LONG TERM RAINFALL-RUNOFF-LAKE LEVEL MODELLING OF THE LAKE NAIVASHA BASIN, KENYA

Lal Perakum Muthuwatta
March 2004

LONG TERM RAINFALL-RUNOFF-LAKE LEVEL MODELLING OF THE LAKE NAIVASHA BASIN, KENYA

By

Lal Perakum Muthuwatta

Thesis submitted to the International Institute for Geo-information Science and Earth Observation in partial fulfilment of the requirements for the degree of Master of Science in Water Resources Survey and Environmental Systems Analysis and Management

Degree Assessment Board:

Prof. A.M.J.Meijerink (Chairman)
Dr.Ir.P.Droogers(External Examiner)
Drs.R.Becht (supervisor)
Ir.A.M.van Lishout (Member)

INTERNATIONAL INSTITUTE FOR GEO-INFORMATION SCIENCE AND EARTH OBSERVATION
ENSCHEDE, THE NETHERLANDS

MARCH 2004

Disclaimer

I dedicate this thesis to my father,
who is no more with us to see the completion of it
though he anticipated this to turn out .

ACKNOWLEDGMENT

I express my sincere gratitudes to my supervisor Drs.Robert Becht for the valuable guidance and support extended to me for the completion of my work.

I acknowledge the debt I owed to the International Water Management Institute, for providing me this opportunity to continue my higher studies.

I wish also to thank the Government of The Netherlands through the Netherlands Fellowship Program (NFP), which made this study possible by its financial support.

My appreciation is incomplete, if I do not mention the support and help extended by Dr. A. Gieske ,Prof. A.M.J. Meijerink and Dr. C.M.M.Mannaerts by means of constructive criticisms , valuable suggestions and encouragements.

Thanks are also due to Dr. R.O.Strobl, for expressing his views on the study and specially for correcting the text of the final report.

I would also like to acknowledge Ir. Gabrial Parodi, Dr. Ben Maathuis, Ir. Arno Van Lieshout, Ing. Remco Dost and Mr. Boudewijn Van Leeuwen for their support during various parts of the study.

Special appreciation should go to Palitha and Lalith, who persuaded me towards higher studies in hydrology, also Yann and SMB for their continuous encouragement.

Inspiration I got from one of the previous studies done by Mr.S.G. Mmbui is invaluable. “Mmbui, even though we have never set eyes on each other, I perceive you through your work”.

I would like to extend my gratitude to Vajira, Sandamali, Inoka, Rupal, Priyantha, Tennakoon, Kitsiri, Wandi, heleen , Jessica, Manuela and Divithuru for their laudable support throughout my stay in The Netherlands.

Above all, I wish to thank my wife Ishika, my mother and my mother-in law, for their morale support, without which I will never be able to complete this study. Apologies to my son, Asanga, for not being able to accompany you for one and half years. “Dearest *Bampa* I wish one day you will forgive me when you realized the real reason behind parting you for such a long period”.

ABSTRACT

The Soil and Water Assessment Tool (SWAT) model, coupled to a GIS, was applied to the Naivasha basin in Kenya, a closed basin experiencing diverse climatic conditions from semi arid to humid, comprises with the only fresh water lake in the grate rift valley. The intent of this study was to estimates the lake water level fluctuations by integrating the SWAT with a water balance model, estimating the lake water level based on the lake volumes.

During the basic data preparation stage of the study, the land use map and the digital elevation model covering the study area was derived with the help of remotely sensed information (TERRA-ASTER). The soil map of the area was updated with the soil physical properties, estimated during the study as well as the information found in the literature. Stream flow analysis was carried out for different areas in the basin to understand the basin hydrological responses, especially the behavior of the base flow. Weather generator, which uses to generate daily rainfall was modified by introducing repetition and adjustment procedures proposed in this study. This modification refined rainfall simulations in terms of the monthly totals. The effect of this new modification is presented in chapter 4.

A combination of trial and error and automatic methods to use calibrate the model using the observed monthly stream flow from 1965 – 1975 period. The Parameter estimation program (PEST) was used for the automatic calibration. Manual calibration was performed until the Nash_Sutcliffe coefficient become grater than 0.5. After that the SWAT model was integrated with the PEST program and the Automatic calibration was carried out. Pre sensitivity analysis showed that curve number (CN), groundwater "revap" coefficient and saturated hydraulic conductivity of the soil (SOL_K) are the most sensitive parameters to the model. In addition to that available water content of the soil layer (AWC), ground water recession coefficient (α_{gw}) and threshold depth of water in the shallow aquifer for return flow to occur (GWQMN) was assigned to the automatic calibration. Because of the high correlation with revap, AWC was omitted from the optimisation. Other 5 parameters showed low parameter uncertainty of the optimisation and the higher sensitivity to the model. Automatic calibration resulted the further development of the association between observed and simulated stream flows. Validation period was selected as 1935 – 1975 and calibrated model was applied to simulate the stream flow during that period. Results showed the good agreement between observed and simulation stream flows after adjusting the CN number by –7%.

Monthly stream flow simulated from 1935 – 1998 was integrated with the EXCEL spread sheet based water balance model to estimate the lake water fluctuations. Root mean square error calculated between estimated and modeled lake water level during 1935 - 1980 showed 0.72 meters. This indicates the model performed with an acceptable accuracy. Further the modeled lake levels, based on SWAT simulated stream flow, were compared with the lake levels estimated by other two models developed in previous studies and showed better performances.

Table of Contents

List of Tables

List of Figures

List of Graphs

1. Introduction

1.1. Background

Much of the community water supply and agricultural water needs are controlled by surface and shallow ground water especially in developing countries (Wolski, 1999). According to the medium projection of the United Nations, the world population will reach eight billion by 2025 with the contribution of a growth rate of 3% from some developing countries. As fresh water resources are limited, the question arises of whether there will be sufficient water per capita available in the 21st century to fulfil the demand generated by this growth. At the same time over exploitation of useable water resources already has threatened the sustainability of the fresh water availability (Zalewski, 2000). On the other hand, goods produced by these exploitations control the substantial portions of the economies in developing world. For example, 75% of Kenya's horticultural exports come from the flower industries established around Lake Naivasha (Sharmo,2002), which consumes an estimated 60 mcm (Becht and Harper 2002) of water annually. Studies on water demand for food, environment, industries and domestic use predicted that more than 20 developing countries would experience chronic and physical water shortage in 2025 (Bastiaanssen, 2000). Meanwhile, some countries especially in the Middle East and Africa are already confronted with a shortage in water supply (Al-Weshah, 2002). For example, figures on per capita available of water for Arab region (Table 1-1) shows declining trend. These figures and findings urge the need of water resources management in a resourceful manner in order to meet future water demands. In this context, developing management plans would be certainly complimented by quantitative descriptions of spatial and temporal distribution of water resources and the processes influencing them.

Parameter	*1990*	*2000*	*2010*	*2030*
Population (million)	226	304	408	758
Renewable water resources ($10^9 m^3/y$)	264	264	264	264
Water demand ($10^9 m^3/y$)	200	269	362	671
Balance of water budget ($10^9 m^3/y$)	64	-5	-98	-407
Per capita water resources ($10^3 m^3/y$)	1.17	0.87	0.64	0.35

Table 1-1: Water budget of the Arab Region up to 2030 (Source: Al-Weshah,2002)

In water resources studies, river basin has long been acknowledged as the appropriate unit of analysis for water resources management and has also been named by the United Nations Conference on Environment and Development (UNCED) as the unit of analysis for integrated water resources management in Agenda 21, chapter 18 (McKinney et.al, 1990). The river basin concept is of fundamental importance in hydrological studies, because the water passing through the stream cross section at the basin outlet originates as precipitation on the basin. Since the characteristics of the basin such as geology, topography and land cover control the magnitude and timing of flows, studying a river basin as a system and obtaining hydrological information on basin scale gives more in-depth understanding of the whole system as well as the interaction between different processes.

As mentioned by many authors (Wolski, 1999,Bastiaanssen, 2000) hydrological studies are often hampered by the unavailability of relevant information. The chief reasons for this are: Most of the river basins all over the world are un-gauged. Also in gauged river basins, finding all the information necessary for understanding the hydrological process is difficult due to the limited range of measurement techniques in space and time (Beven, 1999). In this situation, hydrological models supply an alternative solution. Using models gives two important advantages over relying solely on collected data. First models can be used to understand the processes that are difficult to measure because of complexity of temporal and/or spatial scale. Secondly, a model can be used to study the effects of changes in land cover, water management or climate (Kite and Droogers, 2001).

1.2. Problem definition and importance of the study

Hydrological processes inside river basins are complex due to the combined complex nature of the natural processes and man made features. Also properties of media forming hydrological systems display a degree of heterogeneity at various scales. (Wolski, 1999, Bronstert and Bardossy, 2003). Therefore, attempts to obtain quantitative description of hydrology in river basins must consider these spatial and temporal heterogeneities.

Lake Naivasha is the only fresh water lake within the East African rift valley. Since there is no surface outlet to the basin, the quality of the lake is maintained by ground water outflow. Multiple use of Lake Naivasha water includes agricultural irrigation, domestic, municipal, wildlife and generating geothermal power. Olkaria Geothermal power plant located south of the lake generates about 18% of the country's electricity demand. These geothermal wells are recharged by water from the lake. At the national level the lake is a major source of income through tourism. The lake is among the biggest bird sanctuaries in Africa. Hell's gate national park to the south of the lake is the home of many species of wild animals that depend on the lake for their water requirement and habitat (Mmbui, 1999). Past studies show that lake evaporation is higher than the rainfall (Farah, 2001), so that lake is solely depends on the discharge from the basin. Therefore, the changes in the upper areas of the basin can greatly influence the lake water quantity as well as quality. The greatest threats to Lake Naivasha result from an increased water demand throughout the region. In the late 1970's, horticultural farms were introduced around lake Naivasha and have changed the nature of agriculture around the lake substantially. Sayeed (2001) stated that, flower farms occupy 1560 hectares, which is 31% of the total irrigated area around the lake. The Lake Naivasha management plan (1999) has mentioned the amount of water abstraction each year to be a threat to sustainable utilization of lake water. The same report has mentioned it's objective as "*manage existing human activities in the lake eco system through voluntarily adopted sustainable wise use principle to ensure the conservation of the fresh water resources.* At the same time growing flower farms will increase the water demand of the basin. Because of this economic value for irrigation and the recent drop of water level, the water balance of Lake Naivasha has been of wider interest. Understanding the past and present water balance would facilitate the development of future management scenarios in order to maintain the lake sustainability.

Previous hydrological studies in the basin were hampered by lack of properly distributed spatial inputs such as rainfall and topography. Average rain gauge density of the basin is 1 per 230/km^2. But the uneven spatial distribution of the rain gages leads to a data scarcity situation in upper parts of the

basin. Further, the quality of available topographic information also is a barrier to the proper hydrological investigation of the basin.

This study attempts to model the water balance in the lake by integrating, spatially estimated hydrological parameters, rainfall generated using modified weather generator and digital elevation model derived by remotely sensed information, with the Soil and Water Assessment Tool (SWAT).

1.3. Selected Previous studies and application of GIS in hydrology

As Becht and Harper (2002) mentioned, the water balance of lake Naivasha has been of wider interest for more than 100 years: initially because of scientific curiosity about the causes of its extreme fluctuations but latterly for its economic value for irrigation and supply of potable water to nearby Nakuru area. Nilsson (1932) proposed the lake freshness to be a result of water both entering and leavening the lake via underground seepage. According to Becht and Harper (2002), a first quantification attempt was carried out by Sikes (1936) to estimate the monthly and annual water budget for the lake. He estimated the lake seepage as 43 mcm/year (Million cubic meters per year). McCann (1974) estimated that, about 34 mcm/year of water recharges the shallow ground water aquifer from the lake. The water balance, compiled for 3 years (1973-1975) concluded discharge and direct rainfall, as the major inputs to the lake, while main out flows were evaporation and seepage (Becht and Harper, 2002). A recent study conducted by Mmbui (1999) using historical inflow data from 1932 to 1998 estimated the lake levels based on simple water balance model and were correlated well with the lake levels before 1984. In the same study the average monthly ground water inflow was estimated as 4.6 mcm. Podder (1998) estimated annual average inflow from a sub basin named Malewa, which located in the Eastern part of the basin, as 214.7 mcm. Lukman (2003) used GIS based Soil Water Assessment Tool (SWAT) to estimate surface runoff from different parts of the Naivasha basin.

Early studies on hydrological modeling were significantly suffered by the insufficient computational power. This has been overcome with the rapid development of computer systems and software. Geographic Information Systems (GIS) offer new opportunities for hydrological modeling. Primarily in generating inputs. GIS has the capability of handling large amounts of spatially detailed information derived from various sources such as remote sensing and ground surveys. With the advent of increasing computing power and GIS techniques, physical-based hydrologic modeling has become important in contemporary hydrology for assessing the impact of human intervention and/or possible climatic change on basin hydrology and water resources (Alemaw et.al, 2003). Arnold et.al (1998) has mentioned about the feasibility of simulating large areas simulation by the integration of hydrological models with GIS. In the follow up paper by Sirinivasan et.al (1998), an example of large area hydrologic modeling with the help of GIS based on the data from Trinity river basin, Texas was given. There were many studies found in the literature about using GIS for hydrological modelling. Those studies mainly used GIS for estimating parameters for hydrological models, subdivision of the basins into *hydro logic response units* and visualization (eg. Band,1995)

Many authors (Meijerink et.al, 1994, McDonnel, 1996) stated incapability of representing continuous temporal element of modelling as one of the major limitations of the integration of GIS with hydrological models. Therefore, the most popular way to use GIS in hydrological modelling is a loosely coupled system where the model and GIS maintain two separate databases and interact through some

form of file exchange or conversion processes. With this approach it is possible to obtain better flexibility from the model as well as more GIS functionality for analysis and visualization of model results.

Interpolation techniques are one other major functionality that came with the development of computers. Some information required by hydrological models comes in a form of point data. Often these point data are unevenly distributed throughout the basins. In order to integrate these with the other data sets such as raster data supplied by satellite sensors as well as make estimations for data unavailable locations, interpolation techniques play a crucial role. These estimations generally are based on the surrounding measurements or internal spatial correlation structures of the data sets.

At the same time development of remote sensing (RS) sensors and techniques during the last decades have complimented modelling application by supplying spatially and temporally distributed biophysical and hydrological input parameters in different scales. Satellite sensors have the capability of observing large areas on a regular basis, which overcome the problem of intensive labour and time consumption in large area surveys. In some parts of the world digital contours developed based on existing topographic maps are the only source of elevation data(Hutchinson, 1999). In GIS based hydrological modelling, digital elevation models created using that contour information often create problems in the areas where terrain steepness is low. At present there are quite a lot of new data sources for generating digital elevation data such as Laser Altimeter and Synthetic Aperture Rader (SAR). There were many applications found in the literature about integrating remote sensing information with hydrological modelling. (e.g. Alemaw et.al, 2003, Kite and Droogers, 2000) These applications vary in complexity due to the physics they describe. However those applications show the value of remotely sensed data for hydrological modelling.

1.4. Objectives

The general objective of this study is to apply a basin scale model to estimate spatial distribution of the flow in the basin to lake Naivasha and estimate the lake water level fluctuation. To achieve above objective following specific objectives were addressed

- Modify the rainfall weather generator model (WXGEN) to desegregate monthly rainfall data in to daily time series accurately by introducing repetition and adjustment procedures.
- Estimate the inflow from different subbasins using Soil Water Assessment Tool (SWAT) based on adjusted rainfall data.
- Calibrate simulated inflow from sub catchments with available stream flow data.
- Use the water balance model to estimate lake water levels by incorporating modelled inflow.

1.5. Outline of the Thesis

Chapter 2 presents the brief description of Naivasha river basin and field data collection. This includes the data availability and the preparation of basic data prior to the modelling. Chapter 3 contains the results of stream flow analysis in two different areas of Naivash basin, which gives the physical explanations for initial setting up of the model. Chapter 4 discuses the new rainfall desegregation procedure introduced by this study based on repetition ads adjustment that generates the daily time series of rainfall, which are closely compatible with the observed monthly time series of rainfall. Chapter 5

briefly discussed about the hydrological component of soil and water assessment tool (SWAT). This includes the physical equations and assumptions used by SWAT during the hydrological phase. Chapter 6 deals with the application of the model initially to Gilgil sub basin and then to the Naivasha basin. This chapter also discussed about the manual calibration and the automatic calibration procedure used in this study. Chapter 7 finally discusses the research summary and the conclusions.

2. Study area and data preperation

2.1. Study area

Lake Naivasha, located within the Eastern branch of Great Rift Valley, occupies a basin area of about 3200 km^2. It lies approximately between latitude 0^0 10'S to 1^0 00'S and longitude 36^0 10'E to 36^0 45'E. Basin altitude varies from about 1900 m at the bottom of the valley to 3200 m in the Nyadarua Mountains found on the eastern boundary of the basin. Due to the altitudinal differences, there are diverse climatic conditions found in the basin. The climate of the area is a typical equatorial tropical climate with two rainy seasons followed by a dry season. The Relief controls the precipitation pattern with much more rain in higher altitudes (M'mbui, 1999). The Rift valley floor experienced an average annual rainfall about 640mm while the wettest slopes of the mountains receive about 1525mm.

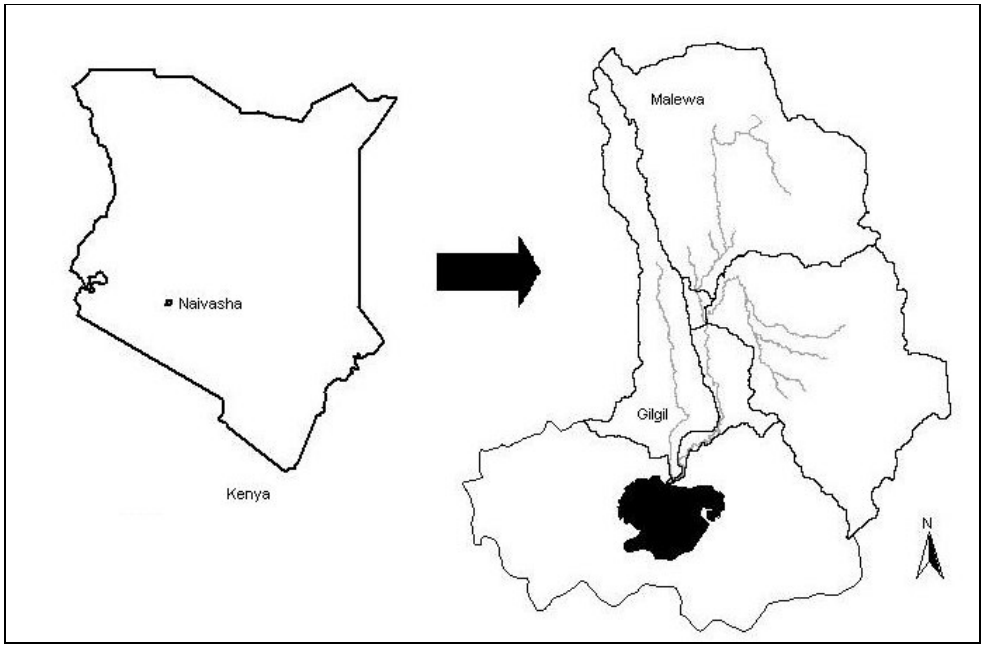
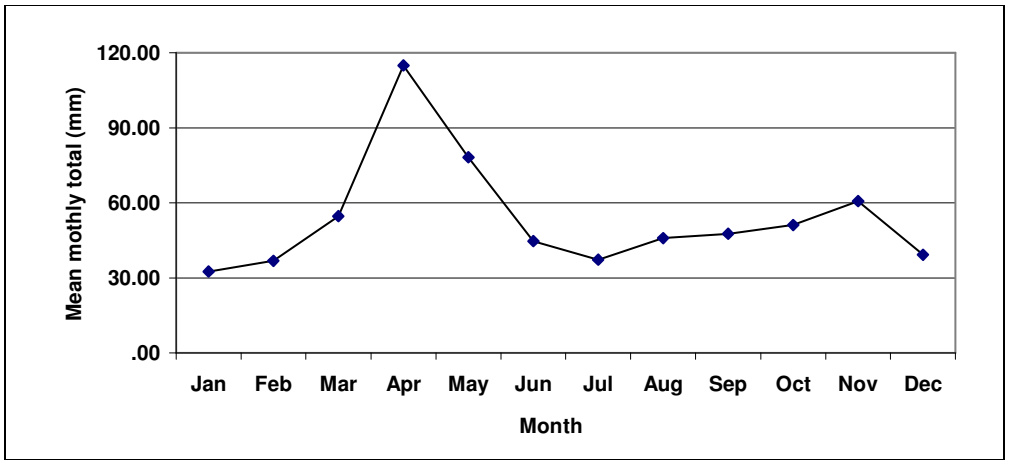


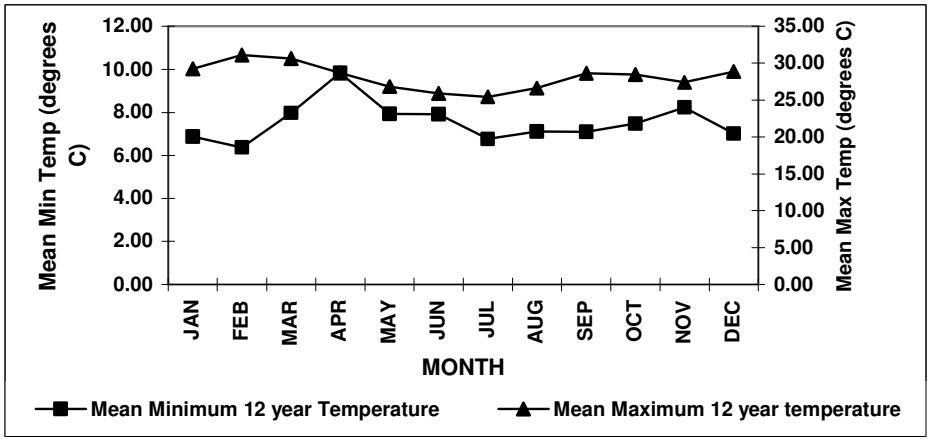
Figure 2-1:Map of the study area

There are two rainfall peaks, which occur in April to May and September to October. The driest months are January, February and December (Graph 2-1).



Graph 2-1: Monthly Average rainfall as measured at the Naivasha District Office, Kenya. (1932 to 1997)

The lowers temperatures are experienced in July, while the highest temperatures occur in March. The potential evaporation is about twice the annual rainfall in the semi arid area while in the upper basin humid areas, rainfall exceeds potential evaporation in most parts of the year (Farah, 2001). The annual temperature range is approximately from 8^0C to 30^0C.



Graph 2-2: Average Minimum and Maximum temperature for a 12-year period (1990-2001) (Source: Lukman, 2003)

Malewa and Gilgil are the two main perennial rives flowing into the lake. In addition to that, Karati, an ephemeral river drains into a lake. Malewa is the major river that feeds the lake, contributing about 90% of the discharge with a basin area about 1600 $km^{2.}$ The Gilgil and Karati basins are about 527 and 450 km^2 respectively; contribute the remaining 10% of the discharge into the lake (Lukman, 2003).

According to the soil map (Sombroek et.al, 1980), the soils in the upper basin area consist of clay loam to clay. These soils are deep (80-120 cm) and well drained. In the lower plains, soils are mainly sandy clay loam to sandy clay, and are deep and well drained. On the mountains, the soils are shallow (<50 cm) to moderately deep and consist of a complex of loam, clay loam and clay.

The land cover of the Naivasha basin can be broadly categorized into four groups, namely Agriculture, Grass, Bush land and Forest. In the humid region, predominant land cover classes are forest and crops. The main crops consist of maize, potatoes and wheat. In addition to that, there are many other vegetables growing by smallholder farmers in the middle part of the basin. In the semi arid region, there are extensive areas of grassland and bush land, which are used for livestock grazing. As mentioned earlier, intensive horticulture farming under irrigation is common around the lake. The North-Eastern part of the lake is predominantly occupied by large-scale vegetable and dairy farms. However, the detailed land use in different parts of the basin is subjected to high heterogeneity. The natural vegetation surrounding the lake is mainly papyrus swamp vegetation while; cactus, acacia, bamboo, shrub and coniferous trees are mainly further away from the lake. The smallholder farmers occupy the upper basin areas. Also, there are big areas covering grass and bushes. The natural vegetation occupies upper Abedears high lands.

2.2. Preparation of basic data

All the spatial inputs required for modelling such as number of sub basins, length of the longest stream in sub basins and hydrological response units were generated through AvSwat2000, which is the Arc View extension of the Soil and Water Assessment Tool ((SWAT). In order to generate these inputs, digital data on land use, soil and elevation were integrated using AvSwat2000. Digital elevation model (DEM) and land use map of the study area were prepared by using remotely sensed information. Soil physical properties were estimated using the data collected during the fieldwork and information found in the literature. Moisture bulk density was estimated based on the soil samples collected during the fieldwork. Other properties such as available water content and silt sand any clay contents of different soils, were estimated using the available techniques and information found in the literature. Following section discussed the preparation of land use and soil maps as well as the digital elevation model, using existing and remotely sensed information.

2.2.1. Land use

Land use classification was attempted by using remotely sensed data. Two images acquired from the TERRA-ASTER sensor on 08.03.2003 over the study area were used. During the field study, GPS coordinates were collected for the purpose of geo referencing and land use classification. Coordinates were collected for different land use categories at 221 locations (Appendix 2.1 gives coordinates of 130 locations). First the geo-referencing of the bands were carried out using GPS data, well distributed over the images. The RMS error of the geo referencing was 1.27 and 1.98 pixels for lower and upper image respectively, which was within reasonable accuracy. In addition to that, by overlaying a road map, positional accuracy of the geo-referenced image was visually checked. However, it was impossible to separate some of the major land use classes in the feature spaces of different band combinations. For example, the Bamboo forest in the upper areas and the Acacia trees in the lower areas and alone the rivers indicate same spectral signatures. Also, the spectral signatures of green grass and the wheat areas also mixed with each other. Therefore, the land use map of the study area was prepared by visual analysis. Different areas were demarcated on the false colour composite (FCC) of bands 3, 2 and 1 (Figure 2-2). Those areas were compared with the GPS locations collected on the ground. Then by using the notes made in the field, different land use classes were assigned to those demarcated areas. The Final land use map consists of 10 different classes as shown in Figure 2-3.

