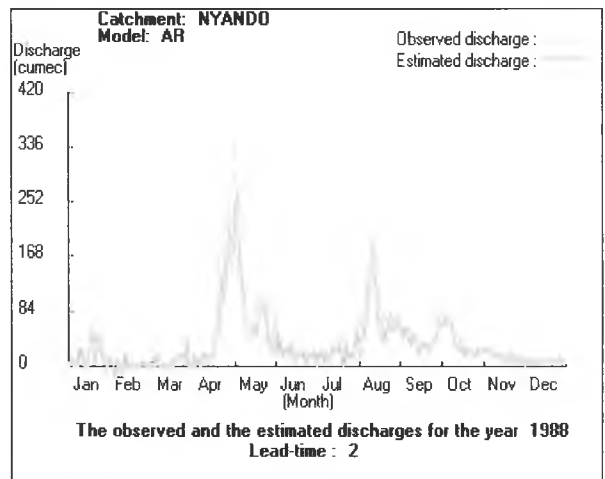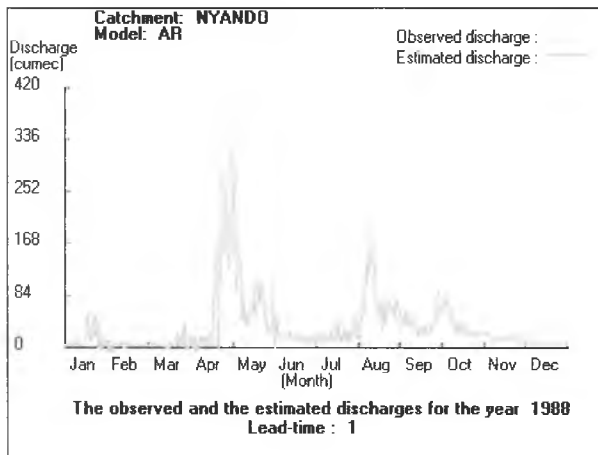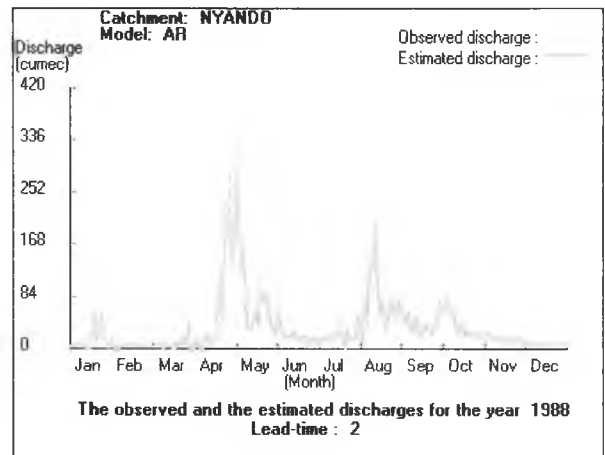
Figures 5.18 (a – f): Updated forecasts for the SLM, LPM and the LVGFM over the Nyando catchment at lead times of one and two days respectively.

From these results, we may conclude that generally the individual model's forecasts in updated mode are better than the forecasts in simulation mode. It is therefore worthy to update the models since all the one-day lead forecast efficiencies are over 70 % during the calibration as it is observed from the table. It is also observed that SLM seems to be the one most responsive to the updating process. In updated mode, the efficiencies of SLM are actually superior to those of LPM in all the lead times both in calibration and verification as observed in Table 5.19.

It is observed from Table 5.19 that though the model performance progressively decreases with the length of lead-time, it generally remains high at above 50 % even at the lead-time of about six days. Of particular interest are forecasts of LVGFM, which remain at over 70 % throughout the six days time lead. This suggests that this model could be used to forecast flow and consequently flood episodes in the Nyando catchment with a lead-time of up to six days. This would give adequate time for possible evacuation of the affected communities or any other corrective measures.

Table 5.19: Summary of AR updating R^2 results in calibration.

Model	R^2 % model performance					
	1day lead	2day lead	3day lead	4day lead	5day lead	6day lead
SLM	73.1	67.0	63.3	59.6	57.1	53.1
LPM	71.4	66.5	62.9	58.6	56.9	52.8
LVGFM	79.0	77.8	75.1	72.0	71.3	70.1

CHAPTER 6

6.0 SUMMARY, CONCLUSIONS AND RECOMMENDATIONS

The aim of this study was to assess a model, from among some existing ones that could be used to forecast flow over the Nyando catchment. This was done through the assessment of performance level of various models on data collected from the Nyando catchment. A brief highlight of the results and conclusions obtained from this study are presented in this chapter.