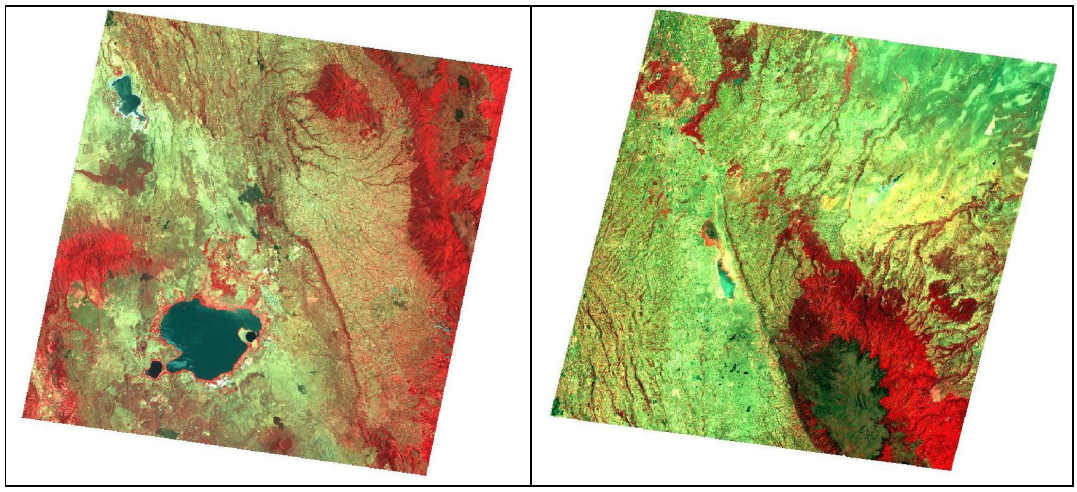


Figure 2-2: TERRA-ASTER images (FCC 321) covering lower (left) and upper (right) areas of the Naivasha basin

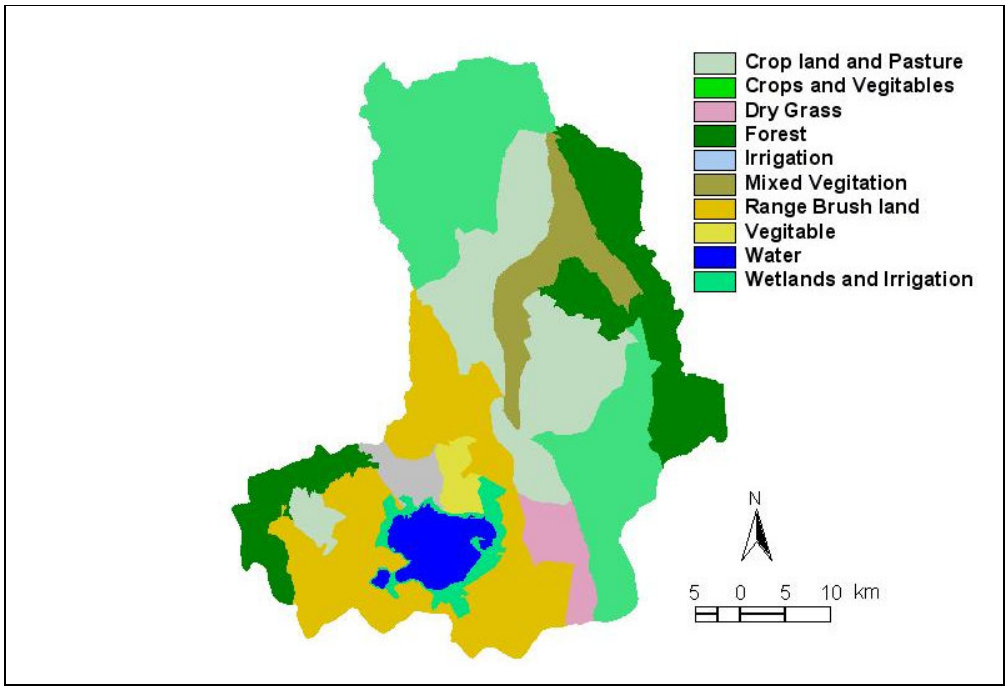


Figure 2-3: Final Land use map of the Naivasha basin

2.2.2. Soils

A digitised soil map of Naivasha area was available from previous studies. This map is based on the **exploratory soil map of Kenya** developed in 1980 on the scale of 1:1,000,000 (Sombroek et.al., 1980). For the classification, Kenyan Soil Survey has used the FAO-UNESCO classification system. The information book attached to the soil map gives indication of the drainage conditions in nominal scale (e.g. well drained, poorly drained) and the effective soil depth in five intervals in centimetres, namely shallow (0-50), moderately deep (50-80), deep (80-120), very deep (120-180) and extremely

deep (more than 180). The Naivasha basin and the surrounding area consist of 42 soil groups. Figure 2-3 shows the major landforms in different colours and different soil groups indicated in letters as used in the Kenyan soil map. Annex 2.1 presents the descriptions of different soil classes.

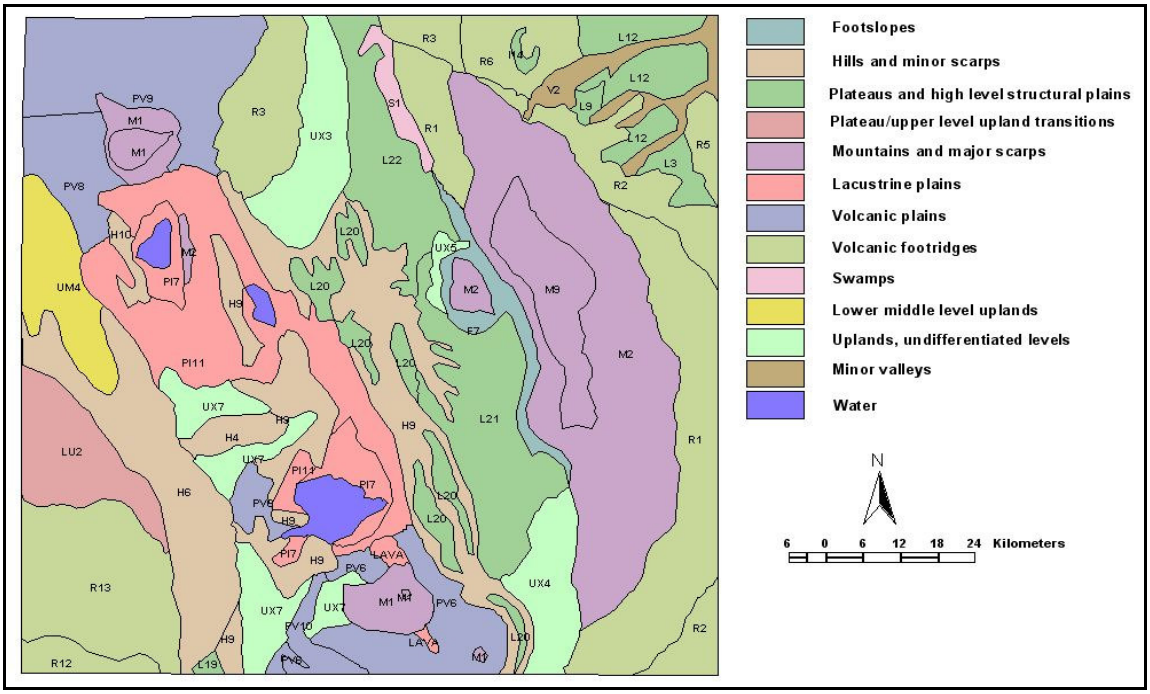


Figure 2-4: Soil map of the Naivasha basin (See appendix 2.1 for the detailed information)

The Soil water assessment tool (SWAT) requires different soil physical properties for modelling (Table 2-1). Since there were no available data on different soil layers, only one soil layer was assumed over the basin during the modelling.

Name	Description
NLAYERS	Number of layers in the soil (min 1 max 10)
HYDGRP	Soil hydrologic group (A, B, C, D)
SOL_ZMX	Maximum rooting depth of soil profile
ANION_EXCL	Fraction of porosity from which anions are excluded
SOL_CRK	Crack volume potential of soil [optional]
TEXTURE	Texture of soil layer [optional]
SOL_Z	Depth from soil surface to bottom of layer
SOL_BD	Moist bulk density
SOL_AWC	Available water capacity of the soil layer
SOL_K	Saturated hydraulic conductivity
SOL_CBN	Organic carbon content
CLAY	Clay content
SILT	Silt content
SAND	Sand content
ROCK	Rock fragment content
SOL_ALB	Moist soil albedo

Table 2-1: Soil physical properties required by the Soil and Water Assessment Tool

A soil hydrologic group defines the groups of soils having similar runoff potentials under similar storm and cover conditions. Based on the information extracted from the soil map (Sombroek et.al, 1980), soil hydrologic groups were assigned to each soil in the area. There is no information on rooting depths. During the fieldwork, rooting depths were identified in some soils especially near the road cuts. Based on these measurements average values were estimated for some areas. For other areas, information was gathered from the published information and personal communication with local authorities. Soil depths were estimated for each soil class based on the soil depth ranges given with the map. Moisture bulk density (ρ_b) was estimated for the different soil samples collected during the field study (Table 2-2) and some values were taken from published data (Zinck, 1986). Soil samples collected in the field were oven dried for 24 hours at 105^0C. To estimate bulk density, weights of dried soil samples were divided by their original weights. Calculated values for 3 soil classes are presented in Table 2.2.

Location ID	X (TM)	Y(TM)	ρ_b	Soil
1	227127	9931062	1.100637	L21
2	223634	9931294	1.202548	H9
3	229557	9940762	1.385987	L21
4	212050	9948958	1.284076	H9
5	212036	9948960	1.335032	H9
6	203574	9942250	1.192357	PI11
7	203591	9942244	1.243312	PI11

Table 2-2: Estimated bulk density values for 3 selected soil groups

2.2.3. Estimation of available water capacity

Available water capacities of the soil layers were estimated for different soils as difference between field capacity (θ_{fc}) and permanent wilting point (θ_{pwp}). Field capacity is the index of the water content that can be held against the force of gravity (Dingman, 1994). This can be computed as the water content corresponding to a pressure head of –340 cm and computed from equation 2.1.

$$\theta_{fc} \equiv \phi(\frac{\varphi_{ae}}{340})^{1/b} \qquad (2.1)$$

where Φ is the porosity, φ_{ae} is the air entry tension in cm and b is the exponent describing the moisture characteristic curve.

Plants cannot exert suctions stronger than about –15000 cm. Therefore, the permanent wilting point was calculated as the water content in the soil corresponding to a pressure head of –15000 cm and computed from equation 2.1.

$$\theta pwp \equiv \phi(\frac{\varphi_{ae}}{15000})^{1/b} \qquad (2.2)$$

The difference between the field capacity and permanant wilting point is considered to be the water available for plant use as in equation 2.3

$$\theta_a \equiv \theta_{fc} - \theta_{pwp} \qquad (2.3)$$

Average values for porosity, air entry tension and b were used as given in table 2-3 (Dingman, 1994). Saturated hydraulic conductivity (K_h) values also assigned to each soil class are based on the same table.

Soil Texture	Φ	K_h (cm s^{-1})	φ_{ae} (cm)	b
Sand	0.395	$1.76* 10^{-2}$	12.1	4.05
Loamy sand	0.410	$1.56* 10^{-2}$	9.0	4.38
Sandy loam	0.435	$3.47* 10^{-3}$	21.8	4.9
Silt loam	0.485	$7.20* 10^{-4}$	78.6	5.3
Loam	0.451	$6.95* 10^{-4}$	47.8	5.39
Sandy clay loam	0.420	$6.3* 10^{-4}$	29.9	7.12
Silty clay loam	0.477	$1.70* 10^{-4}$	35.6	7.75
Clay loam	0.476	$2.45* 10^{-4}$	63.0	8.52
Sandy clay	0.426	$2.17* 10^{-4}$	15.3	10.4
Silty clay	0.492	$1.03* 10^{-4}$	49.0	10.4
Clay	0.482	$1.28* 10^{-4}$	40.5	11.4

Table 2-3: Parameters for estimating the available water capacity in different soils

2.3. Digital elevation model

A Digital elevation model (DEM) describes the elevation of any point in a given area at A specific spatial resolution as a digital file. The most commonly used method of acquiring elevation data is digitising contour maps and transforming them into a raster format using interpolation techniques. The quality of this DEM is often limited due to many reasons. Meijerink et.al (1994) describes the major limitations of DEM based on contour digitisation as: unavailability of relief information between contours and inaccuracies related to the map due to photogrammetric and cartographic errors. However, the development of new satellite sensors provides reasonable solution for generating digital elevation models. Eefforts have been directed toward developing satellites and sensor systems capable of producing global elevation data in digital format. Among them, the Advanced Space borne Thermal Emission and Reflection Radiometer (ASTER) on-board the National Aeronautics and Space Administration's (NASA's) Terra spacecraft provides along-track digital stereo image data at a 15-m resolution. The ASTER sensor is designed to provide image data in 14 visible, near-infrared, short wavelength infrared and thermal infrared spectral bands. Stereo image data are recorded only in Band 3, which is the near-infrared wavelength region from 0.78 to 0.86 μm, using both nadir and aft-looking telescopes. From the nominal Terra altitude of 705 km, the ''pushbroom'' linear array sensor covers a 60-km-wide ground track at a 15-m spatial resolution. There is an approximately 60-s interval between the time the nadir telescope passes over a ground location and the after telescope records the same location on the ground track of the satellite. Images generated from the nadir and aft telescopes yield a B/H ratio of 0.6, which is close to ideal for generating DEMs by automated techniques (Hirano, 2003). This part of the study was conducted using the OrthoEngine module of the Geomatica

Focus software developed by PCI Geomatics**.** The process of generating a digital elevation model (DEM) from an ASTER image consists of several steps:

2.3.1. Data input

OrthoEngine reads the raw satellite data, saves the imagery into a PCIDSK file, and adds a binary segment containing the ephemeris data (orbit information) to the file. OrthoEngine can read the HDF format, which ASTER level 1A data comes with. When reading the HDF file it is possible to extract only the required bands. For example, in case of generating a DEM from ASTER data, only 3N and 3B bands were read.

2.3.2. CGP/TC collection

After reading the images in OrthoEngine, the next step was to enter the ground control points. A ground control point (GCP) is a feature that can be clearly identified in the raw image for which we have a known ground coordinate. The GCPs are used to determine the relationship between the raw image and the ground by associating the pixels (P) and lines (L) in the image to the x, y and z coordinates on the ground.

2.3.3. Model calculation

This step generates a mathematical model that builds a correlation between the pixels and their ground locations. Residual Errors can be used to determine the goodness of fit of the model. Residual errors are the difference between the coordinates that were entered for the ground control points (GCPs) and where those points are according to the computed mathematical model.

2.3.4. Extract the DEM from the overlap between the epipolar pairs

Epipolar images are stereo pairs that are reprojected so that the left and right images have a common orientation. Matching features between the images appear along a common x axis. OrthoEngine uses the math model solution (known exterior orientation) and the pixel and line positions of the points in common in the overlapping images to generate a DEM. The elevations are calculated from the parallax between the corresponding GCPs, tie points and elevation match points in the images. The resulting DEMs are called epipolar DEMs. They are not georeferenced at this stage.

2.3.5. Geocode the epipolar DEMs and stitch them together to form one DEM

The result of this step is, one DEM reprojected to the ground coordinate system. DEMs usually contain poorly correlated areas. It is possible to correct these areas before or after the DEMs are geocoded.

Two ASTER images from 08.03.2003, covering the upper and lower areas of the basin separately, were used to derive the DEM over the whole basin. There were 11 and 7 ground control points used for the lower and upper areas, respectively. First, bands 3B and 3N were extracted from the row level 1A data files. Then the above steps were followed. Since there were not enough ground control points in the upper area, before starting the processing, two images were stitched together. The resulting digital elevation model is presented in figure 2-5.

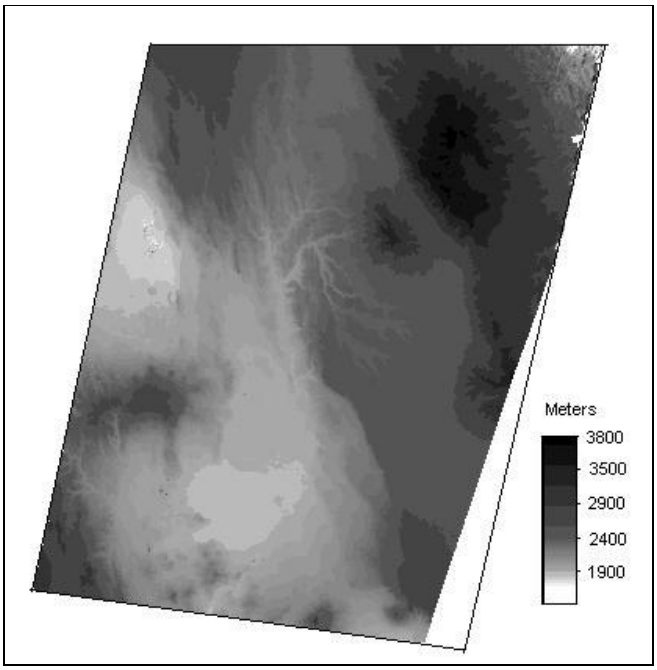


Figure 2-5: Digital elevation model for the Navisha basin and surrounding area

These generated elevation values were compared with the separately collected elevation information using GPS at 20 locations scattered around the basin (Table 2-4).

Point ID	X Coordinate	Y Coordinate	DEM elevation (m)	GPS elevation (m)	Difference (m)
1	223688	9938334	2424	2427	-3
2	204902	9938908	1958	1958	0
3	210707	9932129	1926	1922	4
4	197459	9919794	1932	1934	-2
5	229369	9937438	2436	2427	9
6	207960	9933738	1950	1944	6
7	218412	9918994	2077	2068	9
8	209072	9931434	1913	1917	-4
9	214080	9921226	1916	1907	9
10	205204	9921283	1902	1909	-7
11	211198	9949017	2198	2195	3
12	201864	9953259	2257	2253	4
13	205365	9950182	2188	2191	-3
14	205334	9949084	2163	2159	4
15	223956	9930842	2438	2444	-6
16	224390	9931050	2465	2452	13
17	239554	9942302	3157	3172	-15
18	221236	9956838	2467	2480	-13
19	221757	9930474	2443	2450	-7
20	218699	9918438	2082	2090	-8

Table 2-4: Comparison of DEM elevation with the GPS elevation

This comparison shows that the deviations between elevation values collected using a GPS and DEM is ranging from 0 to 15 m. Those high deviation points (point idS 16,17,18) were compared with the notes made during the fieldwork and found TO BE nearer to the high vegetation areas. Therefore, these deviations could be due to the disturbances from the vegetation during the GPS data collection or due to the height of the vegetation. However in THE lower areas of the basin, the differences were lower than 10 meters. The error associated with the GPS (GARMIN etrex) is also 10m. Therefore, the accuracy of the DEM was within the acceptable range to use with the model. However, for comparison, there was only one point used in the high elevation area (Point 17) and this gave a difference of 15 meters. Therefore, to make a final conclusion about the accuracy, more points should be considered within those regions

All the maps and data prepared during this stage of the study was integrated with the Soil and Water Assessment Tool before the modelling phase of the study. During the calibration, some of the parameters were changed in order to get a better association between the observed and simulated stream flows.

3. Stream flow analysis

3.1. Introduction

Stream flow is the main form of surface water flow and all the other surface flow processes contribute to it. Therefore, determining flow rates in stream channels is a central task of surface water hydrology (Chow et.al, 1988). Combination of base flow, interflow and saturated overland flows generate the stream flow (Maidment, 1992). A stream flow rate integrates all the hydrologic processes and storages upstream of a particular point on a drainage system at a particular time. These flow rates can be used as an integrated index for basin responses to the rainfall. Therefore, it is important to analyse stream flow hydrographs, which explains the functional relationship between flow rates and time at a given location on the stream.

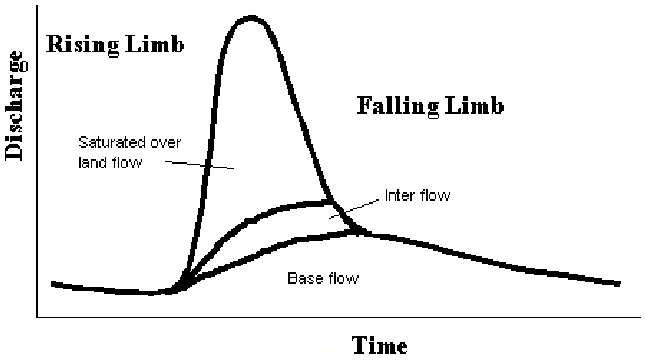


Figure 3-1:Components of a Hydrograph

As shown in figure 3-1, a hydrograph can be separated in to 3 components namely, saturated over land flow, inter flow and base flow. In modelling, these three components can be conceptualised as different reservoirs and integrated into one system by assigning physical properties and transfer rates between them. A physical interpretation derived by analysing stream flow hydrograph for a particular basin can be greatly helpful when setting up a conceptual model realistically. And it supplies a physical basis to hydrological parameter selection in modelling.

Contribution from ground water to the stream flow can be quantified by separating a base flow from the given total hydro graph. This can be useful during the calibration phase of hydrological modelling studies. Further, the knowledge gained by stream flow analysis can be used to approximate empirical parameters such as curve numbers that are crucial to set up a hydrological model closer to the real ground conditions. Variation of the curve number with respect to the relationship between direct runoff and rainfall was developed by U.S. soil conservation service in 1972(Chow, 1998). Therefore by comparing separated direct runoff with the rainfall, curve numbers for different areas of the basin can be roughly estimated.

3.2. Stream flow analysis in Gilgil

Gilgil, a sub basin that approximately occupies 425 km^2, is located in the western part of Naivasha basin. As mentioned earlier 10 – 15 % of the total inflow of the lake originates from the Gilgil sub-basin. To generate datasets of the discharge, daily read gauge heights were transformed to discharges with the help of a rating curve. Those discharge time series were analysed to understand the basin responses to rainfall and later used to calibrate and validate a hydrological simulation model. Available daily stream flow data was analysed in another parallel study (Dapp and Vreugdenhil, 2004). This

analysis showed that the stream flow data from 2GA3 station (Figure 3-1), which is located in the middle part of the subbasin, was reliable with respect to other stations. Therefore, stream flow data were analysed at the gauge number 2GA3.

There were daily data available from 1958 at this station with missing data. Year 1971 had the least data gaps with respect to other years. Hence, a stream flow hydrograph was analysed for 1971 to obtain preliminary understanding about the sub basin. Generally it requires hydrographs for more years to understand the sub basin properties properly. However, due to unavailability of daily data for other years, only 1971 data was analysed. This prior knowledge of responses gave important insight into the model setup.

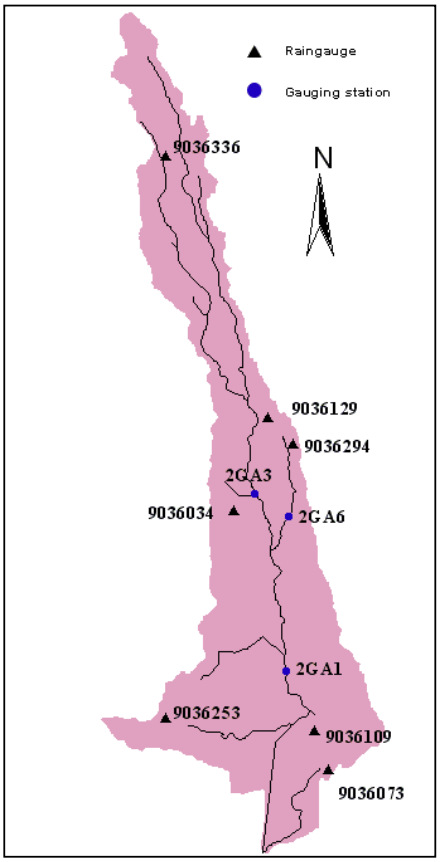
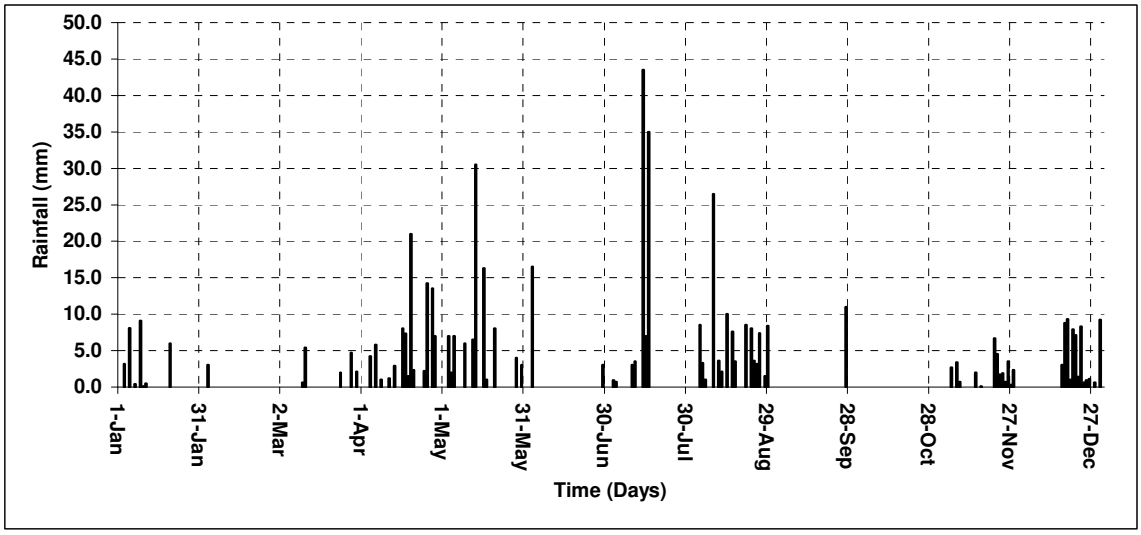


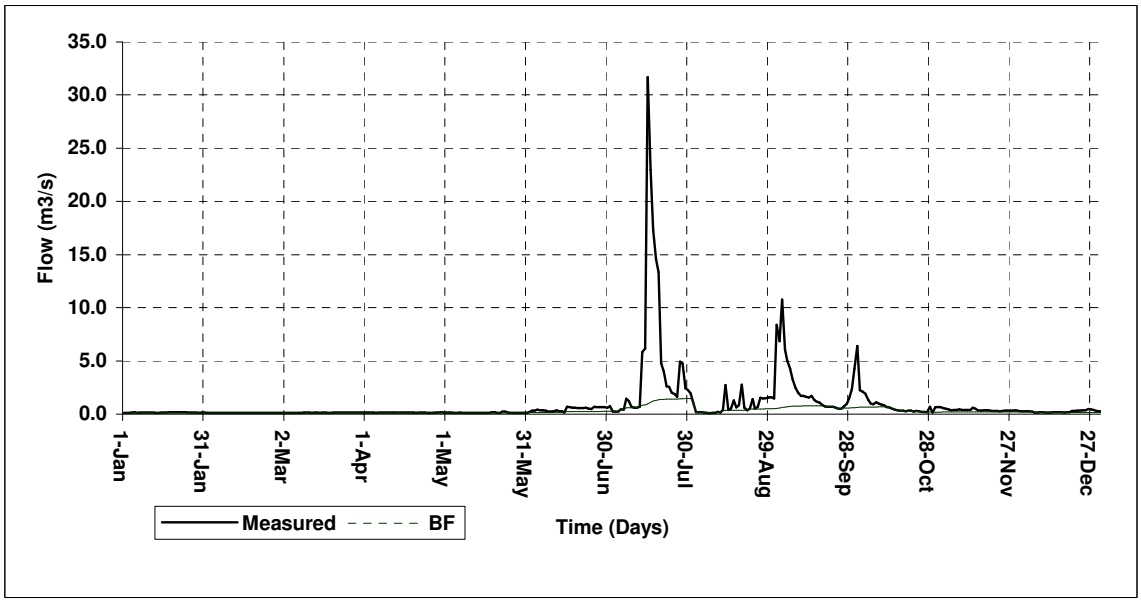
Figure 3-2: Rainfall and river gauging stations in Gilgil subbasin

Measured daily stream flow at GA3 station was compared with the daily rainfall data recorded at the station 9036129. During the fieldwork highly permeable surface soil layers were recognized in the upper areas of the sub basin (Figure 3-3). Impermeable and semi permeable rocks were also identified at shallow depths in the upper areas. In this situation, rainfall is percolating vertically until it encounters the impermeable layer and ponds above that layer to form a saturated zone. This saturated zone is the source of water for lateral subsurface flow. In this area there are no apparent signs of over land flow, but intensive rainfall could generate overland flow in the form of saturated flow.

Figure 3-3: Upper area of the Gilgil subbasin



Graph 3-1: Daily rainfalls at station 9036129 for 1971



Graph 3-2: Stream flow hydrograph at 2GA3 for 1971

An Excel spreadsheet model (TIMEPLOT) based on recursive filter technique (Nathan and McMahon, 1990) was used to analyse the base flow. Graphs 3-1 and 3-2 indicate the daily stream flow at 2GA3 and the daily rainfall at 9036129 stations respectively for 1971. Dotted line in graph 3-2 represents the base flow. From January to June there was some rainfall but no apparent response on stream flow. But the stream flow shows a rapid response to the rainfall event in July. After that in September and October peaks also show the rapid responses to the rainfall.

Annual runoff is 29 mcm for 1971. TIME PLOT Program estimates the base flow and runoff as 9.25 and 19.75 mcm respectively for the same period. Previous studies estimated total Naivash basin annual discharge as 217 mcm (Becht and Harper, 2002) and Gilgil flow as 10-15% of the total flow. In this year it is about 13% of the total flow. Therefore 1971 can be considered as a representative

year. Further, this analysis indicates the base flow contribution to the total flow in Gilgil as 32%. During this period total rainfall recorded at the station was 539 mm, which generates 228 mcm of water to the whole sub basin. Therefore, direct runoff is about 8.5% of the total rainfall. Table 3-1 provides a summary of runoff, base flow and the rainfall in the year 1971.

	mm	MCM
Total rainfall	538	
Total runoff	68	29
Base flow	22	9.25
Direct runoff	46	19.75
Runoff coefficient	8.5%	
Base flow (% of rainfall)	4%	
Base flow (% of total flow)	32%	

MCM – Million cubic meters

Table 3-1: Summary of the stream flow analysis in Gilgil

Due to the complexity of basin processes, use of single recession constant to characterize recession behaviour in a basin, regardless of the season or magnitude of the flow could result certain inaccuracy of the separated base flow (Nathan and McMahon, 1990). This is apparent in the base flow separation graphs. For example, in Graph 3-4, just before 28th September, the program estimated base flow as the lowest flow found in stream flow hydrograph. Graph 3-3 shows some rains in the same period. Therefore the estimated base flow would be lower than the value presenting in the graph. However, the use of TIMEPLOT, is faster than a manual methods and eliminates the subjectivity of the manual methods. Therefore, the estimated values could be used as references without taking absolute values.

There were no rainfall recorded during 30.08.1971 to 26.09.1971 and 28.09.1971 to 04.11.1971. Therefore, the small peaks after the overland flow recession in September and October show the delayed flow of ground water. The base flow recession constant, which is a direct index of ground water flow response to changes in recharge, was estimated using following equation (3.1).

$$Q_t = Q_0 . e^{-\alpha t} \qquad (3.1)$$

Where, α is the base flow recession constant, t is the time lapsed since the start of the recession (days), Q_t is the ground water flow in day t (mm), and Q_0 is the ground water flow at the start of the recession.

In September the ground water recession begins at 15.09.1971 and it declined to the base flow on 21.09.1971. The recession constant estimated using above equation gives the value of 0.12 day. An estimated recession constant for October gives the value of 0.11day.This parameter would affect the shape of the modelled hydro graph during the periods where shallow aquifer receives no recharge.
According to graph 3-1, the interflow component is minor in the Gilgil sub basin. Therefore the base flow was separated leaving the remainder as fast runoff.

3.3. Stream flow analysis in Malewa and Thurasha sub basins

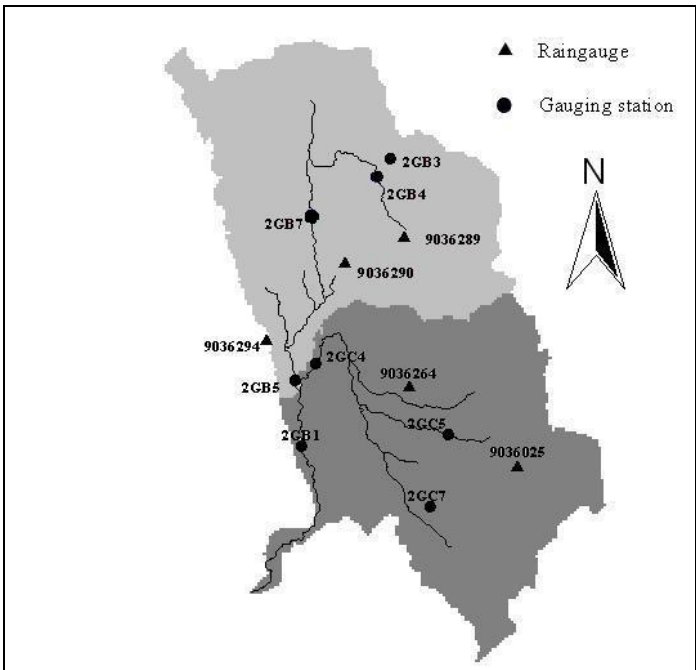


Figure 3-4: Rainfall and river gauging stations in Malewa and Thurasha sub basin

Malewa is the main river that contributes to the lake, which generates about 90% of the total discharge. Total area of the Malewa basin is about 1700 km^2. This river is supplemented by water from the Thurasha river. Al-Sabbagh (2002) analysed stream flow data of the Thurasa and Malewas subbasins and concluded about the data reliability of the gauging stations 2GB1, 2GB5, 2GB4, 2GC4, 2GC5 and 2GC7. During this study, base flow separation was carried out at gauging stations 2GB1 based on the daily data in 1975. Discharge from both the Malewa and Thurasha sub basins contributes to the flow at this station. Daily rainfall recorded at station number 9936290 was used for this analysis. The following table (3-2) and graph (3-2,3-4) presents the results of that analysis.

	mm	MCM
Total rainfall	1091	
Total runoff	65	209
Base flow	27	87
Direct runoff	38	122
Runoff coefficient (%)	3.5	
Base flow (% of rainfall)	2.5	
Base flow (% of total flow)	41.5	

Table 3-2: Summary of the stream flow analysis in Mlewa subbasin

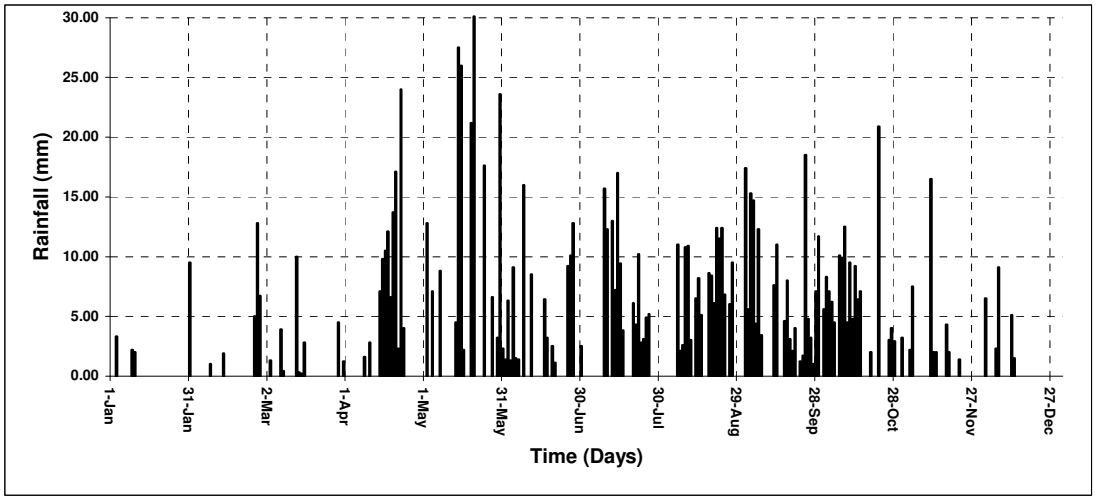


Table 3-3: Daily rainfalls at station 9636241 for 1975

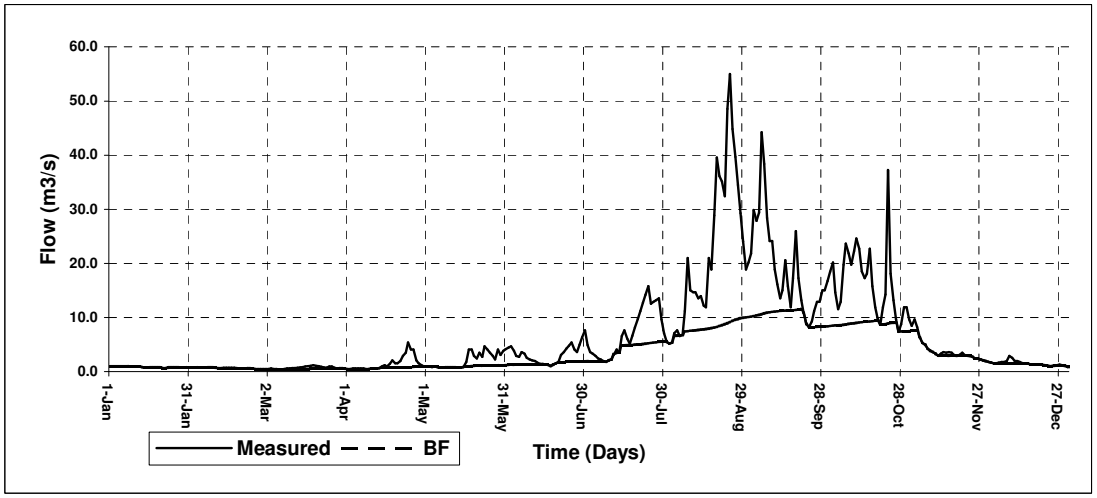


Table 3-4: Stream flow hydrograph at 2GB1 for 1975

Rainfall at the beginning of the year does not create a significant response. Gradual response can be seen from April onwards. High rainfall in June shows low responses. But after the basin becomes wet, rainfall in August and September shows rapid responses. Average annual stream flow was 207 mcm. Analysis of the data showed that the base flow and direct runoff are 92 and 102 mcm respectively. This shows that the base flow contributes about 44% of the total flow. As in figure 3-1, the upper basin areas near Abedears mountain range consist of permeable soil layers. Therefore, a significant amount of rainfall percolates into the soil and unconfined aquifer, which subsequently generates the base flow.

Results of the above analysis and the real ground situation identified during the field work were used to modify the parameters such as curve numbers, saturated hydrologic conductivity and ground water recession during the initial modelling setup. Stream discharge was measured at two points and are presented in Appendix 3.1 and 3.2

4. Weather generator and disaggregation of rainfall

4.1. Introduction

Long term weather data obtained at specific sites on daily and even sub daily basis is essential for water management and simulation modelling in the basins. Most commonly used models require long-term daily values of rainfall, solar radiation as well as maximum and minimum temperature (Soltani et.al, 2000). For example Soil Water Assessment Tool (SWAT) requires daily data on the above-mentioned parameters for the simulations. However, many areas do not have sufficient available weather data due to an unavailability or lack of operation of data collecting stations. Maidment (1992) has mentioned that many studies have attempted to generate time series of rainfall based on the theory of point processes. These studies have assumed storm arrivals are governed by certain random processes. Parameters of these random processes were estimated using available historical data. However, the incapability of these models to generate lower level data that are consistent with the higher-level data is identified as a shortcoming. For example, the aggregated monthly values based on generated daily values are not similar to the observed monthly values in most cases. Furthermore, there are situations where higher level (e.g. monthly) data is available through summary reports, etc. But it is difficult to incorporate this type of data into the models, which are demanding a lower level such as daily data. These situations require disaggregation and adjustment procedures to generate a lower level such as daily or hourly data. Many scientists have recognised these problems and this has led to the development of a range of weather generators such as WGEN (Richarsdon and Wright, 1984), TAMSIM (McCaskill, 1990) and WXGEN (Sharpely and Williams, 1990). The potential of WXGEN and other stochastic weather data generators to simulate long periods of climatological data from a small amount of input data has led to a number of applications (Hayhoe, 1998).

Disaggregation refers to the method of obtaining lower scale time series based on properties from higher scale time series. Rainfall disaggregation is emerging as an important tool for both hydrologists and engineers to understand the hydrological processes that occur in nature (Wong, 2000). The need for daily data for hydrological applications suggests the use of desegregation models to make use of the available monthly information and provide the possible realisations of daily precipitation, which aggregate up to the given monthly values. Although these simulated time series may not be the actual rainfall depths that fell to the ground in the past, their statistics are consistent with the actual time series as well as higher order time scale statistics. Koutsoyiannis and Onof (2001) proposed a method of coupling of a rainfall generator with an adjusting procedure to reduce the error between aggregated lower level data and high-level data. Application of this adjusting procedure with the rainfall generated using the Bartleet-Lewis model showed promising results in various studies (Koutsoyiannis and Onof, 2001; Wong, 2000).

SWAT includes the WXGEN weather generator model (Sharpley and Williams, 1990) to generate climatic data or to fill in gaps in measured records. WXGEN first independently generates the precipitation for the day and then generates the maximum and minimum temperature, solar radiation and relative humidity based on the presence or absence of rain for the day. This chapter gives details about the generating rainfall data using WXGEN model and modification introduced during this study by

integrating adjustment procedures to minimize the differences between observed higher level data with the aggregated lower level modelled data.

4.2. WXGEN Model

The rainfall generator in SWAT (WXGEN) used the first order Markov chain model developed by Nicks (1974). Model inputs include the probability of rain on a given day conditioned on the wet or dry status of the previous day. A wet day is defined as a day with rainfall greater than 0.1mm. Given the wet-dry probabilities, the model stochastically determines the occurrence of rainfall in a particular day. When a rainfall event occurs, the amount is determined by generating from a skewed normal daily rainfall distribution. The Markov chain model consists with two states namely wet and dry days. Simulations require four transition probabilities within and between states as shown in following table (Table 4-1).

Day (*i*) / Day (*i*-1)	Wet	Dry
Wet	$P_i(W/W)$	$P_i(W/D)$
Dry	$P_i(D/W)$	$P_i(D/D)$

Table 4-1 : Different transition probabilities used in WXGEN

$P_i(W/W)$ – Probability of a wet day on day *i* given a wet day on day *i*-1
$P_i(W/D)$ – Probability of a wet day on day *i* given a dry day on day *i*-1
$P_i(D/W)$ – Probability of a dry day on day *i* given a wet day on day *i*-1
$P_i(D/D)$ – Probability of a dry day on day *i* given a dry day on day *i*-1

Once two transition probabilities are inserted into the model, then the program derives the other two probabilities using the following two relationships:

$$P_i(D/W) = 1 - P_i(W/W) \quad (4.1)$$
$$P_i(D/D) = 1 - P_i(W/D) \quad (4.2)$$

To define a day as wet or dry, the model generates a random number between 0.0 and 1.0. This random number is compared to the appropriate wet-dry probability, $P_i(W/W)$ or $P_i(W/D)$. If the random number is equal or less than the wet-dry probability, the day is defined as wet. If the random number is greater than the wet-dry probability, the day is defined as dry.

Then the amount of rainfall in a wet day is calculated by using the following equation (4.3).

$$R_{day} = \mu_{month} + 2 \cdot \sigma_{month} \cdot \left\{ \frac{\left[\left(SND_{day} - \frac{g_{month}}{6} \right) \cdot \left(\frac{g_{month}}{6} \right) + 1 \right]^3 - 1}{g_{month}} \right\} \quad (4.3)$$

Where: R_{day} = amount of precipitation on a given day (mm)

μ_{month} = mean daily precipitation (mm) for the month

σ_{month} = standard deviation of daily precipitation (mm) for the month

SDN_{day} = standard normal deviation calculated for the day

g_{month} = coefficient of skewness for daily precipitation in the month.

The mean daily precipitation (mm) for the month is calculated as:

$$\mu_{month} = \frac{PCPMM}{PCPD} \quad (4.4)$$

Where:

PCPMM = average mount of precipitation falling in month (mm)

PCPD = average number of days of precipitation in month.

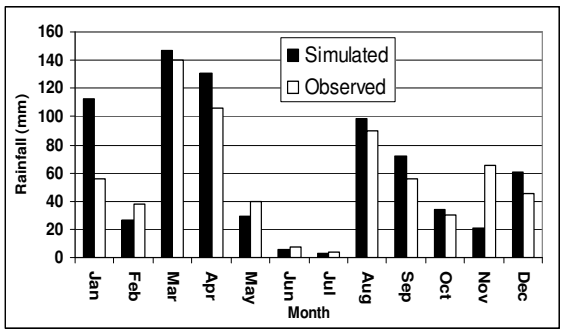
The standard normal deviate (SDN) for the day is calculated:

$$SDN_{day} = \cos(6.283 \cdot rnd_2) \cdot \sqrt{-2 \ln(rnd_1)} \quad (4.5)$$

Where rnd_1 and rnd_2 are random numbers between 0.0 and 1.0

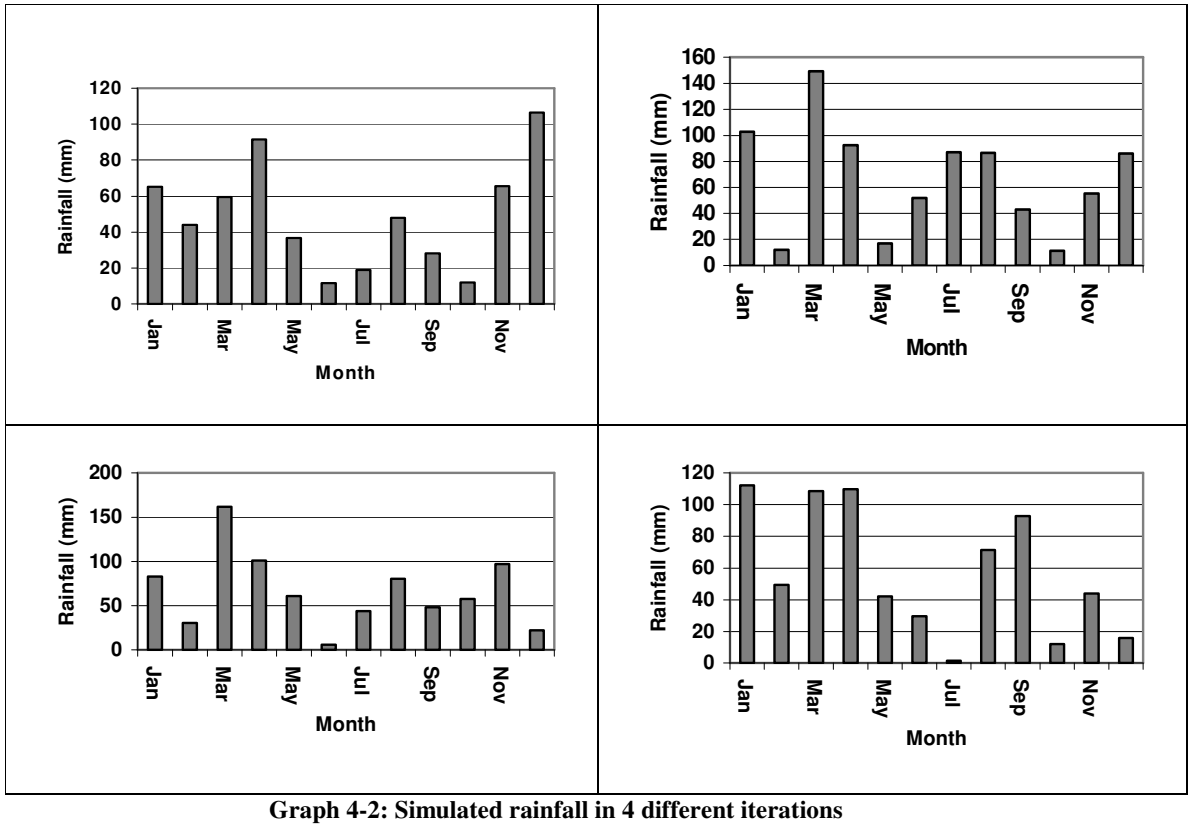
SWAT generates rainfall using the WXGEN model based on historical statistics. It requires the long-term monthly statistics such as mean rainfall and standard deviations. The model uses the same set of monthly statistics for generating long time series. Therefore, the modelling of long-term cycles, which are common in weather data, is not sufficiently addressed. Also, another apparent major problem is the aggregated lower level generated data (daily) are inconsistent with higher level (monthly) observed data. Some authors (Wallis and Griffiths, 1995; Hayhoe and Stewart, 1996) discussed about the unrealistic estimates of WXGEN model and its incapability of generating statistics such as mean and standard deviation consistent with the observed statistics. Therefore, to overcome those problems, an EXCEL spread sheet based WXGEN model developed in one of the previous studies (Lukman, 2003) was used with additional modifications. This new model has the capability to estimate daily rainfall (R_{day}) based on the observed monthly data in cases where monthly-observed data is available. Finally after the simulations, generated daily values in the series were replaced by actual daily values in the case where daily-observed data is available so that only the data gaps were filled by the model. In this way, the consistency between generated daily data and the observed monthly data can be maintained for the months which have a few days of missing data. However, in months with a higher num-

ber of days without data there are significant differences between aggregated daily data and the observed monthly data. Rainfall is the only mean of incoming water to most of the basins. In such cases, it is important to simulate accurate monthly totals in order to maintain the accuracy of the modelled monthly time series of stream flow. Graph 4-1 shows the observed monthly rainfall and aggregated monthly rainfall based on daily data simulated by WXGEN in the year 1960 at station no 9036002 in the Naivasha basin.



Graph 4-1:Simulated and observed monthly rainfall at station 9036002 for 1960

In January, the difference between observed and simulated rainfall is about 50mm. This could significantly affect the modelled stream flow. Only March, June, July and October show the close agreement between monthly totals of simulated rainfall and the observed monthly rainfall. This was based on the one realization of randomly generated rainfall. On the other hand, different repetition of the same simulation can generate different series of rainfall (graph 4-2). Therefore, in order to produce daily rainfall, which is fully consistent with the monthly data, it is vital to integrate the adjustment methodology with the rainfall generation model.



Graph 4-2: Simulated rainfall in 4 different iterations

The methodology proposed by Koutsoyiannis and Onof (2001) based on repetition and proportional adjustments was coupled with the EXCEL based model. This method adjusts the generated daily data

in order to be consistent with the monthly-observed rainfall data. Basic inputs to the rainfall generator model are mean daily precipitation for the month (PCPMM), Standard deviation for daily precipitation in month (PCPSTD), Skew coefficient for daily precipitation in the month (PCPSKW), *Pi (W/D), Pi (W/W)* and Average number of days of precipitation in the month (PCPD). These parameters were estimated (Table 4-2) for the stations where daily records were available for more than 10 years and used to estimate longer time periods by incorporating them into the developed WXGEN model.

4.3. Adjustment procedure

Since disaggregation is a simulation, the statistics generated may fluctuate depending on the parameters used as well as the random number utilised in the simulation. Therefore, repeating a simulation generates a different realization of rainfall. In the process of repetition, the final adjusted series was determined based on the logarithmic distance between the generated and observed high level data. Koutsoyiannis and Onof (2001) defined the logarithmic distance as in equation 4.6 in order to increase the calculation efficiency in their computer program named "Heyto" for hourly data. However, this can be accomplished by minimising the difference between observed and generated monthly total values of rainfall.

$$d = \left[\sum_{i=1}^{n} \ln^2 \left(\frac{Z_i + c}{\hat{Z}_i + c} \right) \right]^{1/2} \qquad (4.6)$$

Where Z_i and $\hat{Z}_i$ are the original and generated daily rainfall. The logarithmic sum is selected to avoid a domination of very high values and "c", a small constant, is inserted to avoid a domination of very low values. The term "*d*" is the difference between generated and observed monthly values. In the model, "*d*" is calculated for every repetition. Minimization of "d" eventually minimized the differences between observed and simulated data for each month so that simulated rainfall approximately preserves the statistical properties sum and mean value of the original series.

Wong (2000) stated that the "d" obtained with this kind of repetition is not a convergent sequence towards zero. Rather it is a random sequence with a certain probability for "d" being equal or smaller than a specific given value. Once this distance reaches the minimum threshold level, the repetition stops and the adjustment procedure is applied. Koutsoyiannis and Onof (2001) proposed 3 adjustment procedures namely *proportional*, *linear* and *power*. The proportional adjustment procedure is the simplest among the three methods. This method modifies the initially generated values to get the adjusted series according to the following equation.

$$X_s = \hat{X}_s \left[\frac{Z}{\sum_{j=1}^{v} \hat{X}_j} \right] \quad s = (1,2,\ldots\ldots\ldots\ldots v) \qquad (4.7)$$

Where $\hat{X}_s$ is a daily synthetic series that has been generated by the basic rainfall model, Z is the monthly amount of observed rainfall and *v* is the number of days in the month.

The following flow chart gives the methodology followed in this study and the case study shows the improved time series of rainfall data for one location in Naivasha basin.

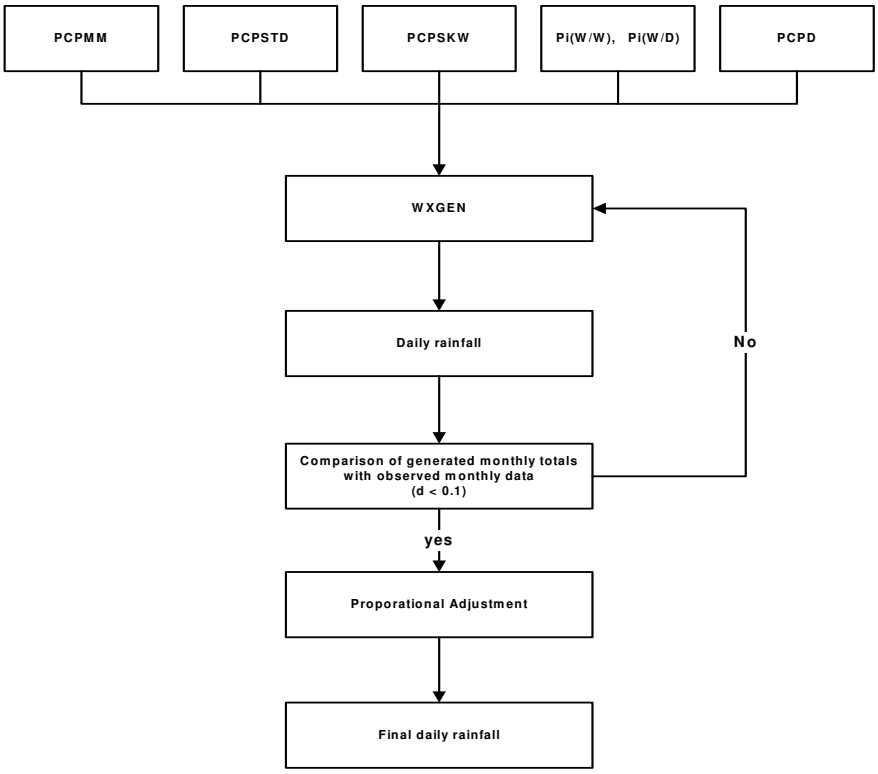


Figure 4-1: Integration of WXGEN with new repetition and adjustment procedure

First, the model generates daily rainfall data based on the given long term statistics and observed daily averages of rainfall. At this point the modelled data is adjusted using the logarithmic distance (Koutsoyiannis and Onof, 2001). The model runs several times until the adjustment creates the closest agreement between the generated rainfall time series and the observed high level (monthly) data. Once it has completed the repetitions, the next step is to proportionally adjust the generated time series using monthly totals. This second adjustment creates the final daily rainfall time series for the month. Therefore, this ensures that the generated results are close to the actual variables, thereby allowing accurate higher order statistics to be obtained.

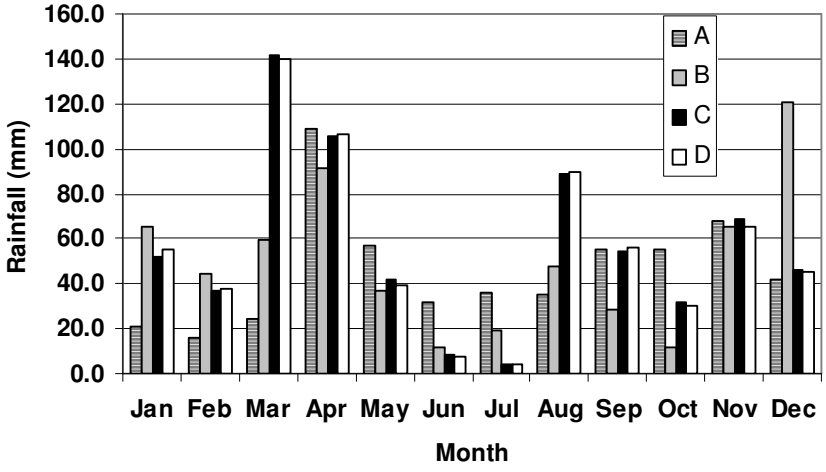
4.4. Case study

A case study was carried out to compare the performance of the weather generator after introducing the adjustment procedure into the original WXGEN model. This section discusses the results of that case study. Data of 1960 from rainfall station number 9036002 was used. All the statistics required by the model was calculated by using 20 years of available daily data. The Following table shows the input data generated based on the long-term statistics from that station.