6.1 Summary And Conclusions

The data used in this study were generally of high quality as was revealed by the homogeneity test results presented in Tables 5.3 and 5.7, and in Figures 5.4,5.5 and 5.6. From the results of the study, all the rainfall, river flow and evaporation stations had highly homogeneous data. Each mass curve fitted on a single straight line suggesting that each of the three sets of data came from the same parent population. The trend line in all cases was a very good fit to the data with R^2 values of over 99% for both rainfall and evaporation data while that of runoff was about 98%. These values indicated a very good fit of the trend line on all the data sets that were used. On the basis of these very high values of R^2, data over the Nyando catchment was declared homogeneous.

When the models described in Chapter 4 of this study were applied to quality controlled data collected from the Nyando catchment, the results obtained indicated that of the four models applied to the Nyando catchment in simulation mode, the SMAR model showed the highest R^2 efficiency values, about 71 % in calibration and about 44 % in verification periods respectively as shown in Tables 5.16 and 5.17. It was closely followed by the LVGFM with R^2 efficiency values of about 69 % in calibration and 28 % in verification periods respectively as shown in Tables 5.13 and 5.14. The SLM gave the lowest R^2 efficiency value of about 47 % in calibration while the LPM gave the lowest R^2 efficiency value, about 25 % in verification period. This is shown in Table 5.17. Thus the conceptual SMAR model, with the highest efficiency values, 71 % and 44 %, in both calibration and verification periods, is apparently superior to the linear systems models that were tested over the Nyando catchment. Compared with the linear systems models, the conceptual SMAR model performed consistently better than the linear systems models in both calibration and verification periods. This confirmed the assumption that conceptual models,

linear systems models.

Among the linear systems models, the LVGFM performed better than the other two models with R^2 value of about 69 % followed by the LPM at R^2 value of about 55 % and SLM at R^2 value of about 47 % in that order during calibration. The SLM, as indicated by the results of the model application in Table 5.17, is the most inferior of all the models applied to the Nyando catchment in simulation mode during calibration period. During verification period however, the results show that the LPM is inferior to all the other models.

Comparing the results obtained by the LVGFM and those of the SMAR model, the best two models for the Nyando catchment during calibration, we see that they are more or less the same during the calibration period. However, the SMAR model performs much better than the LVGFM during verification period. The SMAR model is thus superior to the LVGFM when both are applied to the Nyando catchment.

In the updated mode, the LVGFM is the best model among the linear systems models, with output efficiency R^2 values close to 80 % in calibration and 40 % in verification. The LPM is found to be inferior to the other linear systems models both in calibration and verification as shown in Table 5.18. However, the SLM with efficiency of about 73 % in calibration and 42% in verification is observed to out perform the LVGFM in verification.

In conclusion this study has identified two important models for the Nyando catchmnet; one linear and the other conceptual. The two models, the conceptual Soil Moisture Accounting and Routing (SMAR) model and the linear systems analysis Linearly Varying Gain Factor Model (LVGFM), can give good forecasts of the flow for the Nyando River using rainfall as input. This can be useful for planning and rational use of water supply from Nyando River. In addition the models can be used as tools for early warning systems for the perennial flooding problem in Nyando.

6.2   Suggestions For Further Study

The results of this study are far from perfect. This suggests that there is need to carry out further investigations with a view to improving the model efficiencies further. Among the areas that require further investigations for this purpose are listed below:

(i)   The daily averaged rainfall needs to be recalculated with a view to getting more representative values. The assumption that simple arithmetic mean method was

adequate for the Nyando catchment may not have been correct. More substantive methods such as Thiessen Polygon or Isohyetal methods should be tried.

(ii)   The discharge data should be investigated further with a view to improving the quality. Some discharge records are too low and do not look realistic. Such values should be investigated further and possibly be replaced with estimated values. The available records show that gauging station 1GD04, with a daily average discharge of about 5 cumecs, has data records that are smaller in value than upstream stations such as 1GD07, which has a daily average discharge of about 10 cumecs. This is quite unusual as the downstream stations are expected to record higher values than those of upstream stations, unless there is some kind of a sink in between. 1GD04 is very near 1GD03 and the daily average discharge would therefore be expected to be closer to 17 cumecs, the daily average discharge at 1GD03. This is not the case and this anomaly calls for urgent investigation.