Parameter	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
P(w/d)	0.13	0.14	0.2	0.32	0.24	0.18	0.13	0.18	0.2	0.22	0.28	0.19
P(w/w)	0.45	0.41	0.48	0.65	0.53	0.46	0.43	0.42	0.35	0.48	0.57	0.45
Mean	30.3	35.4	54.0	114.9	70.2	43.8	32.1	39.9	42.0	51.0	63.3	37.5
SDN	3.92	4.14	5.04	7.38	6.3	4.43	4.18	4.26	4.43	4.39	4.82	3.6
Skew	5.34	5.63	4.58	3.04	4.21	5.98	6.48	5.72	5.12	4.30	4.14	4.73
Number of rainy days	4.66	5.19	8.29	13.19	9.49	7.00	5.29	6.84	6.65	8.68	11.16	7.28

Table 4-2: Input statistics for the weather generator at station number 9036002 derived from long term rainfall data

Initially, the original WXGEN model was used to generate the rainfall from January to December in 1960. This model generates daily rainfall using the equation mentioned in the above section. Also, it uses the monthly statistics given in the table 4-2 (Appendix 4.1 presents the statistics for other 3 stations). This method assumes the stationarity within months, i.e, rainfall in the same months of different years has the same statistical properties. As mentioned earlier, in case of monthly available data, this can be modified by replacing mean daily rainfall calculated separately for each month based on actual values. Then in the next step, rainfall was generated for the same period by estimating daily mean rainfall for each month based on observed monthly rainfall values. In the third step, the above model was modified by integrating repetitions and adjustment procedures. First the rainfall generator was repeatedly run until "d" becomes a small value less than 0.1 for each month of the year. Then the proportional adjustment was applied in order to make the generated rainfall values to be consistent with the observed monthly values. The following graphs shows the results of the above simulations



A – Simulated rainfall based on the original WXGEN model

B - Simulated rainfall after introducing the mean daily rainfall based on observed monthly data

C - Simulated rainfall after introducing the repetition scheme

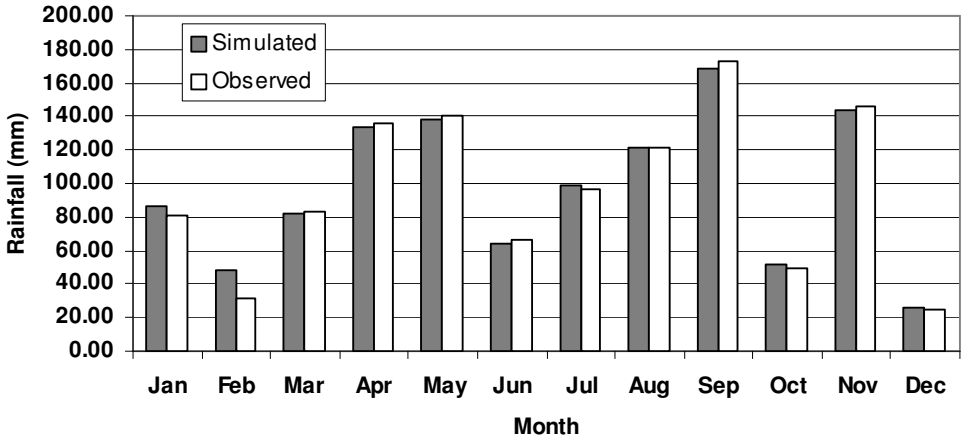
D – Observed monthly rainfall

Graph 4-3: Rainfall generated at the different stages of the modifications and observed rainfall (1960- Station 9036002)

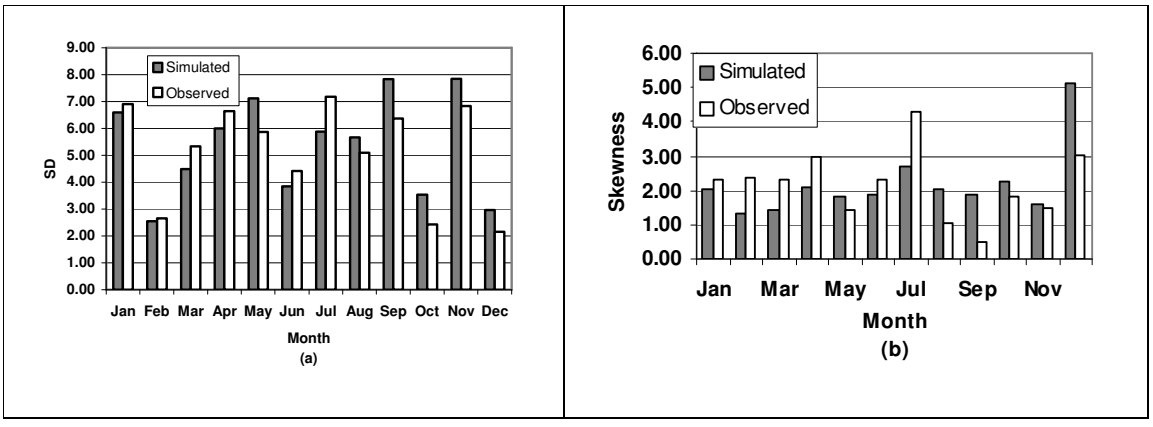
Rainfall simulated using the original WXGEN (**A**) model shows significant variations with the observed monthly data (**D**). This simulation is solely based on long term statistics. Random numbers

generated by the program controls the differences of rainfall amounts generated between different simulations. Since there is no information provided for specific months, generated rainfall can deviate from the observed monthly values. Improvement can be seen after introducing calculated mean daily rainfall for each month using available monthly data (**B**). In this case also the number of rainy days was calculated using available long term data. Mean daily rainfall for each month was calculated by dividing observed monthly total rainfall with the above number. However, in this case also, the effects of random generation are observed. For example, in December the model simulates very high values. Introducing repetitions (**C**) shows the close agreement between observed and simulated rainfall. This can be then adjusted proportionally to generate the exact monthly rainfall. For this year there were no daily data available for some months. Therefore, comparing statistics was impossible.

Observed and simulated data and statistics were compared for the 1960 data at station 9036241. Graph 4-4 indicates the close agreement between the observed and generated data after introducing repetitions. Standard deviations and skewness of the observed and simulated rainfall for each month are presented in graph 4-5**.**



Graph 4-4: Simulated and observed monthly rainfall



Graph 4-5 : Observed and simulated standard deviation, skewnesss (1960-station 9036241)

Preserving the observed statistic is highly important in case of modelling daily runoff based on simulated rainfall. Graph 4-5 (a and b) indicates that model performances are poor in terms of preserving original statistics. This study compares monthly runoff calculated based on daily values simulated by the SWAT model with the observed monthly runoff values. However, this would not significantly affect the simulated monthly surface runoff from the basin other than in cases where the simulated rainfall is extremely high at the last day of the month. Rainfall generated using the modified WXGEN model was integrated into SWAT model to simulate the run off time series of the Naivasha basin and will be presented in Chapter 6.

5. Hydrological Modeling and Soil and Water Assessment Tool (SWAT)

5.1. Introduction

A hydrological model is an attempt to describe the physical processes controlling the transformation of precipitation to runoff (Al-Sabbagh, 2001). Continues stream flow simulation models range in complexity from the very simple antecedent precipitation (API) and tank model to the very complex distributed parameter models such as *Hydrologic Simulation Program-Fortran* (HSPF) and *European Hydrologic system model* (SHE)(Maidment,1992).The major hydrologic processes described by these models include: Canopy interception, Evapotranspiration, Snowmelt, Interflow, Overland flow, Channel flow, Unsaturated subsurface flow and saturated subsurface flow. These processes are subjected to vary through space as well as time. Some models, describe some of the processes by differential equations based on simplified hydraulic laws and other processes are expressed by empirical algebraic equations (Arnold, 1998). Traditionally, hydrologic models have considered watershed to be homugeneous (Kilgore, 1997).Weighted averages or mean values are used as inputs to these "*lumped*" models. The major drawback of lumped models is the incapability to account for spatial variability. As computers had become more powerful and less expensive, many hydrologists began using distributed parameter models. These models offer the possibility of a significant improvement over lumped models due to the ability to integrate spatial variability of hydrological processes.

A wide variety of hydrological models as well as applications of them has been developed over the past decades. The common feature of such models is that they can incorporate the spatial distribution of various inputs and boundary conditions, such as topography, vegetation, land use, soil characteristics, rainfall and evaporation, and produce spatially detailed outputs such as soil moisture fields, water table positions, groundwater fluxes and surface saturation patterns (Troch et.al, 2003). In those models, spatial variations are approximated by spatial variation of precipitation, catchment parameters and hydrologic responses. Temporal variations of hydrologic responses are modelled by introducing threshold values for different processes to occur or not. Representation of the catchments by individual subbasins or grids of individual elements are used to integrate the spatial variability of the abovementioned parameters with the model. Vertical variability is represented by subsurface zones or vertical layers of soil for each grid element. However, these models are yet to become common planning or decision tools. A majority of watershed models simulated watershed responses without or with inadequate consideration of water quality (Arnold, 1998). On the other hand some authors (Jain et.al.,1992, Troch et.al, 2003) stated that the substantial data requirement for the available models was a shortcoming.

5.2. Soil and Water Assessment Tool (SWAT)

The SWAT model—Soil and Water Assessment Tool—(Arnold et al., 1998) is a semi-distributed watershed model with a GIS interface that outlines the sub basins and stream networks from a Digital Elevation Model (DEM) and calculates daily water balances from meteorological, soil and land-use data. The objective of the development of SWAT was to predict the impact of management on water, sediment and agricultural chemical yields in large basins (Fontaine, 2002). Model components include hydrology, weather, sedimentation, soil temperature, crop growth, nutrients, pesticides and agricul-

tural management. This study usedthe SWAT model to estimate spatial variations of surface runoff from the different parts of Naivasha basin. This section gives a brief over view of the hydrologic component integrated in the SWAT model. The important equations are summarized here; complete descriptions are given in Arnold et.al (1998)

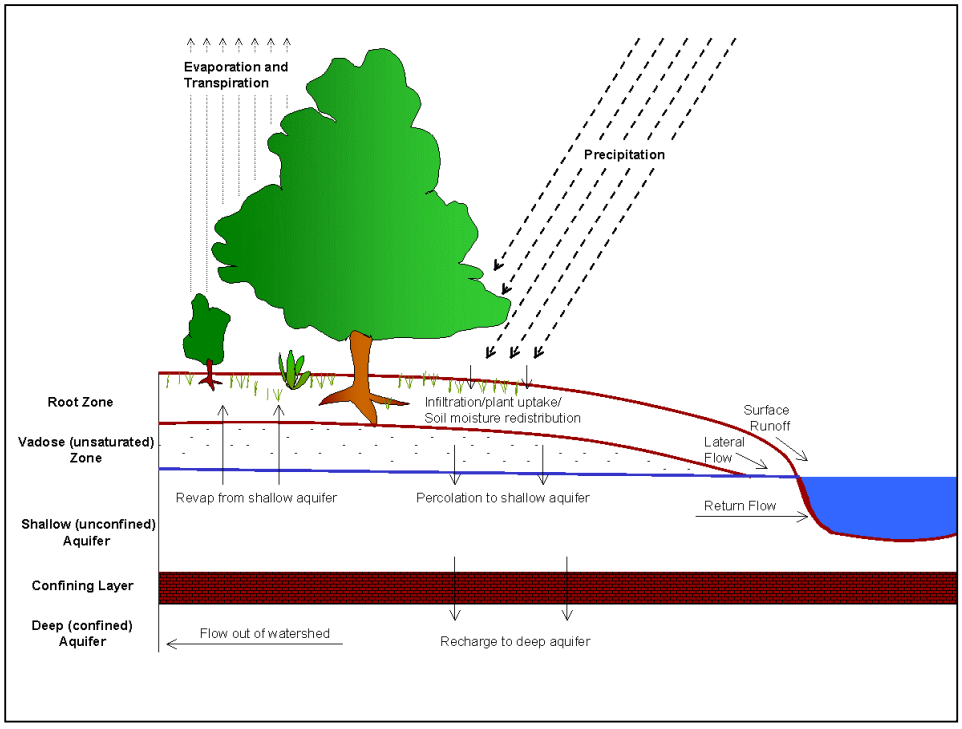


Figure 5-1: Schematic representation of the hydrologic cycle

The computation of hydrologic processes operates in five main steps: (1) Precipitation interception, (2) Surface runoff, (3) Soil and root zone infiltration, (4) Evapotranspiration and (5) Ground water flow (Figure 5.1). The model is based on a routing command language, which allows definition of how the water budget moves inside the catchment, relating the different units spatially (i.e, subbasins, reservoirs, ponds, river reaches) considered (Francos et.al.,2001). The water budget for each subbasin is simulate based on the following equation (5.1).

$$SW_t = SW_0 + \sum_{t=1}^{t} (R_{day} - Q_{surf} - ET_a - W_{seep} - QR_{gw}) \quad (5.1)$$

Where SW_t is the final soil water content (mm H2O), SW_0 is the initial soil water content on day i (mm H2O), t is the time (days), R_{day} is the amount of precipitation on day i (mm H2O), Q_{surf} is the amount of surface runoff on day i (mm H2O), E_a is the amount of evapotranspiration on day i (mm H2O), W_{seep} is the amount of water entering the vadose zone from the soil profile on day i (mm H2O), and Q_{gw} is the amount of return flow on day i (mm H2O).

5.2.1. Surface runoff

SWAT offers the SCS curve number equation and Green - Ampt infiltration method to estimate the surface runoff volume. Fontaine (2002) mentioned the intensive data requirements of Green-Ampt method and the performance of simpler and les data intensive method such as SCS curve number method over the Green-Ampt method. In this study, the SCS curve number method was used to estimate runoff volumes. This method was extensively discussed in various books and research publications (Dingman, 1994, Arnold, 1998, Lukman, 2003). However, one distinct feature in the SCS curve number method with respect to the Green-Ampt method is, it lumps canopy interception in the term initial abstraction. Then SWAT estimates the peak runoff rate, time of concentration for overland and channel flow and surface runoff lag separately for each subbasin (complete description is given in Neitsche, 2002).

5.2.2. Transmission losses

Many semiarid watersheds have alluvial channels that abstract a considerable amount of stream flow (Arnold, 1998). The abstractions or transmission losses reduced the stream flow as it travels towards down stream. SWAT used *Lane* (Neitsche, 2002) method based on channel width, length and flow duration. Both runoff and peak rates are adjusted when transmission losses occur.

Runoff is calculated separately for each subbasin and routed to obtain the total runoff for the basin as shown in figure 5-2. This gives a better physical description of the water balance.

5.2.3. Soil water

Excess precipitation that remains after runoff infiltrates into the soil profile, SWAT calculates the vertical water movement for each soil layer in the profile. Water is allowed to percolate if the water content exceeds the field capacity water content of the layer. The amount of water that moves from one layer to the underlying layer is calculated using storage routing methodology. The Following equation (5.2) is used to calculate the amount of water that percolates to the next layer:

$$w_{perc,iy} = SW_{ly,excess}\left(1-\exp\left[\frac{-\Delta t}{TT_{perc}}\right]\right) \qquad (5.2)$$

Where: $w_{prec,ly}$ is the amount of water percolating to the underlying soil layer on a given day (mm), $SW_{ly,excess}$ is the drainable volume of water in the soil layer on a given day (mm), Δt is the length of the time steps (hrs) and TT_{perc} is the travel time for percolation.

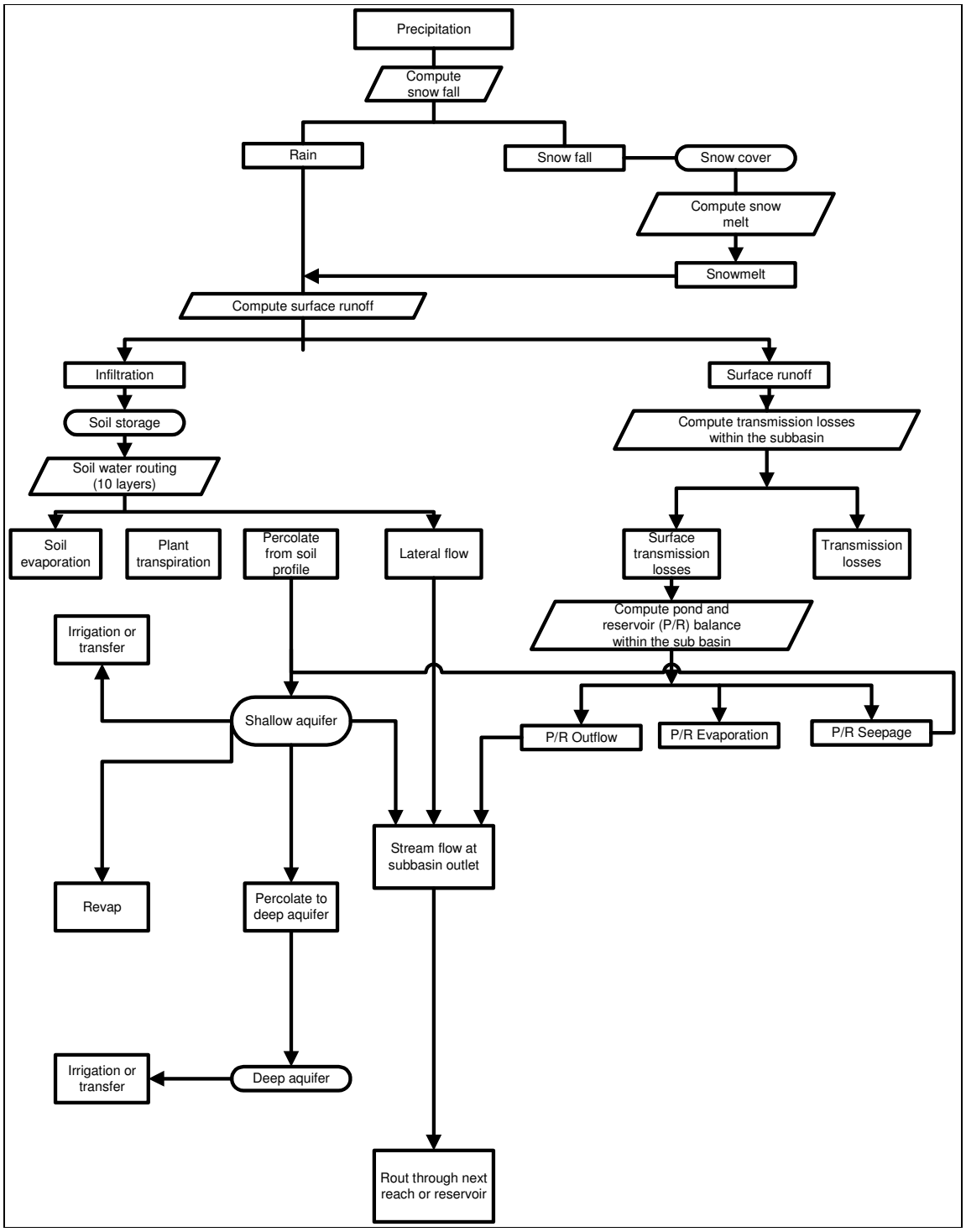


Figure 5-2: Over view of SWAT hydrologic component (Adupted from Arnold et.al,1998)

At the same time SWAT applies a multiplayer storage routing technique to partition drainable soil water content for each layer into other components, which is lateral subsurface flow. Lateral flow is significant in areas with soils having high hydraulic conductivity in surface layers and an impermeable or semi permeable layer at a shallow depth. In such a system, rainfall percolates vertically until it encounters the impermeable layer. This water forms a saturated zone after ponding above the imper-

meable layer. This saturated zone is the source of water for lateral subsurface flow. A kinematic storage routing technique is used to calculate lateral subsurface flow as a function of soil slope, hill slope length, drainable porosity, and excess soil water as given in the following equation (5.3).

$$q_{lat} = 0.024\left[\frac{2 \cdot SW \cdot K_{sat} \cdot \sin(\alpha)}{\phi_d \cdot L_{hill}}\right] \qquad (5.3)$$

Where, q_{lat} is lateral flow (mm d^{-1}), SW is drainable volume of soil water (mh^{-1}), α is slope (m/m), ϕ_d is drainable porosity (m/m), K_{sat} is saturated hydraulic conductivity (mmh^{-1}) and L_{hill} is flow length (m).

In the large subbasins with a time of concentration greater than one day, only a portion of the lateral flow will reach the main channel on the day it is generated (Neitsche, 2002). SWAT calculates a total lateral flow as a summation of two fractions contributing from a given day and the previous day.

5.2.4. Ground water

SWAT assumes two aquifers to address ground water: a shallow unconfined aquifer and a deep confined aquifer. Ground water movement in shallow aquifer is represented by three processes: upward migration (and subsequent ET), seepage to a deep aquifer and ground water flow to a stream (Fontain et.al., 2002) Water percolating to the deep aquifer is lost from the system. Both upward migration and percolation to the deep aquifer are control by the shallow aquifer storage, and by coefficients that regulate the water losses. Contribution of ground water to stream-flow is simulated by creating a shallow aquifer with, storage, recharge by percolation from the unsaturated zone, and discharge to the streams. Water balance for the shallow aquifer is given in the following equation (5-4).

$$Vsa_i = Vsa_{i-1} + Rc - revap - rf - prec_{gw} - WUsa \qquad (5.4)$$

Vsa is the shallow aquifer storage (mm), ***Rc*** is recharge (mm) (percolated from the bottom of the soil profile), ***revap*** is the root uptake from the shallow aquifer (mm), ***rf*** is the return flow (mm), ***WUsa*** is the water use (withdrawal) from the shallow aquifer (mm) and " i" is the day number.

Ground water flow into the main channel on day "*i* "is calculated using following equation.

$$rf_i = rf_i e^{-\alpha\Delta t} + Rc(1 - e^{-\alpha\Delta t}) \qquad (5.5)$$

Where:

α is the recession constant which describes the lag flow from the aquifer and Δt is the time step. This can be estimated by analysing the measured stream flow during periods of no recharge in the watershed.

5.2.5. Evapotranspiration

SWAT offers three options to estimating potential evapotranspiration : Hargreaves , Priestly and Taylor and Penman-Monteith method. These methods vary in the amount of required inputs. Due to the extensive data requirement by Penman-Monteith, this study used Hargreaves method to calculate potential ET.

Hargreaves (1975) proposed following equation (5.6) to estimate potential ET:

$$\lambda E_0 = 0.0023 \cdot H_0 \cdot (T_{\max} - T_{\min})^{0.5} .(\overline{T}_{av} + 17.8) \qquad (5.6)$$

Where: λE_0 is the latent heat of vaporization (MJ kg^{-1}), E_0 is the potential evapotranspiration (mmd^{-1}),H_0 is the extraterrestrial radiation ($MJm^{-2}d^{-1}$), T_{max} is the maximum air temperature for a given day (^0C), T_{min} is the minimum air temperature for a given day (^0C), and $\overline{T}_{av}$ is the mean air temperature for a given day.

Another main feature of SWAT hydrological section is the weather generator. SWAT accepts daily data on precipitation, maximum and minim temperature as inputs. In cases where these data are available, the weather generator estimates them based on long term monthly statistics. The description of the rainfall generator is given in chapter 4.

5.2.6. Revap

Water may move from the shallow aquifer into the overlying unsaturated zone. In periods when the material overlying the aquifer is dry, water in the capillary fringe that separates the saturated and unsaturated zones will evaporate and diffuse upward. As water is removed from the capillary fringe by evaporation, it is replaced by water from the underlying aquifer. Water may also be removed from the aquifer by deep-rooted plants which are able to uptake water directly from the aquifer. SWAT models the movement of water into overlying unsaturated layers as a function of water demand for evapotranspiration. In SWAT terminology, this process has been termed 'revap'. This process is significant in watersheds where the saturated zone is not very far below the surface or where deep-rooted plants are growing. Because the type of plant cover will affect the importance of revap in the water balance, the parameters governing revap are usually varied by land use. SWAT modelled the maximum amount of water than will be removed from the aquifer via 'revap' on a given day as in the following equation (5.7).

$$W_{revap,mac} = \beta_{revap} .E_0 \qquad (5.7)$$

Where: $W_{revap,\max}$ is the maximum amount of water moving into the soil zone in response to water deficiencies (mm H_2O), β_{revap} is the revap coefficient, and E_o is the potential evapotranspiration for the day (mm H_2O).

In addition to hydrology, the SWAT model consists of components for sedimentation, crop growth, nutrients, pesticides and agricultural management, which are not discussed here.

5.3. Model setup

AvSWAT2000 initially generates various spatial data required by the model. The DEM generated using ASTER data, with a cell-size of 15 m, was used to discretize the basin into subwatersheds. Available measured stream flow data from 2 gauging stations were used for calibrating the simulated stream flow for Gilgil and the whole basin. Estimated soil water capacities were assigned to each soil. Initially, CN numbers were assigned approximately based on the results of base flow analysis and the real ground conditions identified during the field investigations. Daily rainfall data was available for 4 meteorological stations for longer periods (Table 5-1). For same stations, monthly rainfall data was available from 1935. Monthly data was disaggregated into daily data based on the procedure described in chapter 4. The following section described the subbasin delineation and the SWAT input preparation.

Station ID	Station Name	X (m)	Y (m)	Elevation (m)	Data availability
9036002	Naivasha DO	214315	9920714	1900	1935-1998
9036025	North Kinangop Forest Station	238582	9935474	2629	1957-1998
9036241	Geta Forrest Station	207148	9948369	2591	1958-1998
9036264	North Kinangop Mawingo Scheme	223586	9944688	2484	1964-1998

Table 5-1: Rainfall data availability at 4 stations

In order to calculate subwatershed parameters such as area, slopes and channel lengths, it is important to project the maps into a correct projection system. Therefore, all the maps were projected to the Kenyan local projection system using the parameters given in table 5-2.

Projection	**Transverse Mercator**
Spheroid	**Clark 1880**
Datum	**Arc 1960**
Zone	**37**
Central Meridian	**39**
Reference Latitude	**0**
Northing (meters)	**10000000**
Easting (meters)	**500000**
Scale factor	**0.9996**

Table 5-2: Transverse Mercator projection parameters for Kenya

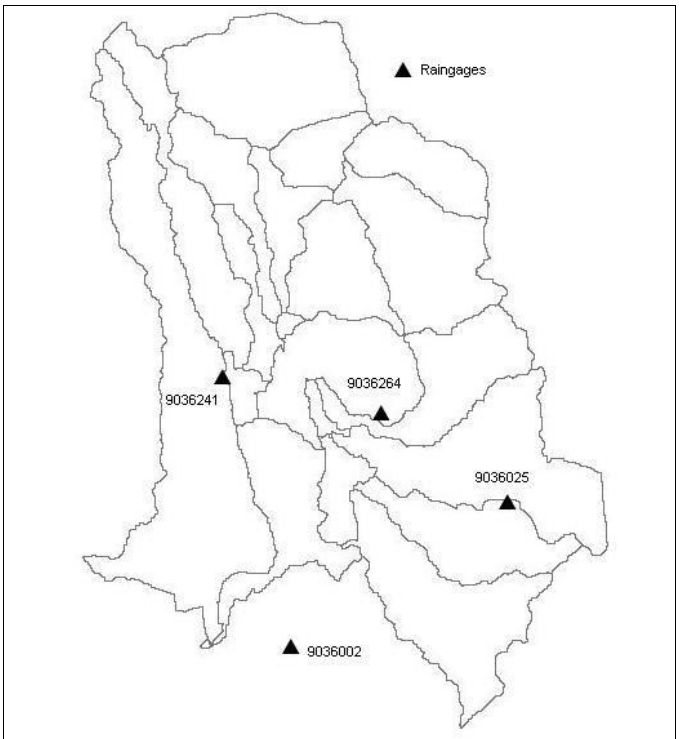


Figure 5-3: Locations of rainfall stations

5.3.1. Subwatershed delineation

Automatic watershed delineation tool in SWAT discretized the basin into subwatersheds and generates the stream network based on flow directions and flow accumulations. This method is based on the eight-pour algorithm proposed by Puecker (1978). During the process, the program fills all the sinks in the DEM in order to create a flow vector. In this way, the elevation of the sinks increases until the overflow occurs. The flow accumulation grid is created by counting number of flowing cells into each unit in the grid. High values of flow accumulation are associated with the cells, which are potentially part of the stream network.

5.3.2. Hydrological Response Units

Subbasins can further be subdivided into hydrological response units (HRU), each of which represents a particular combination of soil and land-cover within the subbasin. HRUs can be defined in two ways namely: the dominant approach and the virtual basin approach (Sirinivasn, 1998). The dominant approach creates one HRU for each sub basin based on the most prevailing land use class and the soil class. The virtual approach creates one or more HRUs for each subbasin based on the given threshold values on percentage areas of land use and soil. This study used 10% and 5% threshold values for land use and soils, respectively. SWAT first determines the land use classes which exceed or are equal to 5% of the subbasin area, then creates a HRU for each soil type that exceeds or is equal 10% of the area inside the land use class. These percentage thresholds causes the number of HRUs to increase, as the number of subbasin is increased Terrain parameters are identical for all HRUs within a given sub-

basin, except for the channel length parameter used to compute the time of concentration, which varies depending on the size of the HRU.

During the delineation, AvSWAT automatically estimates the total number of HRUs, channel length and average slope of the channel, which demand extensive time in manual techniques. These extracted parameters are stored into a SWAT subbasin input file. The model creates a channel network through all delineated subbasins. The flow accumulation threshold can control the number of subbasins. In this study, in addition to the automatically generated subbasin outlets, some of the outlet positions were manually placed at the gauging stations in order to use the discharge for the calibration. A digital stream network of the basin was available from the previous study.