(iii)   The results of unconstrained pulse ordinates show some form of instability in the Nyando catchment. The causes of these apparent instabilities should be investigated further with a view to further improving the models' performance efficiency.

ACKNOWLEDGEMENT

REFERENCES

Adegu, D.B (1999): Rainfall-Runoff Modelling of the Nyando Catchment, Kenya, using Satellite (CCD) data. *MSc Thesis, International Institute for Infrastructural, Hydraulic and Environmental Engineering (IHE), Delft, The Netherlands: 1 –27.*

Anderson, M.P and Woessner, W.W (1992): Applied Ground Water Modelling; Simulation of flow and advective transport. *Academic Press, San Diego, 28-73*

Bird, J.O and May, A.J.C (1994): Mathematics For Electrical Technicians, *Longman Group, England, 390-493*

Bulmer, M.G (1979): Principles of Statistics. *Dover publications Inc., New York, 45-65*

Chaudry, M. H (1993): Open Channel Flow. Prentice-Hall Inc., *New Delhi, 24-48*

Delleur, J.W (1991): Time Series Analysis Applied to Hydrology. *V.U.B-Hydrologie, Brussels, 10-17*

Erricker, B.C (1980): Advanced General Statistics. *Hodder & Stoughton, London, 9-13*

Falconer, R.A (1992): Water Quality Modelling. Department of Civil Engineering, University of Bradford. *Ashgate, Vermount USA, 69-81*

Freund, J.E (1967): modern elementary statistics, 3rd edition. *Prentice Hall, Englewood: 59 – 79.*

Gee, B (1981): Essay writing and project reporting. *Plymouth, 1-15*

Hanke, J.E and Reitsch, A.G (1991): Understanding Business Statistics.

Hoel, P.G (1984): Introduction to Mathematical Statistics.

ICRAF (1987): Catchment and rehabilitation programme. Lake Basin Development Report on phase II

JICA (1987): Water resources /transportation / Energy. The study of integrated Regional Development master plan for the Lake Basin Development area. Final report Vol. 5, sector report 3.

JICA (1992): Sectoral Report (B) Hydrology. The study on the National water master Plan. *Sectoral report (B) Hydrology, Nairobi, Kenya: B-6 - B-29*

Kachroo, R.K, Liang G.C O'Connor K.M (1988): Application of the linear perturbation Model (LPM) to flood routing on the Mekong River. *Hydrological sciences Journal 33, 2, 4/1988*

Kachroo, R.K and Liang, G.C (1992): River flow forecasting. Part 2. Algebraic Development of linear modeling techniques. *J. of hydrol. 133: 17–40*

Kachroo, R.K, Sea C.H, Warsi M. S, Jemenez H and Saxena R.P (1992): River flow forecasting. Part 3. Application of linear techniques in modeling rainfall–runoff transformations. Journal of Hydrology 133: 41 – 97.

Kachroo R.K (1992): River flow forecasting. Part 1. A discussion of the principles. *Journal of Hydrology 133: 1 – 15*

Kato, P (1982): Rainfall-Runoff relationship in the Ruvu Basin of Tanzania. *Msc Thesis, University of Nairobi, Kenya: 1 –- 48*

LBDA (1986): Water resources Database. Phase 1 Report, volume1: 45 - 96

Liang, G.C and Nash J.E (1988): Linear models for river flow forecasting on large catchments. *Journal of Hydrology, 103: 157 – 188.*

Liang, G.C (1990): Linear Perturbation Model for real time flow forecasting on the Yangste River at Heinkou. The hydrological basis for water resources Management (proceedings of the Beijing symposium, October 1990) *I A H S Publ. No 197: 47-55*

Liang G.C (1995): Mathematical methods for River flow forecasting. International Advanced Course / workshop on river flow forecasting lecture notes. *Department of Engineering Hydrology, University College Galway, Ireland: 1 – 49*

Liang G.C and Guo Y.F (1994): Observed seasonal hydrological behaviour used in flow forecasting on Yangtze River above Hankou. *Journal of Hydrology 154: 383 – 402.*

Liang G.C, Kachro R.K, Kang W and Yu X.Z (1992): River flow forecasting. Part 4. Application of linear modelling techniques for flow routing on large catchments. Journal of Hydrology 133: 99–140.