6. Model application and results

6.1. Introduction

As mentioned in the earlier chapters, the GIS integrated SWAT model (AVSWAT 2000) was initially applied to the Gilgil sub basin and then to the total Naivasha river basin. Since there was a simultaneously carried out hydrological study (Dapp and Vreugdenhil, 2004), the Gilgil subbasin was selected for the initial application. First the newly created input data (Land use, DEM and Soil) were integrated with the GIS extension to create spatial inputs to the SWAT model. Then the newly generated rainfall data was integrated. Test period was selected as 1965-1975. Even though more recent stream flow data were available, they are considered to be unreliable, and hence were not utilized for calibration. For the calibration, simulated stream flow was compared with the observed flow at the gauging stations. As mentioned by many authors (Eckhart et.al, 2001 and Arnold), complex hydrologic models are generally characterized by a multitude of parameters. Due to spatial variability, measurement error, etc.,the values of many of these parameters are not exactly known. Therefore, in most cases model calibration is necessary. In order to calibrate the model, different parameters were changed on a trial and error basis until the observed and simulated stream flows satisfactorily agreed with each other. Final calibration was automatically performed using a parameter estimation (PEST) computer software program for the test period. Long-term simulation from 1935-1975 was carried out using the optimised parameters found in the parameter estimation. Those simulated values were incorporated with the lake water balance model to estimate lake water levels.

6.2. Automatic parameter estimation

PEST is a non-linear parameter estimation program, which can easily be linked via templates to any model. Furthermore, it is independent from the base model (e.g ,SWAT) and has advanced predictive analysis and regularization features (Doherty, et.al,2003). PEST runs the particular model through an interface file between itself and the base model until the difference between observed and simulated values approaches to a minimum value. In each run or iteration, it adjusts selected parameters of the base model using an optimisation algorithm.

PEST defines the optimal parameter set as that for which the sum of squared deviations between simulated and observed values, referred to as the *objective function*, is reduced to a minimum. The objective function can be represented mathematically as follows:

$$\phi(\vec{b}) = \sum_{j=1}^{m} \left[\sum_{i=1}^{n_j} \left[w_{ij} (Y_j(t_i) - Y'_j(t_i, \vec{b})) \right]^2 \right] \qquad (6.1)$$

Where: ϕ is the objective function, $\vec{b}$ is the vector with fitting parameters, m represents the different sets of observation groups, n_j is the number of observations in j^{th} observation group, $Y_j(t_i)$ is the observation of type j at time t_i, $Y'_j(t_i, \vec{b})$ is the corresponding model prediction and w_{ij} is the weight associated with a particular kind of measurement at a particular point and accounts for the role of data

type and data point in the objective function. In case of only one observation group, the above equation can be simplified as follows:

$$\phi(\vec{b}) = \sum_{i=1}^{n} \left[w_i (Y(t_i) - Y'(t_i, \vec{b})) \right]^2 \qquad (6.2)$$

PEST is supplied with initial values of the parameters to be optimised and the set of observations, which PEST can compare with the model outputs. First, PEST executes SWAT with the given initial parameter values. Output of SWAT is compared with the given set of observed data and construct the objective function. Then the PEST starts minimizing the objective function by adjusting parameters. To do that, it uses the Gauss-Marquardt-Levenberg algorithm. In case of linear models, optimisation can be achieved in one step. But for non-linear models, parameter estimation is an iterative process. At the beginning of the each iteration the relationship between model parameters and model-generated outputs is linearised by formulating it as a Taylor expansion about the currently estimated best parameter set. This linearised problem is then solved for a better parameter set and this new set of parameters is tested by running the base model (SWAT) again. By comparing the objective function of the current iteration with the previous one, PEST decides whither it is worth undertaking another iteration (Doherty, 2002). Figure 6-1 describes the integration of PEST with the Soil Water Assessment Tool.

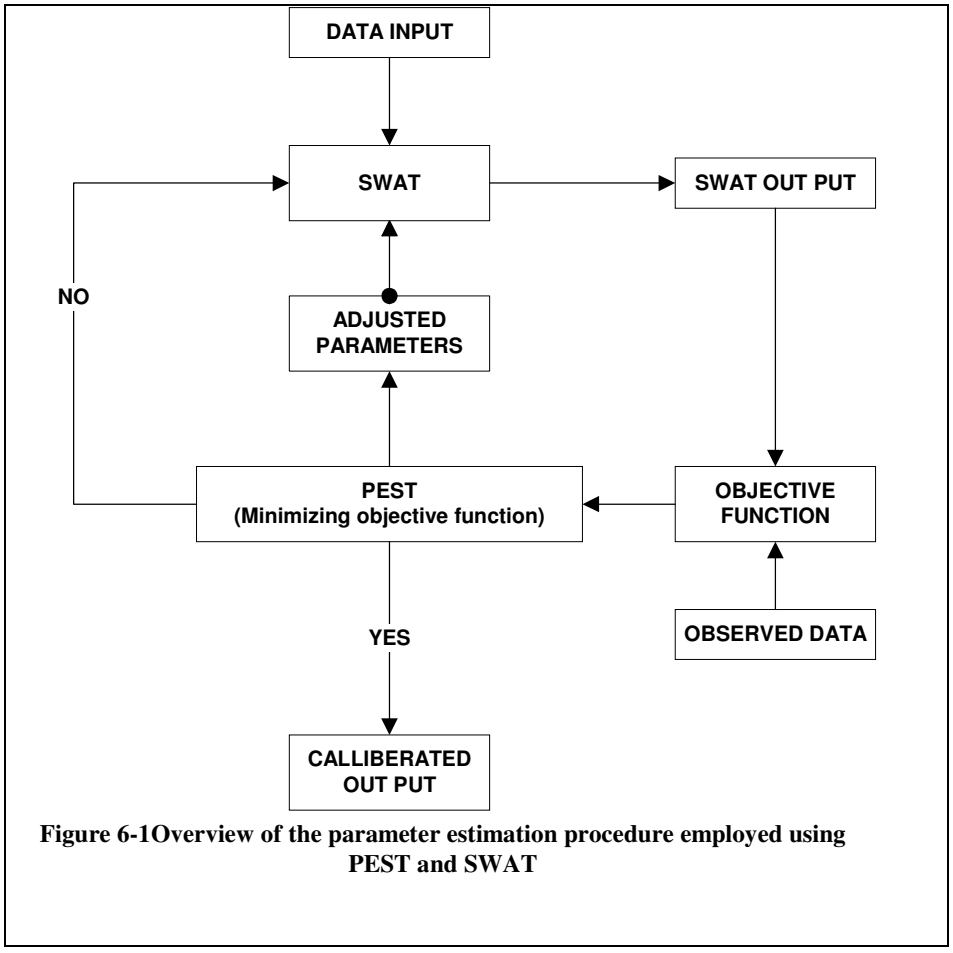


Figure 6-1Overview of the parameter estimation procedure employed using PEST and SWAT

Generally, the model calibration consists of adjusting many parameters until it finds a better agreement between observed and modelled values. Therefore, it generates a system of equations. PEST uses matrix algebra to solve those equations. In such case the derivatives needed in Taylor's expansion is replace by the Jacobian matrix of the functional relationship with respect to the parameters to be estimated (Doherty, 2002). After the manual calibration of the model, PEST was used to perform further calibration.

6.3. Evaluation of model predictions

In order to decide if the simulated results are sufficiently acceptable, various statistical and graphical procedures are available. The Nash_Sutcliffe coefficient is a one way of doing it (Strobl, 2002).

$$R^2{}_{NS} = 1 - \frac{\sum_{i=1}^{n}(Q_i - Q_i')^2}{\sum_{i=1}^{n}(Q_i - \overline{Q}_i)^2} \qquad (6.3)$$

Where, Q_i is the measured monthly discharge, Q' is the model simulated monthly discharge, $\overline{Q}_i$ is the average measured discharge for the period and n is the number of months during the period. This coefficient can vary between negative infinity to one where one indicates a perfect fit between observed and simulated flow values.

In addition to the Nash_Sutcliffe coefficient, there were two other statistical measures calculated, namely percentage deviation of volume (*DV*) and root mean square error (RMSE), which are given by equation 6.4 and 6.5 was calculated.

$$DV = \frac{V - V'}{V} \times 100 \qquad (6.4)$$

V is the observed yearly or seasonal volumes and the V' is the simulated yearly or seasonal volume
For a perfect fit between the observed and modelled series, DV equals zero. Negative and positive values of DV indicate over prediction and under prediction, respectively.

$$RMSE = \left[\frac{\sum(Q_i' - Q_i)^2}{n}\right]^{0.5} \qquad (6.5)$$

Q_i' is the predicted flow, Q_i is the observed flow and n is number of values. Smaller RMS values indicate a better fit. However, the scale of RMSE depends on the units of the variable under scrutiny.
Wilmott (1984) has suggested reporting the unsystematic ($RMSE_u$) and systematic (RMSEs) components of RMSE (Equations 6.6 and 6.7).

$$RMSE_s = \left[\frac{\sum (\hat{Q}_i - Q_i)^2}{n} \right]^{0.5} \qquad (6.6)$$

Where $\hat{Q}_i = a + bQ_i$, a and b are parameters associated with an ordinary least squares simple linear regression between Q_i and Q_i'

$$RMSE_u = \left[\frac{\sum (Q_i' - \hat{Q}_i)^2}{n} \right]^{0.5} \qquad (6.7)$$

Further, these two components are related to the RMSE as follows

$RMSE^2 = RMSE_u^2 + RMSE_s^2$

Ideally, the unsystematic, random error of the model should approach the value of the RMSE, while the systematic component should be very small. Because in the opposite case this would indicate that a systematic error is present the model, which in turn would mean that a better model could be found.

6.4. Modeling the Gilgil subbasin

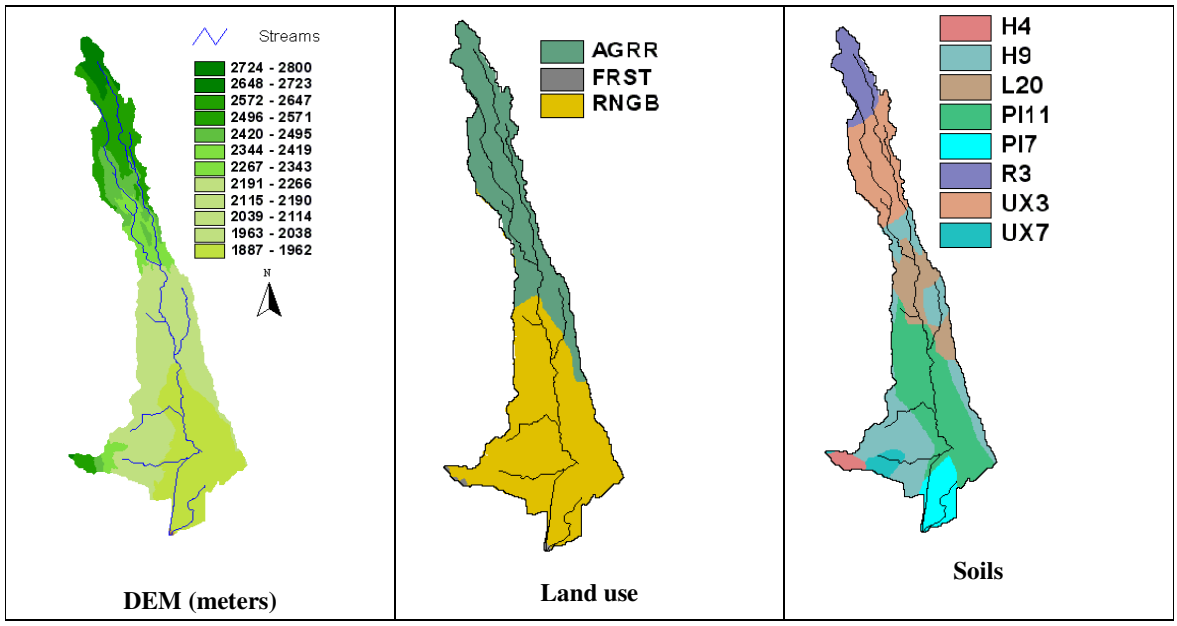


Figure 6-2:a,b,c: Digital elevation model, Land use and soil maps in Gilgil sub basin

Elevation ranges from 1887 – 2800 meters in the Gilgil sub basin. Land use was re classified into 3 broad categories, which are compatible with the SWAT naming convention. Upper areas are predominantly occupied by Agricultural land, while the lower area is mostly Range brush land. There were 8 soil categories found inside the basin. As discussed in chapter 2, these soil classes are supplemented by estimated physical properties such as available soil water, hydrologic soil group etc. The area was delineated into 19 sub watersheds (Figure 6-2) by assuming a flow accumulation threshold of 600 ha.

During the process, one watershed outlet was manually added at the gauging station 2GA5. Simulated flow at the outlet was compared with the observed flow.

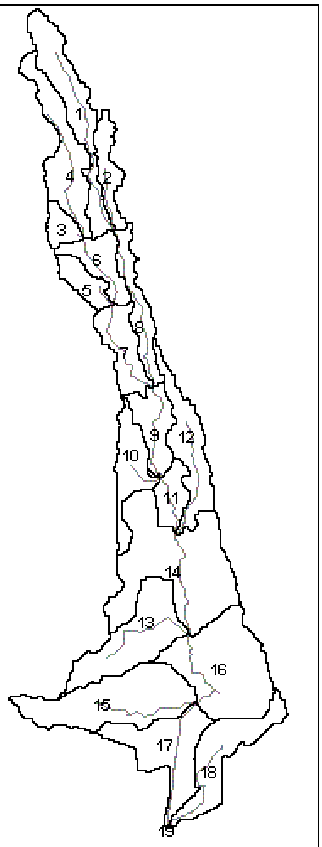


Figure 6-3: Subwater sheds in Gilgil

During the watershed delineation, AvSWAT200 automatically assigned the parameters such as Manning's roughness coefficient and curve number for each delineated sub watershed. Initially, the simulation was carried out by modifying some of these parameters based on the findings of stream flow analysis (Chapter 3) and the identifications made in the field.

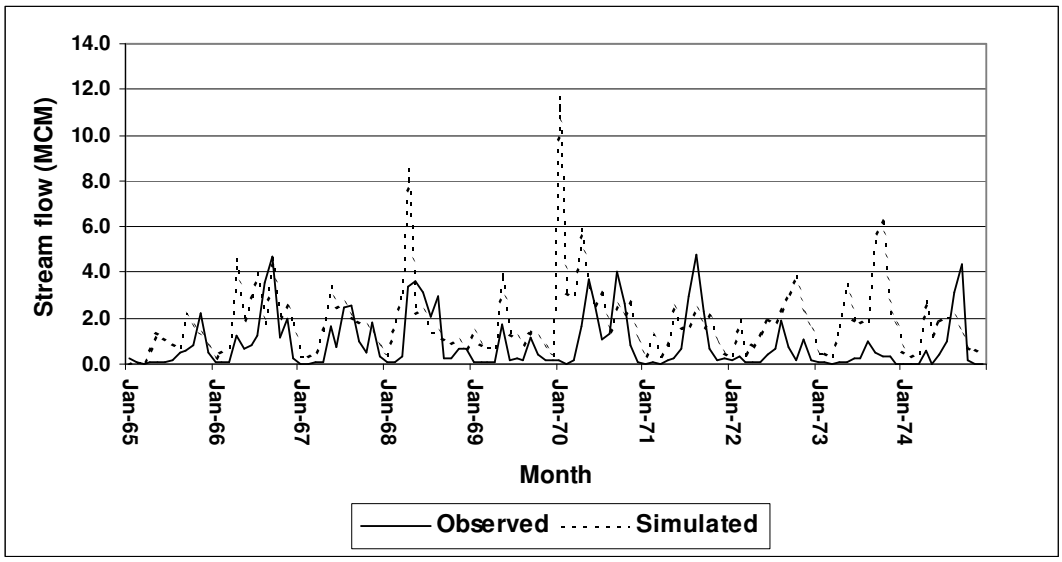
Available daily rainfall data at the stations 9036002, 9036025, 9036241 and 9036264 were used. Gaps of the rainfall data were filled by using the Excel based WXGEN model. Temperature, Solar radiation and Wind speed were estimated through an original weather generator based on long term statistics. SWAT assigns rainfall at the nearest stations to sub watersheds. Since 9036241 and 9036202 are the closest stations to the Gilgil subbasin, SWAT used rainfall and weather data only from those 2 stations for simulations.

Dapp and Vreugdenhil (2004) identified CN number, revap coefficient, and sol_k as most sensitive parameters to the model. As shown in table 6.1, for the initial simulation, the CN number was assigned as 75 for the whole basin. The GW_REVAP coefficient was kept at the lowest possible value of 0.02 and Saturated hydraulic conductivity of the soil was assigned as 40 mm/hr.

Parameter	Value	Description
CN2	75.00	Curve number
GW_REVAP	0.020	Groundwater "revap" coefficient
SOL_K	40 mm/hr	Saturated hydraulic conductivity of the soil

Table 6-1:Initial parameter values selected for the manual calibration

Saturated hydraulic conductivity is a little higher for the clay loam soils according to the values found in the literature. But, as mentioned in chapter 3, the upper areas of the basin produced a substantial amount of lateral subsurface flow. Therefore, to incorporate that into the model, a higher conductivity value was initially assumed. Also for the convenience for the calibration, initially all 3 parameters were assumed equal for all sub watersheds delineated.



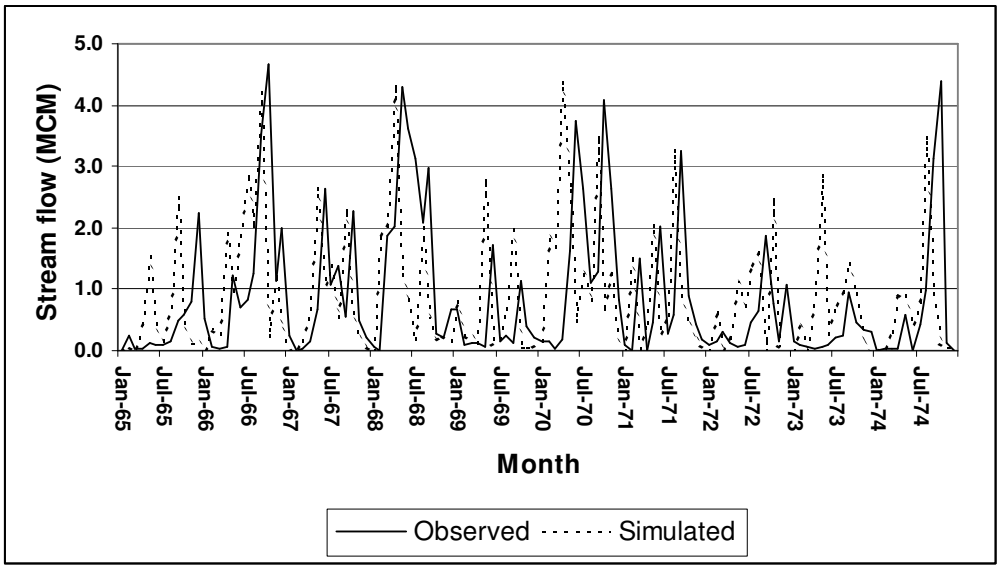
Graph 6-1:Observed and simulated flow in the Gilgil sub basin before the calibration

According to Graph 6-1, the model simulates higher values of direct runoff and base flow with respect to the observed values for most of the months. This is due to the selection of parameters. As mentioned in chapter 3, the upper areas of the subbasin are highly permeable. Therefore, these areas are not generating much surface flow. A high CN number (75) and low GW_REVAP (0.02) generate higher values of direct runoff and base flow, respectively. Also simulated data shows a faster recession of the ground water component than in observed data. In January 1970 the model simulated an exceptionally high stream flow. This was due to the over estimation of rainfall produced by the previous weather generator. According to records observed, monthly rainfall was 234 mm at station 9036241 for this month, whereas the weather generator produced 529 mm of rainfall before the repetition and adjustment procedure proposed in chapter 4 was introduced.

Manual calibration was carried out based on a trial and error method. Parameters were changed until the simulated flow showed good agreement with the observed flow. Table 6.2 shows the adjusted parameter values. Two different values of CN numbers were assigned for the upper and lower areas of the subbasin, while other parameters were assumed equal for the whole Gilgil subbasin. Adjusted rainfall data was also integrated into the model. Graph 6.2 shows the results after adjusing the parameters.

Parameter	Value	Effect on the simulation
CN2 (Upper)	65	Decrease the direct runoff
CN2 (Lower)	70	
GW_REVAP	0.15	Increase the water transfer from shallow aquifer to the root zone of a soil
SOL_K	30 (mm/hr)	Adjust soil water movements

Table 6-2:Adjusted parameters after the manual calibration



Graph 6-2:Observed and simulated flow in the Gilgil sub basin after the calibration

Correcting the exceptional rainfall value of January 1970 visually enhanced the Graph 6-2 with respect to the graph 6.1. Also the simulated stream flow follows the same patterns shown in the observed series A lower value of ground water recession coefficient gives a close agreement between observed and simulated base flows. Since the main aim is to calibrate the model for the entire Navisha basin, further calibration was performed, with the flow from the whole basin, by assigning the above values (Table 6-2) as initial parameter values.

6.4 Modelling the entire Naivasha basin

Figure 6-4:Mask created around the lake

Since Naivasha basin does not have a surface outlet and the lake is situated within the basin, all the streams drain into the lake. The SWAT delineation tool identified the lake also as a sink and filled it during the process and modelled the rivers through the lake, extending it until the lowest elevation found in the basin boundary. This is an unavoidable shortcoming in the delineation algorithm. This has been overcome by creating a mask as in figure 6-3. Mask is generally used to emphasize the approximate basin boundary to the model in situations where the spatial extent of the DEM is larger than the spatial extent of the basin. This avoids the unnecessary processing time and significantly decreases the computational time. The mask boundary was created around the lake as well so that the model-identified boundary of the lake also as a basin boundary. Since the lakebed has the lowest elevation in the basin, the model generates the stream outlets at the lake

In order to generate the SWAT input parameters, a DEM as well as land cover and the soil maps were integrated into the AVSWAT200 extension. Because of the newly introduced mask, the program de-lineated the basin boundary only in the upper part of the basin (Figure 6-4) and ignored the southern part of the basin. Since there is no significant inflow from those areas to the lake, this would not affect the simulation results.

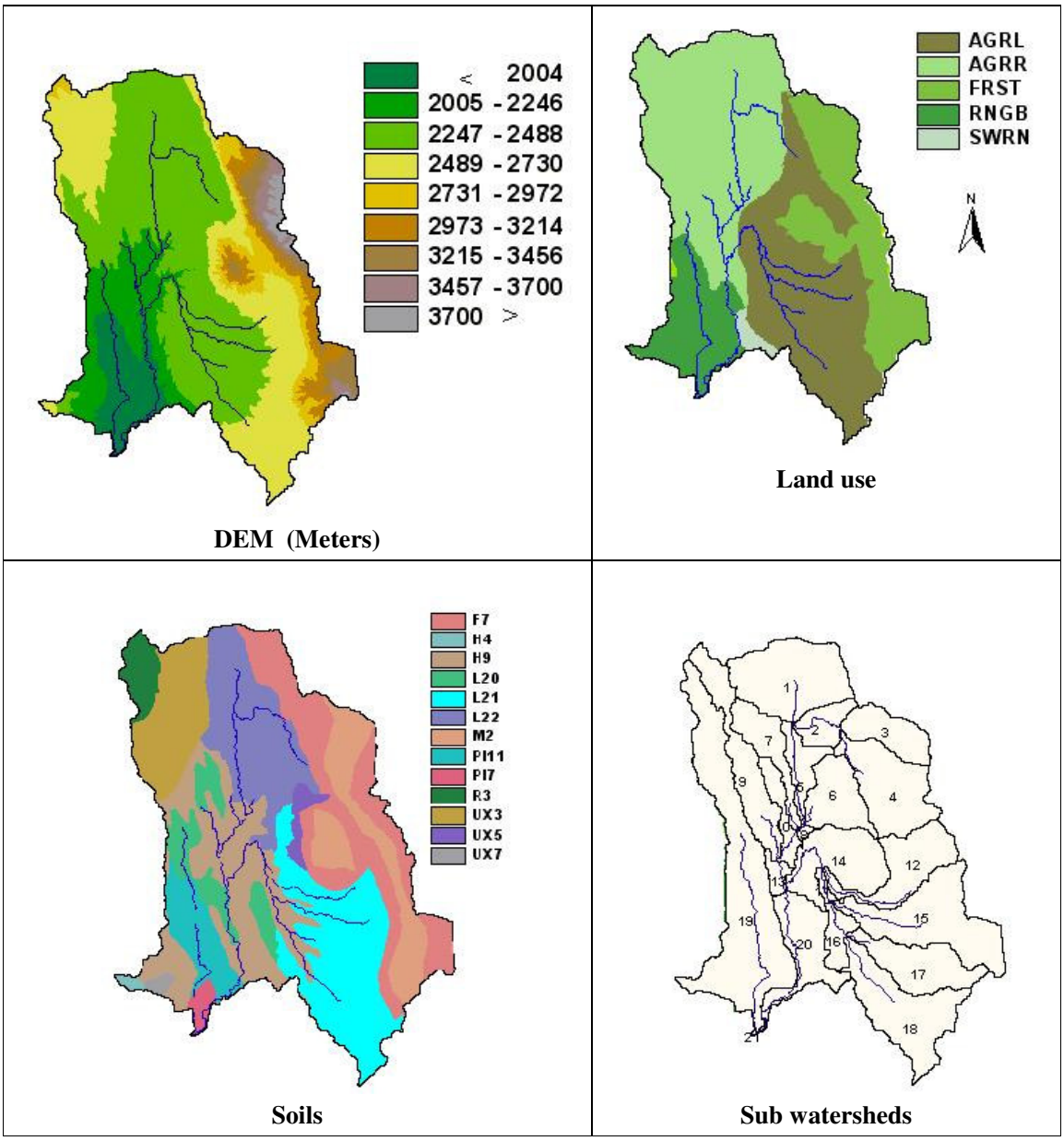


Figure 6-5: DEM, Land use, Soils and Sub watersheds in the study area

The Land use map legend shows the naming convention used by SWAT. Their physical names are given below:

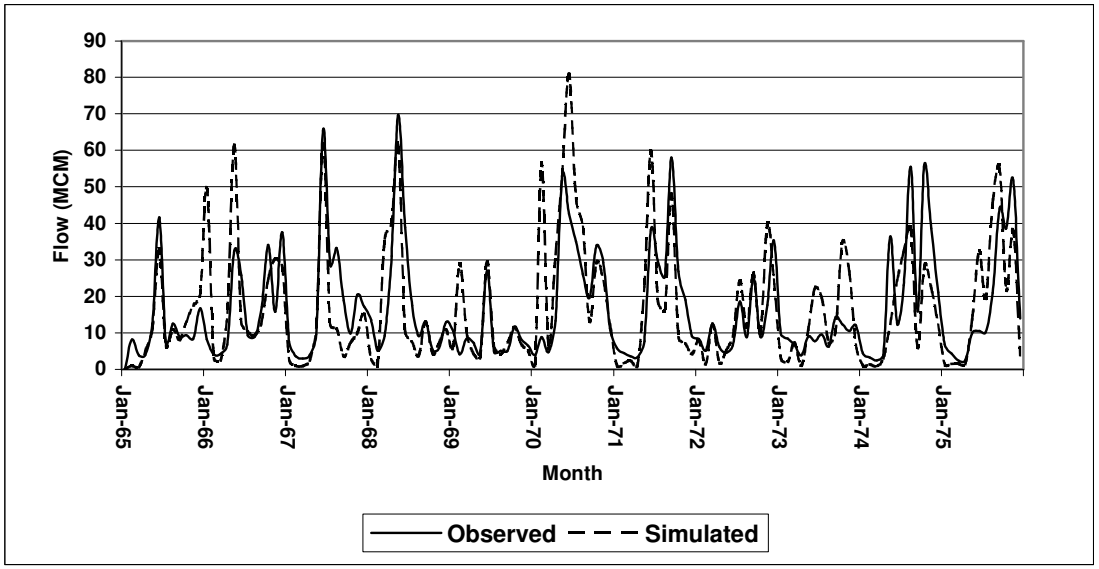
AGRL: Agricultural land (close grown) , **AGRR**: Agricultural land (raw crops),

FRST: Forest (mixed), **RNGB**: Range brush, **SWRN**: Arid range land

Soil map legend uses the Kenyan soil map naming convention. Description of these soil classes are given in Appendix 2.1

The area comprises 13 soil classes. The land use map, described in chapter 2, was classified into 5 broad categories, which are consistent with SWAT terminology. Bain was discretized into 21 sub watersheds (Figure 6-4) by assigning a flow accumulation threshold area of 7000 hectares.