LVEMP (2002): Integrated Water Quality/Limnology Study for Lake Victoria. *Final Report. Part II Technical Report, Nairobi, Kenya: 45-96.*

Maidment, D.R (1993): A Handbook of Hydrology. *McGraw-Hill Inc., New York.*

Mainul, A and O'connor K.M (1994): A simple linear rainfall–runoff model with a Variable gain factor. *Journal of Hydrology, 155: 154 – 183.*

Mamdouh, S (1993): Statistical Analysis in Water Resources Engineering: 63-103.

Manning, C.J (1992): Applied Principles of Hydrology. *Prentice Hall, New Jersey, 171-216*

McClave, J.T and Sincich, T (2000): Statistics, eighth edition. *Prentice-Hall Inc,Upper Saddle River, 537-542*

Moore, P.G (1974): Principles of Statistical Techniques 2nd edition. *Cambridge University Press, Cambridge*: 106 -107

Mutua, F.M (1979): Rainfall runoff model on Nzoia Basin. *Msc Thesis, University of Nairobi, Kenya: 1 – 19, 54 – 68.*

Mutulu, P.M (1984): Stochastic Modeling of rainfall-runoff data for Sondu and Yala River basins. *Msc Thesis, University of Nairobi, Kenya: 95 –106.*

O'Connor K.M (1997): Background to The Seventh Advanced Course/Workshop on River flow forecasting. *Department of Engineering Hydrology. University College Galway, Ireland: 1-23*

Ogallo, L.A.J (1980): Time series analysis of rainfall in East Africa. *PhD Thesis, University of Nairobi, Kenya: 18 – 41*

Ogallo, L.A.J (1977) Periodicities and trends in the annual rainfall series over Africa *Msc Thesis, University of Nairobi, Kenya: 18 – 29, 45 – 47.*

Opere, A. O (1991):The use of rainfall–runoff model in forecasting of discharge from Rainfall from the upper Athi–River Catchment. *Msc Thesis, University of Nairobi, Kenya: 6 –8, 18 –28, 37 –42, 86 – 93.*

Opere, A.O (1998): Space-Time characteristics of stream flow in Kenya. *PhD Thesis, University of Nairobi Kenya: 13 –35.*

Ponce, V.W (1989): Engineering Hydrology: Principles and Practices. *Prentice Hall, New Jersey: 389-451*

Frank, H and Althoen, S.C (1994): Statistics-concepts and applications. *Cambridge University Press, Cambridge: 58 - 71*

Shaw, E.M (1988): Hydrology in Practice. *Chapman and Hall, London: 322-345.*

Singh, V.P (1992): Elementary Hydrology. *Prentice Hall of India, New Delhi, 737-751*

Tiessen, H (1997): Measuring The Real World – a textbook of applied statistical methods. *John Wiley and Sons, New York: 225 - 226*

UNDP/ WMO, (1974): Hydro meteorological survey of the catchments of lakes Victoria, Kyoga and Albert. Hydrological studies of selected River Basins. Technical Report 1, Vol IV. *UNDP/WMO, Geneva: 115-245.*

UNEP, (1992): The EL Nino Phenomenon. *UNEP/GEMS Environmental Library No. 8: 4-5.*

Viessman, W and Lewis, G.L (1996): Introduction to Hydrology. *Harper Collins College Publishers, New York, 505-666.*

Wanjohi, D.M (1999): Water flow and quality modelling of the Nyando River in Kenya. *MSc Thesis, International Institute for Infrastructural, Hydraulic and environmental Engineering (IHE), Delft, The Netherlands: 6-10.*

Ward, R.C and Robinson, M (1990): Principles of hydrology. *McGraw-Hall Book Company, London, 216-288.*

WMO/FAO (1963): Proposal for a comprehensive hydro meteorological survey of Lake Victoria catchment: 37 - 78

WMO (1975): Hydrological Forecasting Practices. *Operational hydrology Report No 6. WMO No. 425, 13-19*

WMO (1975): Intercomparison of conceptual models used in operational hydrology. *Operational hydrology report No. 7 WMO No. 429, 1-25*

WMO (1990): Hydrological Models for water resources systems design and operation *Operational hydrology report No 34, WMO No. 740, 1-42*

WMO (1992): Simulated real time intercomparison of hydrological models. *Operational hydrology report No. 38, WMO No. 779, 6-14*