Flow time series prepared for the whole basin in one of the previous studies (Mmbui, 1999) was used to compare the SWAT simulated flow. After adjusting parameters as mentioned in table 6.2, agreement between observed and simulated flow time series was close in most parts of the period. In addition to the above parameters the base flow recession coefficient (α_{gw}) was adjusted from its original value of 0.14(days) until 0.1 (days) on trial and error basis. This modification adjusted the recession parts of the stream flow graph closer to the observed values. Results of this simulation are presented in Graph 6-3.

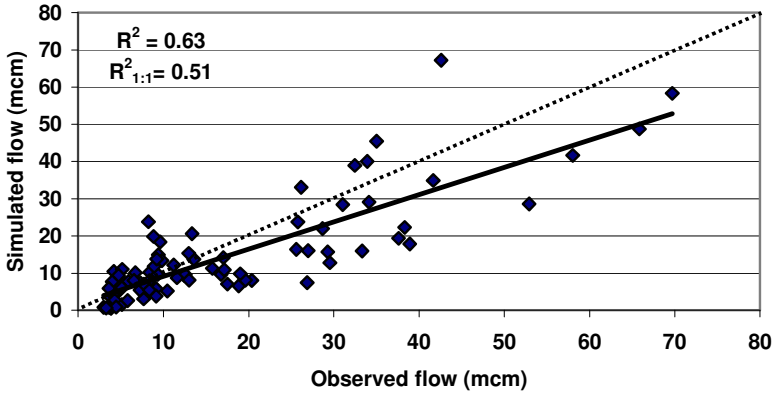


Graph 6-3:Observed and simulated flow in the total Naivasha basin after the calibration

From 1965 to 1975, simulated flow values are in close agreement with the observed values. In some months, the model simulates the shape of the observed values, but the difference between observed and simulated values is relatively higher. This could be due to errors in the observed rainfall data or stream flow data. However, Becht and Harper (2002) mentioned the unreliability of flow records after mid 1970. Therefore, that could be one reason for the deviations between observed and simulated flow data in the latter part of the calibration period. Graph 6.4 shows the scatter plot and linear regression line between the observed and simulated stream flow values for the above simulation.

Willmott (1981) mentioned the insensitivity of the R^2 value to the various additive and proportional differences that can exist between observed and modelled data and he proposed computing a $R^2_{1:1}$ which gives the association of the modelled data with the 1:1 line. Therefore, in addition the R^2 value

Nash_Sutcliffe coefficient and $R^2_{1:1}$ was also calculated. Nash_Sutcliffe coefficient was 0.51 for above simulation. R^2 $R^2_{1:1}$ was 0.63 and 0.51 respectively. This goodness of fit results agreed well with the targets set before the calibration process, which is achievement of all R^2 grater than 0.5. Therefore, the calibration was stopped at this point



Graph 6-4:Scatter plot of monthly observed versus simulated flow

After the manual calibration, the SWAT model was integrated with PEST to calibrate the stream flow automatically for the 1965 to1975 period. In addition to the parameters indicated in table 6-2, the available water content of the soil layer (AWC), ground water recession factor α_{gw} and threshold depth of water in the shallow aquifer for return flow to occur (GWQMN) was assigned to PEST as parameters to be optimised. Since SWAT delineated 21 sub watersheds, the number of parameters characterising the basin was high. It is neither possible nor meaningful to optimise all these parameter values. Therefore, the PEST control file was prepared in such a way that, it enables to formulate constraints and interdependencies for parameters. Therefore, it is possible to optimise only a few selected parameter values while others simultaneously are adjusted in previously defined ratios.

The CN value was assigned as 70 to the sub watersheds in the upper humid areas and 80 to the rest of the basins. Initial *revap* value was assigned as 0.15 for all the basins. GWQMIN was assigned as 700 and 1000 mm for different parts of the basin. Saturated hydraulic conductivity was assign as 25 mm/hr and 15 mm/hr for the upper areas and rest of the areas, respectively. Based on the results of manual calibration, initial AWC was selected as 0.07 mm/mm. Initial α_{gw} was assigned as 0.11 (days). During the optimisation, PEST minimized the initial objective function of 6892.8 to 2478. The 95% confidence limits show that the uncertainty associated with CN2 and GWQMN are low with respect to the other variables. Soil_k indicated the highest uncertainty. Also AWC indicated high correlation with revap, which was, is -0.9763. Therefore, for further optimisations, AWC was omitted from the above 6 parameters and calibration was carried out with 5 parameters. Table 6-3 gives the final estimated parameters and their 95% confidence intervals and Table 6-4 presents the correlation coefficients between parameters. Other than Soil_k, for all parameters the 95% confidence interval bounds are closer to the estimated values. This implies the low uncertainty of parameters. Moderately high correlation between ALFA and Sol_K (Table 6-4) could be one reason for relatively high uncertainty

associated with soil_k. However, the final adjusted soil_k values are in the same order of magnitude as in values found in the published literature (Dingman, 1994).

Parameter	Initial value	Adjusted value	95% confidence interval
CN2	80	72	71 .0 – 73.8
	70	65	
Revap	0.15	0.116	0.109 – 0.123
GWQMN	1000	1009	940 – 1078
α_{gw}	0.1	0.016	0.011 – 0.02
SOL_K	40	37.5	24.0 – 50.7
	25	23	

Table 6-3:Optimised parameters and there 95% confidence intervals

	CN	revap	GWQMIN	alfa	SOL_K
CN	1.0	-0.18	0.06	0.08	0.04
Revap	-0.18	1.0	-0.17	0.22	0.16
GWQMIN	0.06	-0.17	1.0	0.1219	0.16
Alfa	0.08	0.22	0.1219	1.0	0.32
SOL_K	0.04	0.16	0.16	0.32	1.0

Table 6-4:Parameter correlation coefficient matrix

	Vector 1	Vector 2	Vector 3	Vector 4	Vector 5
CN	5.5974E-04	-6.3358E-04	1.000	3.0867E-03	1.1697E-03
Revap	0.4564	-0.8898	-8.1960E-04	1.4927E-04	-6.7469E-05
GWQMIN	3.6824E-05	-6.1677E-05	-1.0631E-03	-3.4346E-02	0.9994
Alfa	-0.8898	-0.4564	2.0854E-04	1.0090E-04	8.3063E-06
SOL_K	2.1194E-05	1.7881E-04	-3.1250E-03	0.9994	3.4343E-02
Eigenvalues	3.8845E-06	8.9195E-06	0.5042	48.87	1233.0

Table 6-5:Normalized eigenvectors of parameter covariance matrix

Table 6-5shows the normalized eigenvectors of parameter covariance matrix. Domination of by one element in a eigenvector, especially for those eigenvectors whose eigenvalues are largest, indicates that the individual parameter values are well resolved by the estimation process (Doherty ,2002). In Table 6-5, vector 4 and vector 5 shows the highest eigenvalues of 48.87 and 1233.0 respectively. Also, those two vectors are dominant by one element, which indicates goodness of the optimisation.

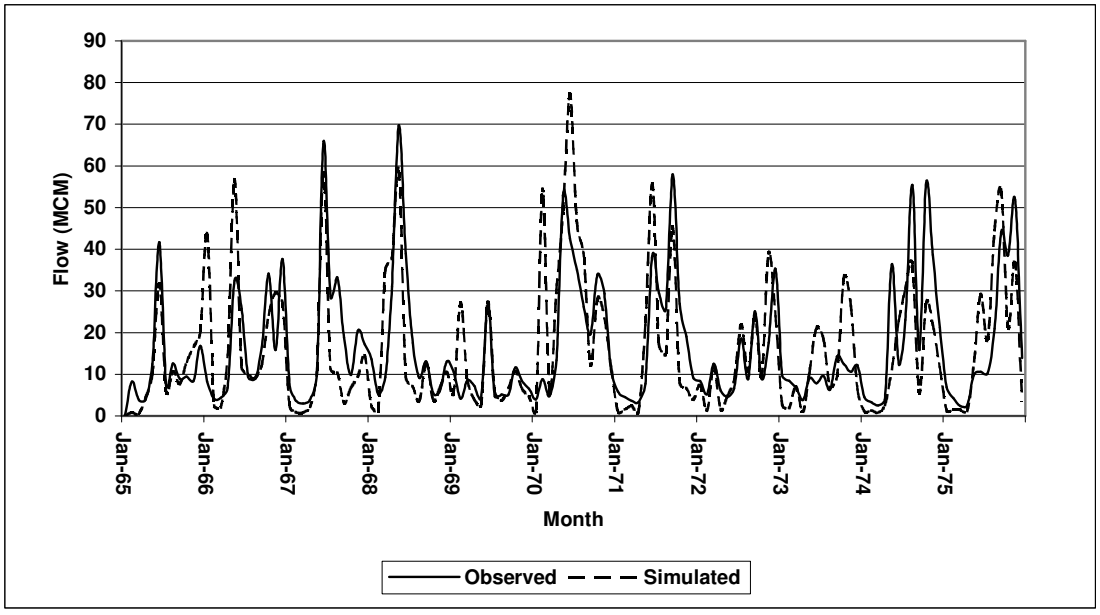
Stream flow values simulated with optimised parameters were compared with the observed values. Cumulative values between observed and simulated clearly showed that simulated values were little higher than observed values. In order to alter the simulated values, CN numbers were reduced by –

7%. Therefore, the new CN numbers for the lower and upper areas are 67 and 61 respectively. Then the simulation was carried out with new parameter values and the result of this simulation is presented in graph 6.5.

During the optimisation, PEST estimates the relative sensitivity (**RS**) of different parameters. Since the parameters are of different types and vary in different magnitudes this assists in comparing the effects that different parameters have on the parameter estimation process. Lenhart et.al.(2002) categorised the relative sensitivity into 4 classes. According to that classification, RS value between 0 - 0.05 and 0.05-0.2 categorised as negligible and medium, respectively. In higher side, the RS values between 0.2-1.0 and grater 1 classified as high and very high, respectively. The relative sensitivity values found in the parameter estimation process were categorised into those 4 classes and presented in table 6-6.

Parameter	Relative sensitivity	Category
CN	3.9	Very high
Revap	1.6	Very high
GWQMIN	1.4	Very high
Alfa	0.28	High
SOL_K	0.22	High

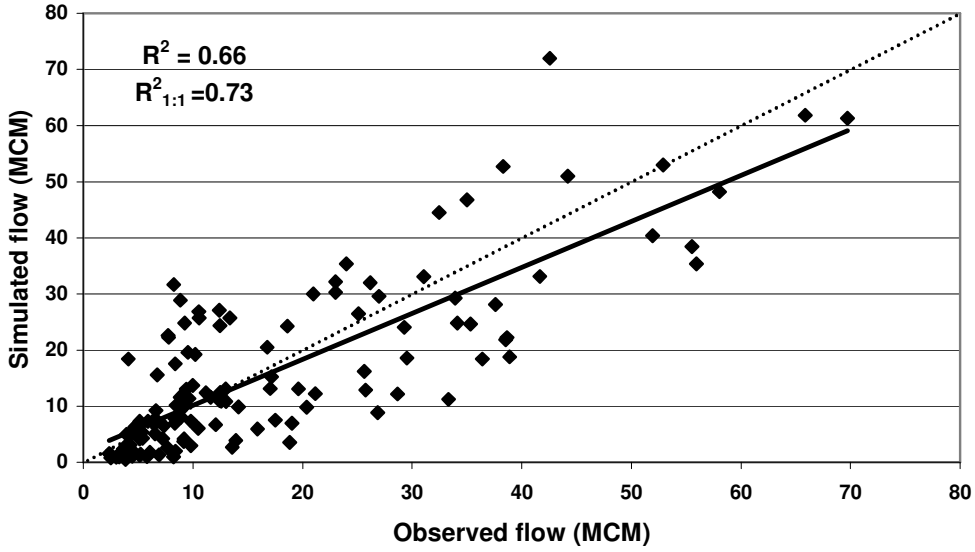
Table 6-6:Relative sensitivity values of the optimised parameters



Graph 6-5:Observed and simulated flow after the automatic calibration

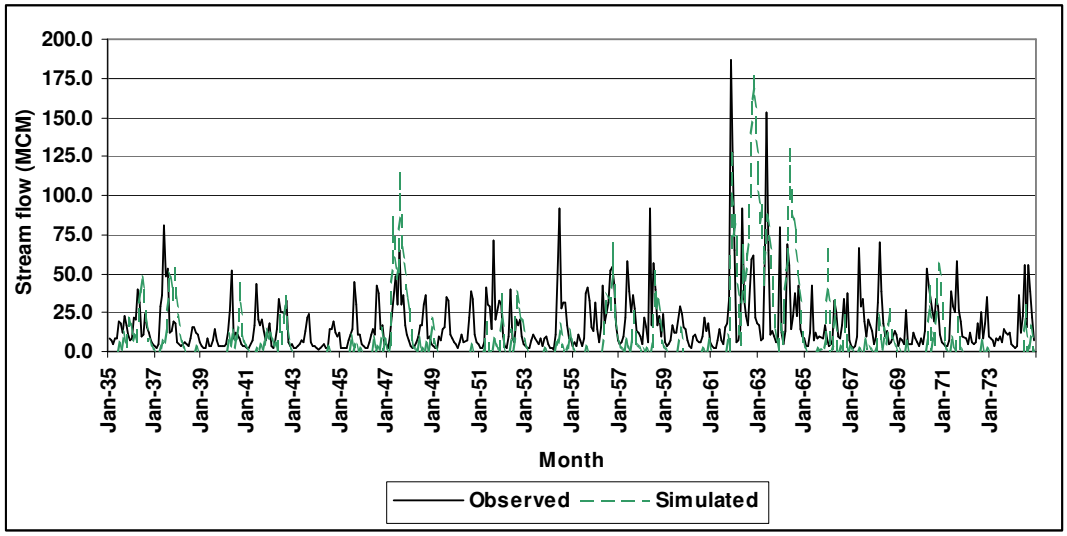
During the calibration, PEST minimized the initial objective function of 5867.0 to 2249.0
As shown in table 6-3, the confidence interval for each parameter shows the least margin of the upper and lower limits, which are closer to the optimised values. This indicates the certainty of the optimised parameters. Part of the pest run record file is attached in appendix 6-1.

The Nash_Sutcliffe coefficient was calculated and was 0.54 for the above case. This indicates that the SWAT model simulates flow data at an acceptable accuracy during the period and it gained more accuracy by doing an automatic calibration with respect to the manual calibration. The scatter plot between observed and simulated monthly values (Graphs 6.6) also shows a good agreement between observed and simulated flow values during the period. Linear regression of the simulated values against the observed values to the one to one line yielded an $R^2_{1:1}$ of 0.73.



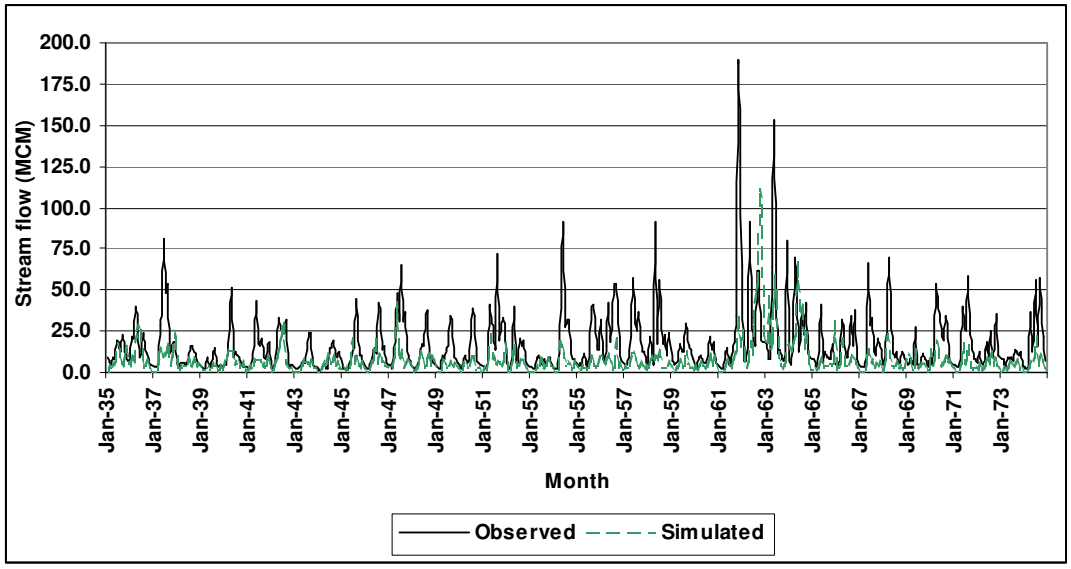
Graph 6-6:Scatter plot of monthly observed versus simulated flow after automatic calibration

Then PEST was used to further calibrate the model, by comparing simulated and observed stream flow from 1935 to 1975. Since the quality of stream flow data is unreliable after the mid 70's, the year 1975 was considered as the last year of the calibration period. Graph 6.7 shows the simulated stream flow after the calibration.



Graph 6-7:Observed and simulated flow from 1935 to 1975 after the automatic calibration

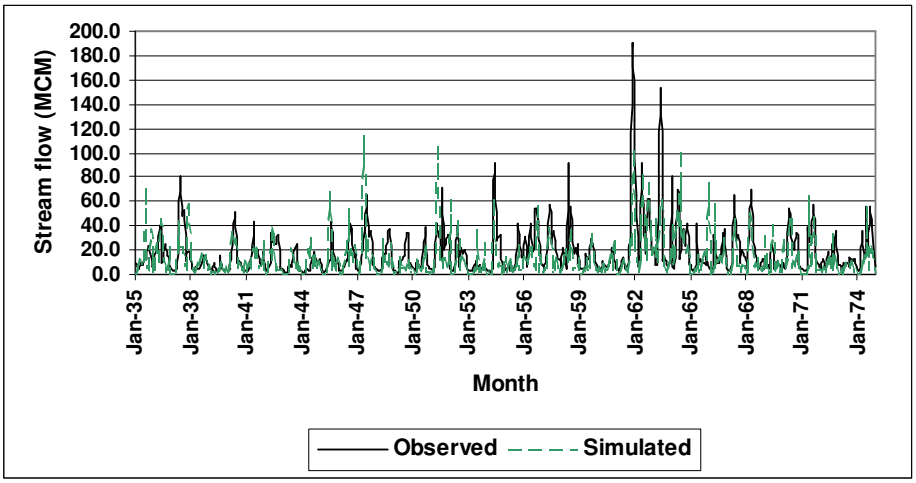
In this case, observed and simulated peak flows are in good agreement for most months. But the model simulates lower base flows with respect to the observed values. This is due to the inherent shortcoming of the algorithm used by the parameter estimation program. The optimisation algorithm used by PEST is derived from the Gauss-Marquardt-Levenberg method that has been criticized for being too easily trapped in local objective function minima (Abbaspour et.al.,2001). This method starts with searching mainly along the steepest gradient of the objective function surface. Therefore, the objective function can be converged to the local minima rather than searching for global minima. Doherty (2002) suggested to assign weights to the individual flow observation based on the inverse values of a particular observation. In this case, peak flow values get the lower weights while low flow values are assigned the higher weights. For the next simulation, weights were assigned to each observation as suggested earlier. A result of that simulation is presented in Graph 6-8.



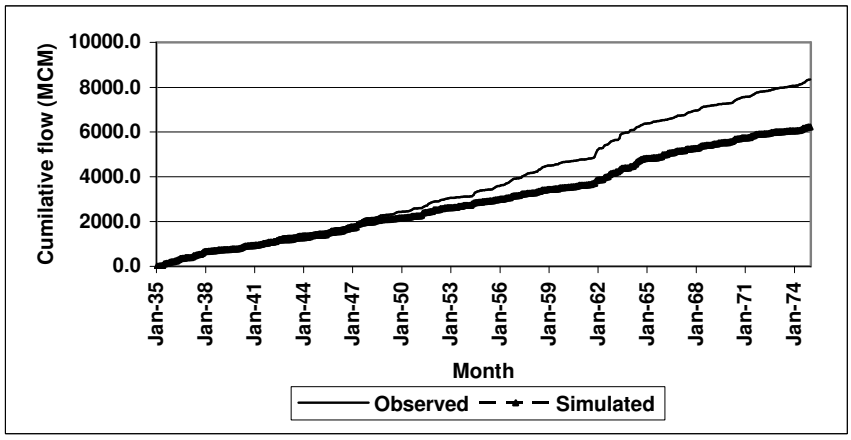
Graph 6-8:Observed and simulated flow from 1935 to 1975 after assigning weights

From Graph 6.8, it is clear that the model simulates the base flow closer to the observed values. At the same time pea, flows are not properly simulated. Due to the lower weights assigned at the peak flows, the steepest values of the objective function were shifted to the low flow values so that the PEST algorithm identifies steeper gradients of the objective function at low flow months. Therefore in contrast to the previous case, the algorithm was trapped at the local minimum values of the objective function at the months with low flow values.

In the next calibration exercise, in order to avoid under estimating at peak flow values due to the high weights assigned to low flow values, weights were only assigned for the peak flow values (inverse value of the flow). This would control the high variations of the objective function at peak flow months. The result of this simulation is presented in Graph 6-9. A cumulative observed and simulated stream flow value for this case is presented in graph 6-10.



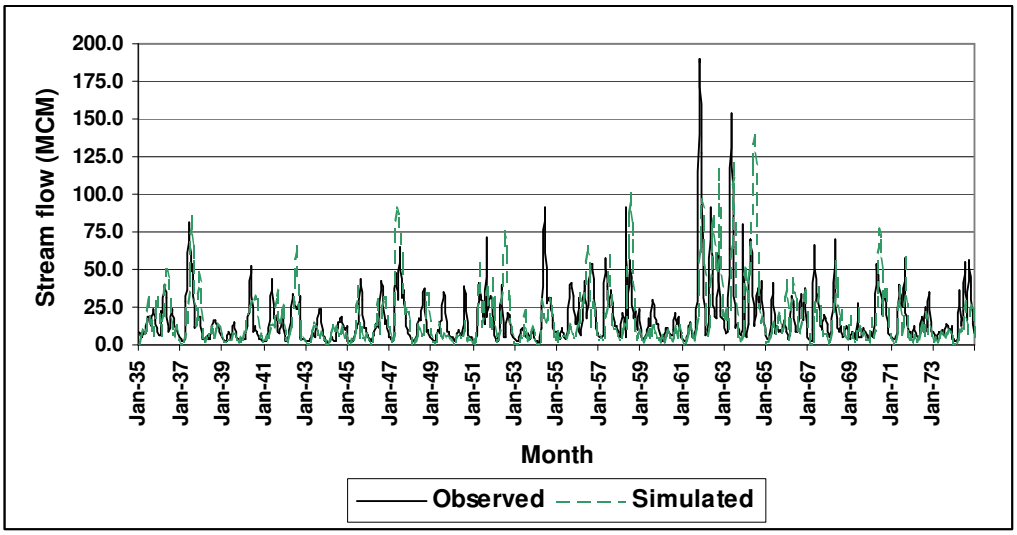
Graph 6-9:Observed and simulated flow from 1935 to 1975 after assigning weights only at peak flow months



Graph 6-10:Observed and simulated cumulative flow from 1935 to 1975 after assigning weights only at peak flow months

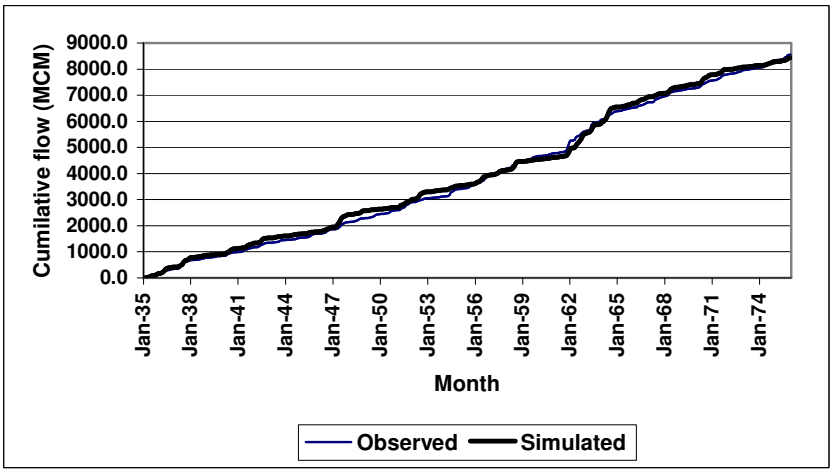
Graph 6-9 shows a better agreement between observed and simulated stream flow with respect to the other two graphs (6-7 and 6-8). Low weights at high peaks filtered out the effect of peak flows in the calibration. However, there are few peak flow months with observed and simulated stream values having significant differences. In a cumulative graph, the values of observed and simulated stream flow closely agreed with each other during the early part of the simulation period, but do not show good agreement in the latter part. This is due to the incapability of model to simulate some peak flow values.

Then the basin was simulated for the 1935-1975 period by assigning the optimum parameter values found in the first simulation (Table 6-3 and (-7%) adjusted CN numbers) where the calibration period was 10 years from 1965 – 1975. A result of this simulation is presented in Graph 6-10.



Graph 6-11:Observed and simulated flow from 1935 to 1975 using parameters found in 1965-1975 calibrations

According to Graph 6-11, apart from few a peak values, the SWAT simulation is closely agreeing with the observed stream flow series. However, it is difficult to make a visual comparison between Graphs 6-9 and 6-11 by only looking at the graphs. Therefore as in the previous case, cumulative flows were investigated and presented in Graph 6-12. This shows a better agreement of cumulative flows with respect to the previous case. On the other hand, in this simulation, period 1935-1965 can be considered as the validation period. That means that the simulation carried out, based on the parameters found during the calibration period (1965-1975), simulates rest of the period (1935-1965) closer to the observed stream flow values.

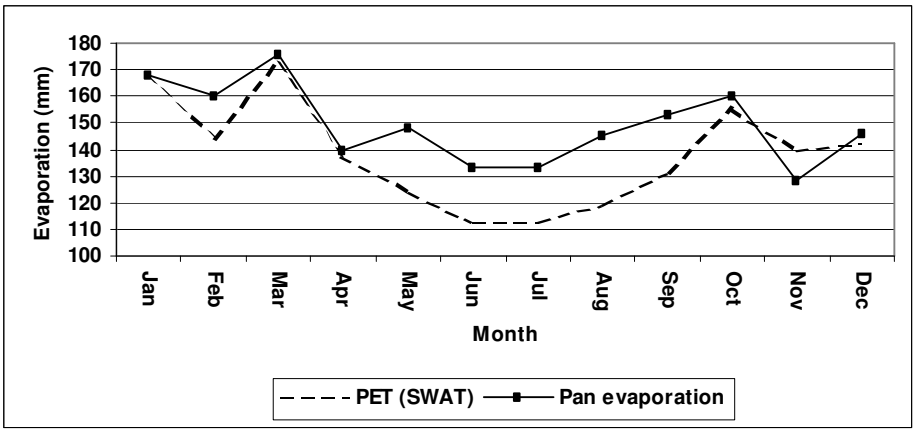


Graph 6-12:Observed and simulated cumulative flow from 1935 to 1975

Therefore, to simulate stream flow from 1935 to 1998, the same sets of parameters, which were used in the previous simulation, were utilized. Simulated stream flow was integrated with the lake water balance model to compare the observed and modelled lake water level.

Due to unavailability of weather data, Hargreaves and Samani method was used to estimate potential eavpotranspiration (PET). Monthly average values of estimated PET from 1965 – 1975 were com-

pared with the pan evapotration values in lower semi arid part of the basin. Graph 6.13 shows the results of this comparison.



Graph 6-13:Observed PAN evaporation and SWAT estimated PET in lower part of the Basin

During dry months, Hargreaves and Samani method under estimates pet with respect to pan eavporation. In wet months, the agreements between two values are higher. There is no data set to compare the PET in upper basin areas. However the estimated PET values in upper humid areas are closer to that in lower semi arid areas which contradict the findings in previous studies (Farah, 2001). Droogers and Allen (2002) stated about the same situation, which is the over estimation of PET by Hargreaves and Samani method in high humid conditions.

Simulated average daily steam flow from 1965 to 1970 was used to investigate the base flow component of the discharge in Malewa subbasin using a TIMEPLOT program mentioned in chapter 3. Since, the SWAT model used simulated rainfall data, it is difficult to compare the base flow of the simulated stream flow with the observed values presented in Graph 3-4. However, the contribution of the base flow to the total stream flow was 39%, which is closely agreed with the observed value of 44%.

6.5. Lake water balance

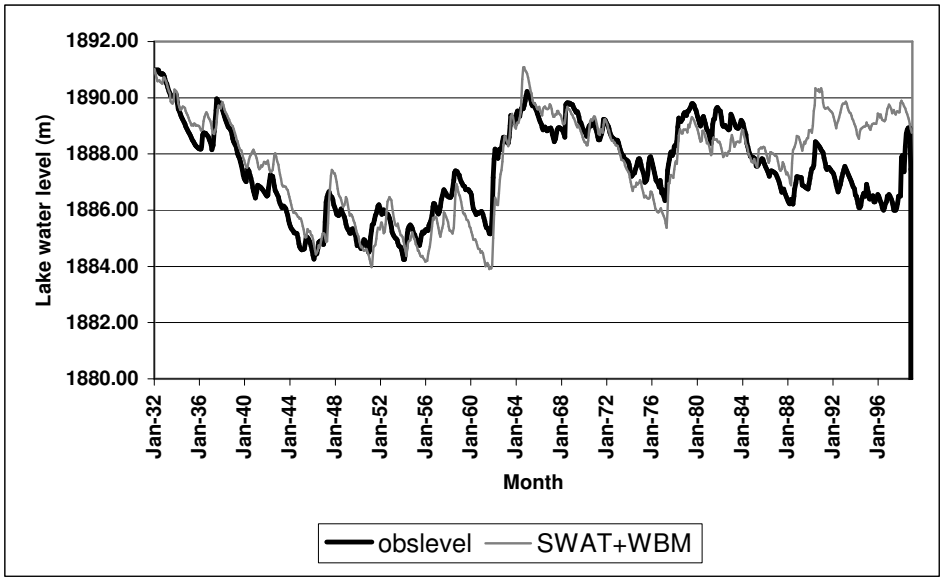
The main objective of this study is to simulate the lake water balance by using SWAT simulated stream flow. In SWAT there is no direct way to model the lake water level. Therefore, the simulated flow was integrated with a lake water balance model developed in one of the previous studies (Mmbui, 1999). This model first calculates the lake water balance based on the mass conservation equation as given in the following equation:

$$\Delta S = P - ET + R + GW_{in} - GW_{out}$$

Where: ΔS is the change in lake storage, P is the direct precipitation into the lake, ET is the lake evaporation, R is the surface runoff in to the lake, GW_{in} is the ground water flow into the lake and GW_{out} is the ground water flow from the lake.

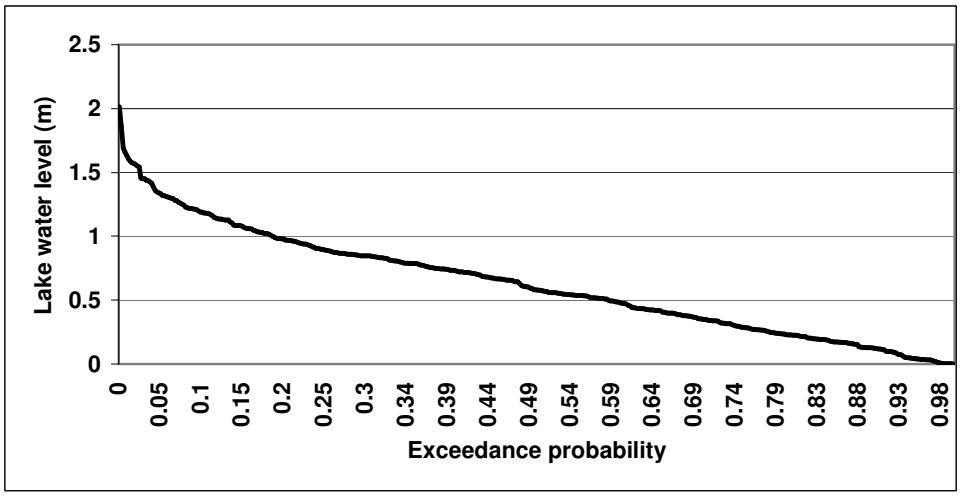
These estimated volumes are then converted into lake water levels by integrating a previously developed volume elevation curve with the water balance model. In order to validate the calibrated SWAT

model, simulated stream flow was integrated with the water balance model. Graph 6.14 presents observed and modelled lake water levels between 1935 and 1975.



Graph 6-14:Observed and simulated monthly lake water levels from 1935 to 1998

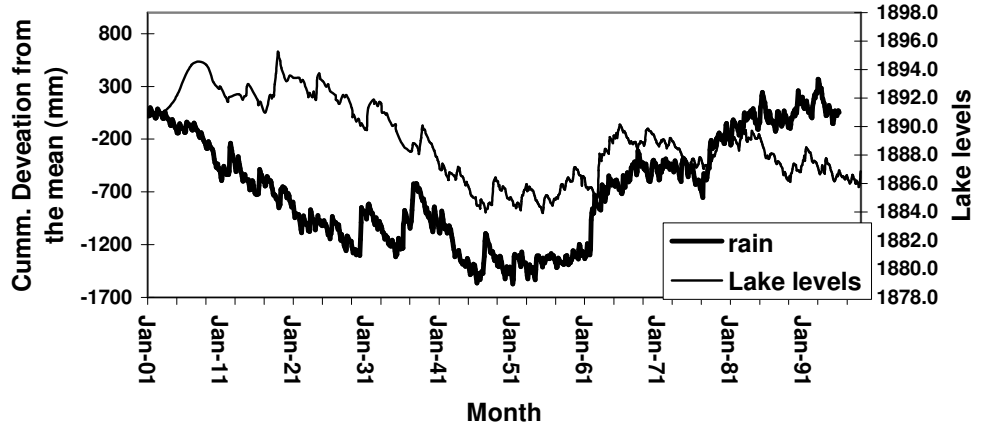
Apart from few months, modelled lake levels follow the same fluctuations as in the observed lake levels. Also there are under estimations and over estimations in some periods. Especially from 1955 to 1961, model under estimates the lake level up to two meters. However, in most months, the difference between observed and modelled lake levels is below 1 meter. Monthly differences between observed and modelled lake levels were ranked and plotted with their associated exceedance probability values for the period from 1935 – 1980 (Graph 6.15). This graph provides information on the probability of exceeding the difference of observed and modelled lake water level particular height. According to the graph, there is about 80% chance that the lake level differences are within 1 meter height.



Graph 6-15:Exceedance probability of the difference of modelled and observed lake levels

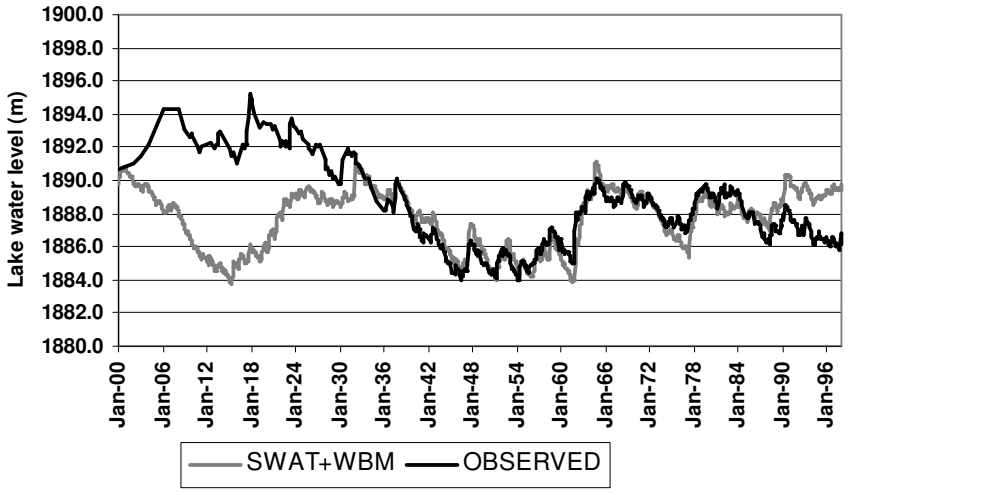
Percentage deviation of volume (*DV*) was calculated for the periods 1935 – 1965 and 1965 – 1980 separately and was 17.9% and 9.8%, respectively.

To investigate the reliability of the rainfall data, the rainfall values recorded at station number 9036002 and the lake levels were compared. The long term (1935-1998) mean value of the rainfall was subtracted from the cumulative rainfall data for each month. By doing that, extreme rainfall events were filtered out from the data set and it would enhance the trends of the time series in different time periods. This separated component was compared with the lake level time series (Graph 6-16).



Graph 6-16:Cumulative deviations of rainfall and lake levels

From 1930 to 1954, both graphs have similar fluctuation patterns. From 1955 to 1959, lake level is following an increasing trend, while it is not visible in the rainfall data. Lake level records are widely accepted as reliable (personal communications, Robert Becht). Therefore rainfall during this period is identified as unreliable.

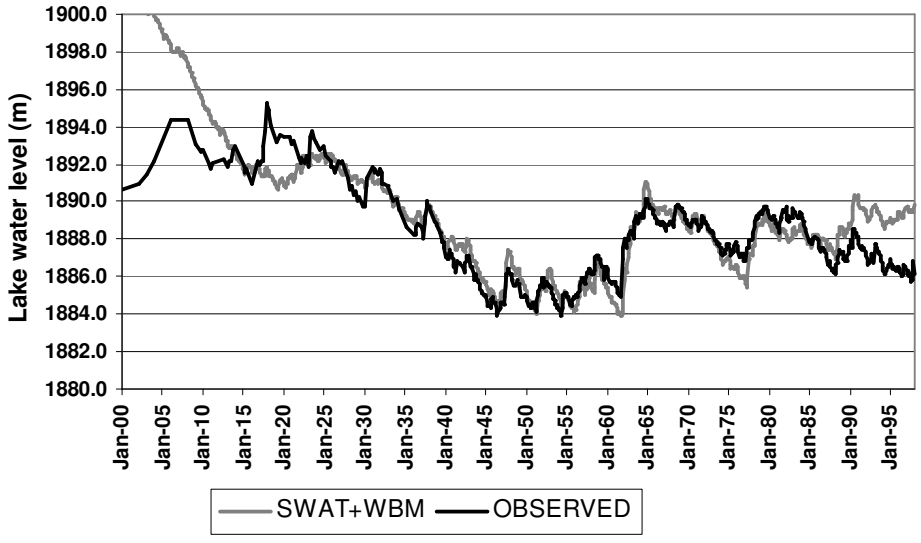


Graph 6-17: Observed and calculated lake levels after adjusting the initial lake volume

Further, the lake levels modelled by integrating the SWAT generated stream flow, from 1900, with the water balance model was compared with the observed lake levels. As mentioned in section 6.5, the

WBM model estimates the lake water level based on previously developed volume elevation curve. Initially the, starting value of the volume was adjusted in such a way that, the observed and estimated lake level are close to each other during the starting period. Comparison (Graph 6-17) shows that, during 1900 and 1932, model significantly under estimate the lake levels.

Between 1900 and 1915, modelled water level shows the decreasing trend while the observed levels are fluctuating without showing any long-term trend. After 1915, the modelled graph shows an increasing trend. But the fluctuations in observed series after 1915 are not apparent in modelled series. The lake water levels have been monitored since 1908. However, the Lake level records are widely accepted as reliable only from around 1915(personal communications, Robert Becht). Therefore, the initial lake volume was changed until the observed and modelled lake levels get closer to each other (Graph 6-18) at 1915. Due to the high initial condition, model estimates high lake water level initially. But it rapidly decreases during 1900-1915 without showing the fluctuations indicates in observed series. Even after 1915, the association between two graphs are poor until early 30's. For example, the peak water levels presented in observed series is not represented in modelled series. Further, the observed series shows an increasing trend between 1916-1918, and decreasing trend between 1918 and 1923. In contrast, the modelled series shows the decreasing trend during the first period and the increasing trend during the second period. Therefore, even after assigning higher initial lake volume value, model is still incapable of modelling the patterns found in the observed data. This clearly attributes the unreliability of rainfall data during that period.



Graph 6-18: Observed and calculated lake levels after increasing the initial lake volume

6.6. Comparison with other models

The original lake water balance model (WBM), discussed in section 6.5,used observed monthly inflow data, from 1932 – 1998, to model the lake water levels. There were two other models, developed in previous studies, modelling lake water levels based on rainfall (Wolskiy, 1999) and corrected stream flow data based on statistical relationships with rainfall and stream flow(M'mbui, 1999). In the second model, data gaps in the stream flow series were statistically generated based on the correlations between the yearly flow and yearly rainfall. Then the generated yearly stream flow was disaggregated into monthly values based on fixed coefficients assigned for each month.

Lake water levels modelled using SWAT generated stream flow was visually and quantitatively compared with the above mentioned 3 cases (Figure 6-5). A visual comparison of the result is presented in Graphs 6-19 –6-21 and the results of quantitative comparison are described and presented in table 6-6.

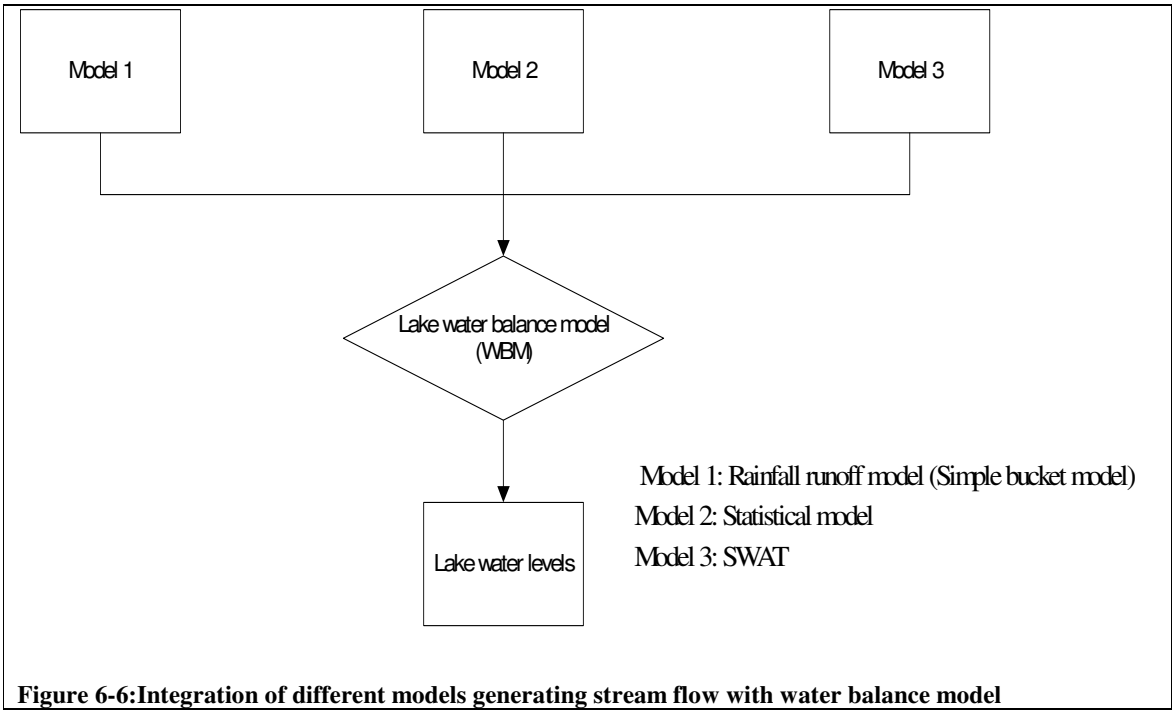
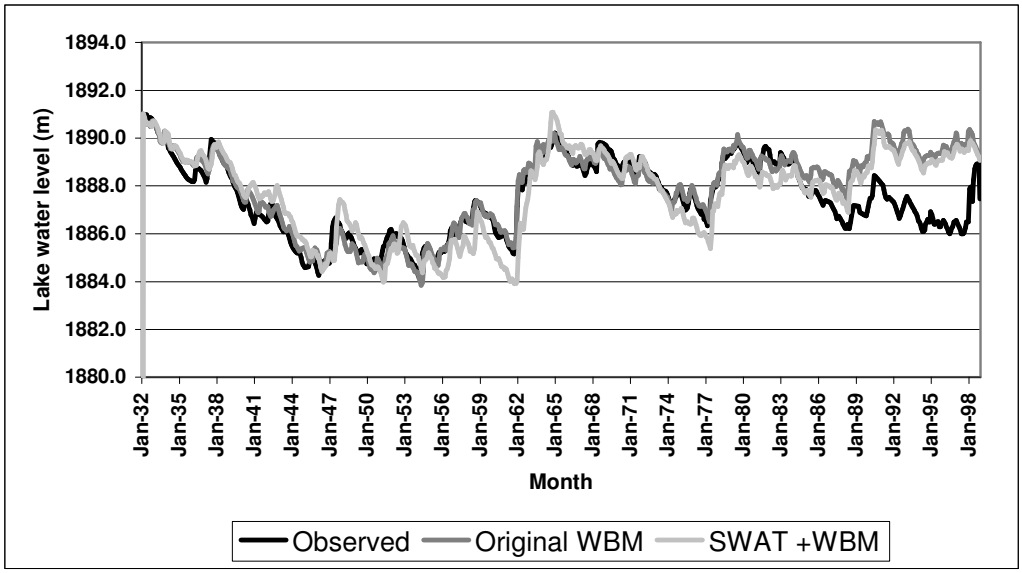
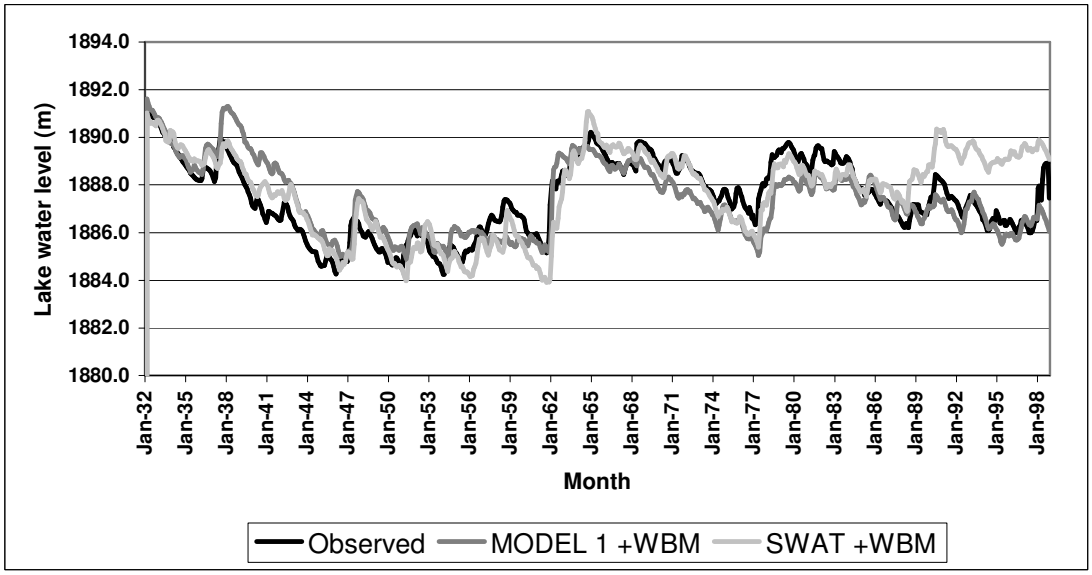


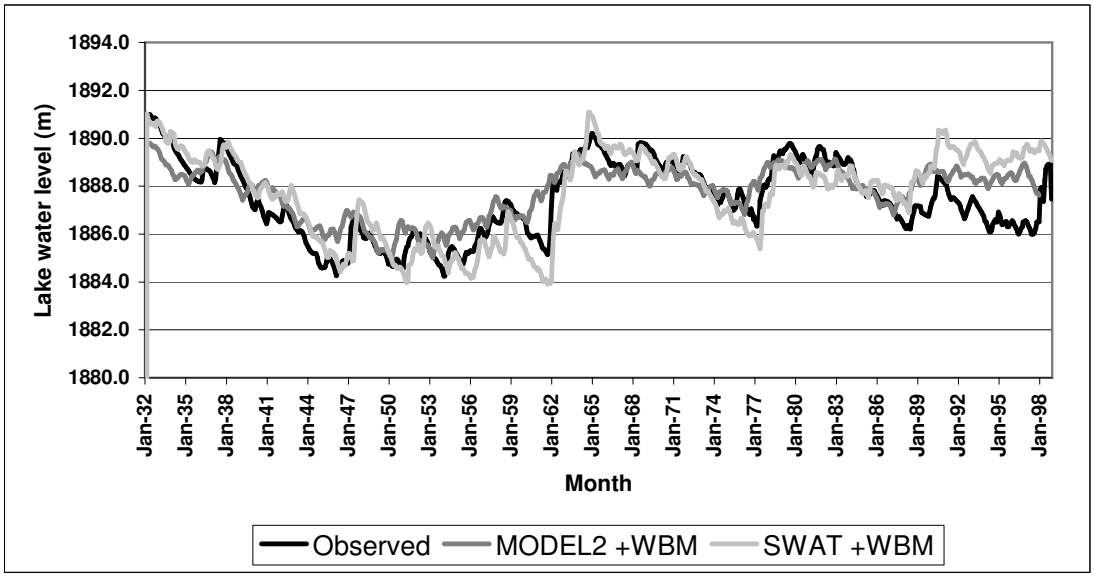
Figure 6-6:Integration of different models generating stream flow with water balance model



Graph 6-19:Observed and calculated lake levels using Original water balance model and SWAT model



Graph 6-20:Observed and calculated lake levels using rainfall a runoff model and the SWAT model



Graph 6-21:Observed and calculated lake levels using a statistical model and the SWAT model

With the observed stream flow data, the water balance model estimates the lake water level very close to the observed levels (Graph6-17). After 1984, the estimates start deviating from the observed. This divergence is attributed to the increased abstraction of water from the horticultural activities around the lake. As explained earlier, lake water level series estimated based on the stream flow simulated by the SWAT model follow the same pattern, but for some periods, estimates deviated from the observed lake levels. However, after 1984, the simulated lake levels start deviating from the observed values as expected. This is not apparent in the estimates based on the rainfall runoff model (Model1, Graph 6-18). In addition to that, lake levels estimated based on SWAT simulated stream flow, closely estimate the lake water levels with respect to Model 1. For example, lake level fluctuations between 1957-1961 was not properly estimated by Model1.

Lake water levels estimated by Model2, based on statistically generated stream flow data, also show deviations from the observed lake levels during most parts of the period. However, with respect to Model 1, it simulates the water levels after 1984, more realistically. Quantitative comparison was also carried out based on the observed and modelled lake levels from 1935-1975 and the results are presented in table 6-6.

Model	N_S	RMSE (m)	$RMSE_S$ (m)	$RMSE_U$ (m)	R^2	$R^2_{1:1}$
WBM	0.952	0.38	0.1	0.38	0.95	0.95
Model 1(Bucket)	0.662	1.02	0.36	0.96	0.69	0.65
Model 2(Statistical)	0.747	0.88	0.48	0.73	0.81	0.46
SWAT	0.831	0.72	0.1	0.72	0.86	0.86

N_S: Nash_Sutcliffe coefficient, RMSE: Root mean square error, RMSEs: Systematic component of root mean square error, RMSEu: Unsystematic component of root mean square error

Table 6-7:Statistical results generated from different models

According to the statistics presented in Table 6-6, performance of the original water balance model developed using observed stream flow data is significantly closer to the observed lake water level with respect to the other models. However, the SWAT model simulates the data also closer to the observed values with respect to other two models. As mentioned in section 6.3, for a good model, the RMSEs should be relatively small while RMSEu should approach the total RMSE. RMSE associated with WBM is 0.38 m, while SWAT is reporting a 0.78 m for the total RMSE. Both cases report low systematic components. Unsystematic error components for both WBM and SWAT are approaching the total RMSE. Therefore the errors of both models are due to the random noise of the observed data and performances of the models are in an acceptable region of accuracy.

7. Conclusions and recommendations

7.1. Conclusions

As mentioned in chapter 1, the general objective of this study was to apply a basin scale model to estimate the flow in to Lake Naivasha by considering the spatial distribution of the parameters responsible for the hydrological process and estimating the lake water level fluctuation. The study first concentrated on the preparation of the basic input parameters. In that perspective, the study was complimented by using remotely sensed information. As mentioned in chapter 2, section 2.3, a digital elevation model based on remotely sensed data certainly supplemented the study. For example, a previously used digital elevation model in the Naivasha basin (Lukman, 2003) was based on contour information and suffered the problem of not including a sufficient number of proper elevation values in the lower part of the basin. The DEM developed during the study was based on ASTER-TERRA images and solved that problem and hence enhanced the simulation results.

One major achievement of this study is the modification of a weather generator by introducing repetition and adjustment procedures. As mentioned in chapter 4, weather generators and disaggregation procedures are extremely useful in situations where weather data is not available or the temporal resolution of the available data is insufficient for a particular modelling purpose. However, the random weather generators simulate different possible realizations of weather in the particular area based on the information given. Because of the random behaviour of the weather generator, there can be over estimations and underestimations of the results (Graph 4.2). In this situation, as demonstrated in chapter 4, the introduction of repetition and adjustment procedures plays a crucial role in terms of properly desegregated rainfall. The effect of this improvement is presented in graphs 4-3 and 4-4 and clearly attributed in the simulated stream flow in graph 6-2 with respect to graph 6-1.

The parameter estimation program (PEST) is an additional tool used during this study. It is extremely helpful in overcoming the tedious and time-consuming task of manual calibration. However, as mentioned in chapter 6, careful selection of parameters and initial conditions would yield properly optimised parameters. As mentioned in chapter 6, selecting initial parameter values away from the optimised values and unrealistic boundary conditions would result in high parameter uncertainty and unrealistic optimisations. In this study, parameters values, which simulate stream flow closer to the observed values, were found by manual calibration. That enhanced the results and the optimised parameter values in automatic calibration. Assigning realistic lower and upper limits for parameters was also an important factor in the automated calibration process. It minimised the objective function with minimum parameter uncertainty. Otherwise, the model could have simulated the stream flow values close to the observed values only based on the mathematical relationships between parameters. In this way, the real physical base of the model would have been probably lost. This would certainly contradict the fundamental point of modelling (discussed in chapter 1) which should mimic an understanding of the hydrological processes prevailing inside the basin. As shown in Graphs 6-10 and 6-12, after each calibration, cumulative values between observed and simulated stream flow were compared. This was used to improve the optimised parameters further, as mentioned in chapter 6.

The Soil and Water Assessment Tool (SWAT) is a conceptual model that operates on daily time steps. The objective in model development was to predict the impact of management on water, sediment and agricultural chemical yields (Berihun, 2004) in large ungaged basins. Impacts can be assessed between baseline and alternative scenarios. This can be accomplished with little or no calibration. In this study, SWAT was integrated with the water balance model to study the lake water level fluctuations for more than 50 years. This was possible due to the ability of SWAT to simulate long periods on a continuous scale. However, calibration was necessary in this study, since it modelled the absolute values of lake water levels rather than assessing between scenarios. The Arcview extension (AvSWAT2000) of the SWAT model used the capability of GIS to generate and process spatial information required by the model. Each subwatershed is parameterised for SWAT, based on a series of hydrologic response units (HRUs), each of which corresponds to a particular combination of soil and land-cover within the subwatershed. In this perspective, the use of AvSWAT2000 gives a substantial advantage over the manual methods in terms of time. Also, it reduced the possible errors associated with manual methods and reduced the subjectivity of using manual methods. As mentioned in chapter 1, there is high spatial heterogeneity associated with the hydrological processes within a river basin. Delineating subwatersheds and HRUs derived using AvSWAT200 incorporate the heterogeneities with the modelling with respect to the lumped model. However, it depends on the spatial resolution of land use and soil information.

According to Graph 6-11, SWAT simulates the stream flow from 1935 – 1965 with an acceptable accuracy based on the parameters optimised during the calibration period 1965-1975. Therefore, the set of optimised parameters along with the other parameters can be accepted as the representative set of parameters for the whole system.

7.1.1. Limitations

One of the major limitations to large area hydrological modelling is unavailability of data to account for spatial variability associated with rainfall. For example, during this study data, from only 4 rainfall stations were used. There is no rain gauge station in the upper areas of the basin having long-term records. For each subbasin, SWAT assigns rainfall from the nearest rain gauge station. This would decrease the accuracy of simulations. The Hargreaves and Samani method estimated the PET in an acceptable order of magnitude in the wet months with respect to the long term observed pan evaporation data around the lower part of the basin. But for the dry months, it underestimates the PET up to 19% of the observed pan eavpotranspiration (Graph 6-13). However, the spatial variation was not correctly represented in the simulated data with respect to the ET map prepared by Farah (2001). Due to the unavailability of long-term data, the original weather generator inside the SWAT was used to estimate temperature. This could be the reason for poor spatial variations of simulated potential evapotranspiration values. The Hargreaves and Samani method estimates the potential evapotranspiration as a function of extraterrestrial radiation and air temperature, SWAT makes a broad assumption that roughly 20% of the extraterrestrial radiation is lost while passing through the atmosphere under cloudless skies. Using this assumption, the maximum possible solar radiation at a particular location on the earth's surface is calculated. This could be a reasonable assumption in the wet months. But in the dry months, this percentage can be lower than the assumed value. This could probably be the reason for relative under estimations of PET in the dry period.

The SWAT weather generator corrects the estimated daily solar radiation, minimum and maximum temperature by adjusting those according to the wet and dry condition of the day. As mentioned in chapter 4, the modified weather generator, introduced in this study, estimates the accurate monthly total rainfall but it is incapable of preserving the statistics such as skewness (Graph 4-4). Therefore in simulated rainfall data, wet and dry day sequences could be different from the real rain falls on the area. This would affect the generation of solar radiation and temperature, hence affecting the potential evapotranspiration. This problem of poorly generated rainfall as well as the statistical properties closer to the observed or real rainfalls on the ground could be achieved by incorporating a model such as Brtlett-Lewis which has proven capabilities of generating daily rainfall data closer to observed rainfall as well as the statistics of them.

7.2. Recommendations

Even though the modified weather generator addresses the problem of the temporal variation of rainfall data, spatial variability of the rainfall still remains a question. These circumstances urge the requirement of a spatially correlated weather generator. Development of a spatially correlated weather generator could be certainly complimented by the geo-statistical tools such as Kriging as it has developed many applications in the field of spatial estimation of rainfall. In case of the spatial variability of rainfall, one other possibility is to use remotely sensed data such as METEOSAT, as it observed all the precipitation clouds within its field of view in 15 minutes cycles.

The ground water component of SWAT is one-dimensional and does not address the flow between subbasins. In that sense it is a lumped system and thus is not detailed enough to handle distributed parameters and variable pumping. Hence, ground-water levels cannot be simulated accurately. As mentioned by Sophocleous (1999) this could be overcome by integrating SWAT with a sophisticated groundwater simulator such as MODFLOW.

Broad land use information derived from remotely sensed data does not seem to have significantly affected the simulations in this study. However, in addition to hydrology, it is possible to model the agricultural chemical yields in stream flow using SWAT. In such situations, it is important to have a properly classified land use map showing the extent of agricultural lands, which is a main source of non point source pollution.

As mentioned in chapter 1, the development of water resources plans would certainly compliment the quantitative description of water resources in basins. For example, the stream flow series simulated with the SWAT model can be used as an input to the model such as Water Evaluation and Planning System (WEAP), which consider both the demand and supply side planning in an integrated manner.

Becht and Harper (2002) clearly mentioned about the need of a well-calibrated hydrological model, which is capable of evaluating hypothetical scenarios. Mmbui(1999) also mentioned about the unavailability of proper stream flow data and indicated the need of a well calibrated rainfall runoff model for the Naivasha basin. The results of this study strongly indicate the capability of the SWAT model to fulfil that requirement.

Even though this study was concerned about the hydrological component of the SWAT model, as mentioned in chapter 4, the SWAT model consists of components for nutrients, pesticides and agricultural management. A model in which all components are well calibrated would be a useful tool for future lake management activities.

References

Abbaspour,A., Matta,V.,Huggenberger.P.,Jonson,C.A. ,2000. A contaminated site investigation: comparison of information gained from geophysical measurements and hydrogeological modeling, *Journal of Contaminant Hydrology, (40) , Pages 365-380*

Alemaw, S.B.F. and Chaoka,T.R. , 2003. A continental scale water balance model: a GIS-approach for Southern Africa, *Physics and Chemistry of the Earth, Parts A/B/C, Volume 28, Issues 20-27, Pages,957-966.*

Al-Sabbagh, M.,2001. Surface runoff modeling using GIS and remote sensing. MSc. Thesis, International Institute for Aerospace Survey and Earth Science, Enschede, The Netherlands.

Al-Weshah, A., 2002. The role of UNESCO in sustainable water resources management in the Arab World. *Desalination*.152, 1-l 3.

Arnold,J.G.,Sirinivasan,R.,Muttiah,R.S, Williams,J.R. 1998. Large area hydrologic modelling and assessment, Part 1: Model development. *Journal of the American water resources association*, 34(1).

Band,L.E and Moore,I.D., 1995,Landscape attributes and Geographic Information Systems. *Hydrological processes*, 9(4).

Bastiaanssen,W., 2000, Shared water resources information from space. Inaugural address, International Institute for Aerospace Survey and Earth Science, Enschede, The Netherlands.

Becht, R., Harper, D., 2002. Towards an understanding of human impact upon the hydrology of lake Naivasha. *Hydrobiologia*, 488:1-11.

Berihun,A.T.,2004. modelling water quality using Soil and Water Assessment Tool , Naivasha basin ,Kenya. MSc. Thesis, International Institute for Aerospace Survey and Earth Science, Enschede, The Netherlands.

Beven, K.J., 1999, Rainfall-runoff modeling. John Wiley & sons, Ltd.

Bronstert A. and, Bardossy.A. , (2003),Uncertainty of runoff modelling at the hill slope scale due to temporal variations of rainfall intensity. *Physics and Chemistry of the Earth 28: 283–288.*

Chow, V.T., Madiment, D.R., Mays, L.W., 1988. Applied hydrology. McGraw hill.

Dapp,J.T., Vreugdenhil, H.S.I.,2004. Contribution to the Management of Lake Naivasha, by focussing on the Hydrology of the Gilgil River. Bsc. Report,Technical university of Delft.

Dingman, S.L., 1994. Physical Hydrology.Prentice Hall, New Jersey.

Doherty, J., 2002. PEST user manual. Watermark numerical computing, Brisbane, Australia.

Doherty, J., Johnston,J.M., 2003. Methodologies for calibration and predictive analysis of a watershed model. *Journal of the American water resources association* (JAWRA) 39(2): 251-265.

Droogers, P., Allen, R.G. ,2002. Estimating reference Evapotranspiration under inaccurate data conditions. *Irrigation and Drainage System,* 16(1), 33-45.

Eckhart,K.,Arnold, J.G.,2001.automatic calibration of distributed catchment model. *Journal of Hydrology*, 251 103-109.

Farah, H.O., 2001.Estimation of regional evaporation under different weather conditions from satellite and meteorological data: A case study in the Naivasha basin, Kenya. PhD Thesis, Wageningen University, The Netherland.

Fontaine,T.A., Cruickshank,T.S., Arnold J.G., Hotchkiss, R.H,2002. Development of snowfall snowmelt routine for mountainous terrain for the soil and water assessment tool (SWAT). *Journal of hydrology*, 262:209-223.

Franeos, A. , Bidoglio,G. , Galbiati,L. , Bouraoui,F. , Elorza,F.J. , Rekolainen,S. , Manni,K. , Granlund ,K., 2001.Hydrological and water quality modeling in a medium-sized coastal basin. *Physics and Chemistry of the Earth (B*), 26(1): 47-52.

Hargreaves, G.H, 1975, Moisture availability and crop production, *Transaction of the American society of Agricultural engineering*, 18.

Hayhoe, H.N., Stewart,D.W., 1996.Evaluation of CLIGEN and WXGWN weather data generator under Canadian conditions. *Canadian Water Resources Journal*, 21(1): 53-67.

Hayhoe, H.N., 1998. Relationship between weather variables in observed and WXGEN generated data series. *Agricultural and Forest Meteorology* 90,203-214.

Hirano,A., Welch,R., Lang,H., 2003. Mapping from ASTER stereo image data: DEM validation and accuracy assessment. *ISPRS Journal of Photogrammetry & Remote Sensing* 57: 356– 370.

Hutchinson, M. F., Gallant,J.C. (1999). *Representation of terrain*. M.F.: Longley

Jain,S.K.B., Storm,J.C.,Bathurst,J.C.,Refsgaard,J.C.,Sing,R.D.,1992.Application of the SHE to catchmants in India. Part 2. Field experiments and simulation studies with the SHE on the Kolar subcatchment of the Narmada river. *Journal of hydrology* 140:25-47.

Kilgore, J.L., 1997. Development and evaluation of a GIS based spatially distributed unit hydrograph model. MSc. Thesis**,** Virginia Polytechnic Institute and State University

Kite, G and Droogers, P., 2002.Integrated basin modelling, Research report (43). International Water Management Institute, Colombo, Sri Lanka.

Koutsoyiannis, D., Onof, C., 2001. Rainfall disaggregation using adjusting procedures on a Poisson cluster model. *Journal of Hydrology,* 246, 109-122.

Lake Naivasha riparian owners association, 1996. Lake Naivash management plan. P.O.Box 1011, Nivasha, Kenya

Lenhart,T., Eckhardt,K., Fohrer,N.,Freed, H.H., 2002.Comparison of two different approaches of sensitivity analysis. *Physics and Chemistry of the Earth* 27 : 645–654

Lukman, A.P., 2003, Regional impact of climate change and variability of water resources (Case study, lake Naivasha basin , Kenya). MSc. Thesis, International Institute for Aerospace Survey and Earth Science, Enschede, The Netherlands.

Maidment ,D.R,1992. Handbook of hydrology. McGraw-Hill, Inc.
McCann, D.L., 1974. Hydrological investigation of Rift valley catchments

McCaskill,M.R., 1990. TAMSIM – a program for preparing meteorological records for weather driven models. Tropical Agronomy Technical Memorandum N0.65, CSIRO, Div.of Tropical Crops and Pasture, Brisbane, 26 pp.

McDonnell,R.A., 1996. Including the spatial dimension: Using geographic information systems in hydrology. *Physical geography*, 20(2), 159-177.

McKinney,D.C., Cai,X., Rosegrant, M.W., Ringler,C., Scott,C.A., 1990.Modeling water resources management at the basin level: Review and future directions. SWIM Paper 6, International Water Management Institute, Colombo.

Meijerink A.M.J., de Brouwer, H.A.M., Mannaerts,C.M.,Valenzuela,C.R., 1994. Introduction to the use of Geographic Information Systems for practical hydrology. Plublication number 23, International institute for Aerospace Survey and Earth Sciences, Enschede, The Netherland.

Mmbui, S.G., 1999, Study of long-term water balance of lake Naivasha, Kenya. MSc. Thesis, International Institute for Aerospace Survey and Earth Science, Enschede, The Netherlands.

Nathan,R.J.,McMahon,T.A.,1990. Evaluation of automatic techniques for base flow and recession analysis. *Water resource research*. 26(7): 1465-1473.

Neitsche,S.L.,Arnold, J.G.,Kini,J.R.,Wlliams, J.R.,King,K.W., 2002. Soil and Water Assessment Tool, Theoretical documentation. Grassland, Soil & Water Research Laboratory, Temple, Texas GSWRL Report 02-01.

Nicks, A.D. 1974. Stochastic generation of the occurrence, pattern, and location of maximum amount of daily rainfall. p. 154-171. *In* Proc. Symp.Statistical Hydrology, Aug.-Sept. 1971, Tuscon, AZ. U.S. Department of Agriculture, Misc. Publ. No. 1275.

Nilsson, E., 1932. Quarternary Glaciations and pluvial lakes in British East Africa. PhD thesis, Central tryckeriet Stockholm: 101 pp.

Peucker,T.K., 1978. Data structures for digital terrain models: discussion and comparison. I[st] international study symposium on topographical data structures for Geographical Information Systems. Harvard paper on GIS, (Eds. Dutton,D.)

Podder, Md.A.H.,1998, Estimation of long-term inflow into lake naivasha from the malaewa catchment, Kenya. MSc. Thesis, International Institute for Aerospace Survey and Earth Science, Enschede, The Netherlands.

Richardson, C.W., Wright,D.A., 1984. WGEN: A model for generating daily weather variables. U.S.Dep.of Agric.,Agric.Res. Service, ARS-8, 83pp.

Sharmo, U.G, 2002, 1999. Hydrological investigation for water harvesting potential using GIS, RS and runoff models. MSc. Thesis, International Institute for Aerospace Survey and Earth Science, Enschede, The Netherlands

Sharpley, A.N. and J.R. Williams, eds. 1990. EPIC-Erosion Productivity Impact Calculator, 1. model documentation. U.S. Department of Agriculture,Agricultural Research Service, Tech. Bull. 1768.Williams, J.R. 1995. Chapter 25. The EPIC Model. p. 909-1000

Sirinivasan,R., Ramanarayanan,T.S.,Arnold, J.G. and Bednarz,S.T. 1998. Large area hydrologic modelling and assessment, Part 2: Model application. *Journal of the American water resources association*, 34(1).

Sombroek,W.G.,Braun,H.M.H., and Van der Pouw,B.J.A, 1980, exploratory soil map and agroclimatic zones map of Kenya. Exploratory soil survey reports No. E1, Kenya soil survey, Nairobi.

Soltani, A., 2000.Evaluation of WEGEN for generating long term weather data for crop simulation. *Agriculture and Forest Meteorology* 102, 1-12.

Sophocleous, M.A., Koelliker, J.K., Govindaraju,R.S.,Birdie,T.,Ramireddygari,S.R.,Perkins,S.P., 1999. Integrated numerical modeling for basin-wide water management: The case of the Rattlesnake Creek basin in south-central Kansas. *Journal of Hydrology* 214: 179–196.

Strobl, R.O., 2002. Water quality monitoring network design methodology for the identification of critical sampling points. Ph.D. Dissertation. Dept. of Agricultural and Biological Engineering. The Pennsylvania State University.

Syeed, A., 2001, Economy versus environment: How a system with RS and GIS can assist in decisions for water resources management. MSc. Thesis, International Institute for Aerospace Survey and Earth Science, Enschede, The Netherlands.

Troch P.A., Paniconi,C., McLaughlin,D.,2003. Catchment-scale hydrological modeling and data assimilation, *Preface / Advances in Water Resources* 26 :131–135.

Wallis,T.W.R., Griffiths, J.F.,1995. An assessment of the weather generator (WXGEN) used in the erosion/productivity impact calculator. *Agriculture and Forest Meteorology,* 73:15-133.

Willmott, C.J, 1984. On the evaluation of model performance in physical Geography. Spatial statistics and models, D.Reidel publishing company,(Eds. Gaile, G.L and Willmott,C.J.).

Wolski, P., 1999. Application of reservoir modelling to hydrotopes identified by remote sensing. Phd Thesis, International Institute for Aerospace Survey and Earth Science, Enschede, The Netherlands.

Wong, K. M., 2000.Disaggregation of rainfall time series using adjustments, Diploma thesis, Department of Civil and Environmental Engineering - Imperial College, London, London.

Zalewski,M.,2000. Ecohydrology — the scientific background to use ecosystem properties as management tools toward sustainability of water resources. *Ecological engineering* 16 :1–8.

Zinck,J.A, 1986, Physiography and soils, lecture notes. International institute for Aerospace Survey and Earth Sciences, Enschede, The Netherland.

Appendix 2.1 : Coordinates collected during the field work on different land use classes

Point ID	X (meters)	Y (meters)	Elevation (m)	Land use
1	216184	9918191	2014	kws
2	219575	9923685	2115	Scrub land
3	219940	9929505	2327	Scrub land + trees
4	220487	9930511	2430	Small farms
5	223688	9938334	2424	Small farms
6	227659	9938930	2434	Hospital
7	235843	9945974	2610	Small farms
8	239554	9942302	3157	Benchmark
9	220577	9930489	2421	Dry grass
10	220189	9930504	2407	Dry grass
11	223273	9930655	2409	Grass
12	222973	9930172	2423	Wheat
13	222953	9930075	2422	Wheat
14	222916	9929821	2432	Maize
15	223053	9929847	2428	Grass
16	225301	9931235	2412	Grass
17	229369	9937438	2436	Grass
18	203230	9941069	1979	Small trees
19	203392	9942294	1975	Grass
20	205334	9949084	2163	Scrub
21	206417	9950515	2150	Maize
22	205641	9954619	2242	
23	207060	9959533	2335	Wheat
24	207593	9965749	2252	Scrub
25	205232	9971472	2483	Grass
26	198862	9971722	2523	Grass + trees
27	207197	9960144	2276	Grass
28	214613	9972619	2358	Grass
29	211198	9949017	2198	Scrub
30	220344	9928926	2297	Trees
31	221757	9930474	2243	Grass
32	223017	9930753		Grass
33	223956	9930842	2438	Grass
34	224390	9931050	2465	Grass
35	225215	9931147	2449	Grass
36	225284	9931283	2444	Grass
37	226733	9931548	2478	Grass
38	227099	9931090	2483	Grass
39	227661	9930673	2484	Grass
40	228021	9932254	2462	Grass
41	229369	9933167	2472	Grass

42	229499	9940814	2455	Grass
43	229570	9940744	2460	Small farm
44	230027	9941134	2490	Grass
45	233435	9944147	2563	Trees
46	233439	9951107	2741	Needle trees
47	232876	9951563	2774	Small town
48	231594	9953334	2816	Mixed
49	231532	9953563	2825	Mixed
50	231126	9954237	2834	Grass
51	221152	9957799	2434	Grass
52	221236	9956838	2467	Grass
53	213210	9914445		Grass
54	212097	9914675	1894	Grass
55	211381	9914324	1902	Mixed
56	210290	9910149	1917	Mixed
57	206267	9908603	1900	Grass
58	206197	9908675	1898	Grass/bushed
59	206230	9908735	1897	Grass/bushed
60	195944	9910657		Home garden
61	214080	9921226	1916	Urban
62	205204	9921283	1902	Grass
63	205349	9920937	1901	Grass
64	205282	9920717	1901	Grass
65	203077	9941018		Small trees
66	203572	9942317		Dry grass
67	205517	9949103		Maize
68	206261	9950578		maiz+trees
69	206228	9950417		Mixed
70	206479	9931839		Grass
71	205918	9931260		Dry grass
72	205925	9931263		Mixed
73	204377	9945718		Bare land
74	196504	9982321		Bush land
75	193523	9981234		Bush land
76	194171	9981051		Small farms
77	194428	9981254		Bush land
78	194809	9975121		swamp
79	198840	9971755		Bush land
80	206536	9971820		Grass
81	203607	9953206		Bush land
82	209149	9929287	1903	Trees
83	209072	9931434	1913	Brushes/dry grass
84	207960	9933738	1950	Dry grass/trees
85	207600	9935078	1942	georeference/dry grass
86	208024	9935366	1942	Bushes/dry grass

87	209162	9933111	1956	Dry grass
88	208945	9935700	1937	Grass/busches
89	204902	9938808	1958	Trees
90	205016	9941877	1975	Maize
91	205529	9949506	2175	Maize
92	205365	9950182	2188	Maize
93	204469	9953329	2280	Bushes/high grass
94	203561	9957130	2442	High grass
95	203613	9958296	2453	Maize
96	201531	9962725	2444	Grass/bushes
97	200791	9957921	2363	high grass
98	200836	9956646	2313	Bush land
99	201864	9953259	2257	Bushes
100	217950	9911673	2044	Dry grass
101	218213	9911647	2069	Dry grass
102	219243	9911708	2125	Dry grass
103	223179	9911794	2248	Small farm
104	223750	9912230	2288	Scrub
105	225704	9911499	2411	Small farm
106	226571	9911479	2464	Trees
107	226272	9911533	2438	Maize
108	224718	9921304	2487	Trees
109	223644	9923563	2476	Small farm
110	223259	9924261	2467	Grass
111	218412	9918994	2077	Maize
112	218699	9918438	2082	Grass
113	218069	9917254	2065	Grass
114	206927	9934767	1956	Grass
115	207884	9932592	1946	Grass
116	205722	9931066	1931	Scrub
117	205212	9930501	1906	Scrub
118	204724	9929656	1931	Acacia
119	202498	9927926	2030	Scrub
120	197459	9919794	1932	Scrub
121	194158	9912602	1940	Grass
122	194106	9912624	1940	Scrub
123	195073	9910075	1932	Acacia
124	215077	9917769	1945	Grass
125	204533	9945730	1990	scrub
126	206919	9957802	2247	Wheat
127	207549	9965280	2312	Wheat
128	207645	9966203	2309	Wheat
129	197852	9971196	2522	Eucalyptus
130	197948	9970580	2539	Wheat

Appendix 2.2 : Soil properties in Naivasha basin

H Hills and minor scarps

H4: Soil developed on ashes and other pyroclastic rocks of recent volcanoes; somewhat excessively drained, shallow, dark brown-to-brown, friable and slightly smeary, rocky and stony, clay loam

H9: Soil developed on undifferentiated tertiary volcanic rocks; well-drained, shallow, dark reddish brown, friable, very calcareaous, bouldery or stony, loam to clay loam, in many places saline

L Plateaus and high-level structural plains

L20: soil developed on ashes and other pyroclastic rocks of recent volcanoes; well drained, deep to very deep, dark brown, friable and slightly smeary, clay loam to clay

L21: imperfectly drained, deep very dark greyish brown, mottled, firm clay,abruptly underlying a thick topsoil of friable silty clay loam (solodic PLANOSOLS).

L22: well drained, deep to very deep, very dark greyish brown to dark brown,friable and slightly smeary, clay loam (ando –luvic PHAEOZEMS)

M Mountains and major scarps

Somewhat excessively drained, shallow to moderately deep, brown to dark brown, firm and slightly smeary, strongly calcareous, stony to gravely clay loam: in many placessaline and/or sodic and with inclusions of lava fields (ando-calcaric REGOSOLS,partly lithic phase)

Pl Lacustrine plains

Pl 7: soil developed on sediments from volcnic ashes and other sources; imperfectly drained to poorly drained, ver deep, dark greyish brown to dark brown, firm to very firm, slightly to moderately calcareous, slightly to moderately saline, moderately to strongly sodic, silt loam to clay; in many places with a humic topsoil; subrecent lake adges of the Central Rift Valley

Pl11: soil developed on sediments mainly from volcanic ashes; complex of

- Well drained, moderately deep to deep, dark brown, frible and slightly smeary, fine gravely, sandy clay loam to sandy clay, with a humic topsoil
- Imperfectly drained, moderately dee to deep, strong brown, mottled, firm and brittle,sandy clay to clay (Gamblian lake of the Central Rift Valley)

R Volcanic footridges

R3: soil developed on tertiary basic igneous rocks; well drained, extremely deep, dusky red to dark reddish brown, friable clay with inclusions of well drained, moderately deep, dark red to dark reddish brown, friable clay over rock, pisoferric or petroferric material

Ux Uplands, undifferentiated levels

Ux3: soil developed on basic igneous rocks, but with influence of vocaninc ash predominant; well drained, deep to very deep, dark reddish brown to dark red, firm clay with inclusions of imperfectly drained, moderately deep, dark greyish brown clay

UX5: well drained, very deep, dark reddish brown to very dark greyish brown, friable and slightly smeary clay, with a humic topsoil (ando-luvic PHAEOZEMS)

UX7: Well drained, shallow, dark brown, friable, strongly calcareous, strongly saline and moderately sodic, stony loam with a stone surface (dissected older piedmont plain) (calcaric REGOSOLS, stone-mantle and saline-sodic phase)

F7 : soils developed on colluvium from acid igneous rocks (rhyolites) with volcanic ash admixture. Moderately well dearine, deep to very deep, reddish brown, friable clay with an acid humic topsoil (ando-humic ACRISOLS)

Appendix 3.1 : Discharge measurements at Gauging station 2GB07 of Malewa river

Date: 03.10.2003
Coordinates (TM) : 212101,9964638

Measure-ment num-ber	Distance from initial point	Width	Depth	Average velocity	Area	Discharge
		cm	cm	m/s	cm2	m3/s
1	0	15	0	0.0	150	0.00
2	30	35	20	0.4	700	0.03
3	70	45	45	0.9	2025	0.19
4	120	50	32	0.7	1600	0.12
5	170	50	33	0.9	1650	0.15
6	220	50	26	0.8	1300	0.10
7	270	50	25	1.0	1250	0.13
8	320	50	38	1.1	1900	0.21
9	370	50	36	1.5	1800	0.27
10	420	50	46	1.2	2300	0.27
11	470	50	41	1.0	2050	0.20
12	520	50	36	1.4	1800	0.25
13	570	50	43	1.1	2150	0.23
14	620	50	45	0.8	2250	0.18
15	670	50	27	0.6	1350	0.08
16	720	40	27	0.4	1080	0.05
17	750	15	0	0.0	202.5	0.00
					25558	**2.46**
		Total Discharge		**2.46m3/s**		
		Average Velocity		**0.96m/s**		

Appendix 3.2 : Discharge measurements of Gilgil river

Date: 25.09.2003

Coordinates (TM) : 204571, 9946052

Measurement number	Distance from initial point	Width	Depth	Average velocity	Area	Discharge
		cm	cm	m/s	cm2	m3/s
1	0	25	31	0.3	387.5	0.01
2	50	50	40	0.7	2000	0.13
3	100	50	44	0.8	2200	0.17
4	150	50	42	0.9	2100	0.20
5	200	50	38	1.0	1900	0.18
6	250	50	37	1.0	1850	0.19
7	300	50	40	1.1	2000	0.22
8	350	50	38	1.2	1900	0.23
9	400	50	37	1.0	1850	0.18
10	450	50	30	0.5	1500	0.07
11	500	25	33	0.5	412.5	0.02
						1.61
		Total Discharge		**1.61 m3/s**		
		Average Velocity		**0.8 m/s**		

Appendix 4.1 Weather Generator Statistic and Probability Value

Station number: 9036025

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Pwd	0.12	0.16	0.21	0.36	0.31	0.25	0.21	0.26	0.27	0.27	0.30	0.17
pww	0.54	0.51	0.56	0.68	0.63	0.61	0.53	0.55	0.62	0.60	0.65	0.54
Mean	52.5	61.20	80.40	166.2	151.5	99.6	71.40	88.2	105.0	100.8	111.9	62.7
SD	4.83	5.24	6.10	8.42	7.97	6.12	5.33	5.81	6.1	6.31	6.0	5.33
Skew	4.01	4.24	3.88	2.36	2.03	2.88	4.49	3.47	2.55	3.36	2.37	4.46
Number of rainy days	4.38	6.64	9.95	15.50	13.98	11.38	9.26	10.86	12.12	12.14	13.9	8.05

Station number: 9036241

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Pwd	0.12	0.13	0.18	0.36	0.29	0.28	0.29	0.31	0.30	0.27	0.28	0.15
pww	0.51	0.49	0.54	0.67	0.67	0.64	0.59	0.65	0.65	0.65	0.65	0.45
Mean	54.0	46.8	70.5	168.6	151.8	116.7	102.0	110.4	126.6	107.7	96.6	49.2
SD	5.13	4.67	5.92	9.13	8.01	6.51	6.19	6.02	6.65	5.96	5.50	4.94
Skew	4.45	4.87	4.03	3.37	2.17	2.22	3.27	3.43	2.29	2.43	2.41	6.41
Number of rainy days	4.05	4.89	8.11	14.18	13.29	12.03	11.79	13.53	12.95	11.71	11.61	5.89

Station number: 9036264

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Pwd	0.11	0.10	0.13	0.28	0.32	0.27	0.21	0.23	0.23	0.25	0.25	0.11
pww	0.48	0.54	0.49	0.72	0.66	0.55	0.60	0.61	0.58	0.62	0.68	0.48
Mean	44.1	39.6	47.7	165.6	151.2	100.5	74.4	90.3	79.2	82.2	82.8	44.7
SD	4.45	3.88	4.6	9.01	8.24	6.46	4.74	5.94	5.05	5.04	4.43	5.51
Skew	4.31	4.30	4.51	2.43	2.55	3.43	3.08	3.79	2.75	3.16	2.49	7.32
Number of rainy days	3.09	3.69	5.28	12.16	13.97	10.06	10.16	10.84	9.63	11.88	13.03	4.44