



**UNIVERSITY OF NAIROBI
SCHOOL OF ENGINEERING**

**PEDOTRANSFER FUNCTIONS FOR SATURATED HYDRAULIC
CONDUCTIVITY FOR SURFACE RUNOFF MODELING**

By

John Paul Odhiambo Obiero

B.Sc (Hons). Agric. Eng.:(Egerton), M.Sc Agric. Eng.; (Nairobi)

**A thesis submitted in fulfillment for the degree of Doctor of Philosophy
in the Department of Environmental and Biosystems Engineering in the
University of Nairobi**

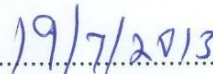
July 2013

Declaration

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

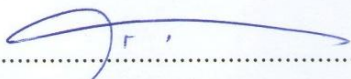

.....

John P. O. Obiero

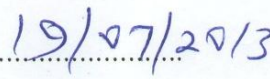

.....

Date

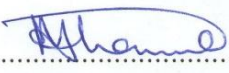
This thesis has been submitted for examination with our approval as university supervisors.


.....

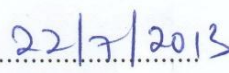
Prof. Lawrence O. M. Gumbe
University of Nairobi
Nairobi, Kenya


.....

Date


.....

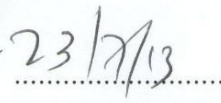
Dr. Mohammed A. Hassan
Technical University of Kenya
Nairobi, Kenya.


.....

Date


.....

Prof. Bernard N. K. Njoroge
University of Nairobi
Nairobi, Kenya


.....

Date

Dedication

This thesis is dedicated to my Wife, Jane Auma Odhiambo, my daughters Grace Laura Akinyi and Janet Mercy Achieng, late son Caleb Jeffia Obiero, parents, brothers and sisters for their moral support, encouragement and patience with me while carrying out this research work to its final conclusion.

Acknowledgement

I wish to thank my Supervisors Prof. L. O. Gumbe, Dr. M. A. Hassan and Prof. B. N. K. Njoroge for their continued support, encouragement, and guidance in the period of this research. I acknowledge the support of the University of Nairobi in funding the research work through Dean's committee research grant. I also wish to extend my sincere appreciation to departments and individuals at the University of Nairobi for their assistance and contributions that enabled this research work make significant progress to its conclusion. Gratitude to Prof. J. O. Onyando for his assistance and guidance in hydrological model selection process and provision of relevant literature. I am grateful to the contribution made by the staff in the department of Geospatial and Space technology for training me in the area of GIS which was essential in this research, and who also assisted me in processing of my geospatial data through photogrammetric mapping. My sincere gratitude to Prof. G. C. Mulaku for guidance in GIS related work in the project. Special thanks to David Siriba for his contributions in the mapping process, Richard Omwega for a detailed and reliable mapping of the contours, river networks etc., Naftali Oyugi, for tracing of contour maps, G.T.O. Osewe among others. I am grateful to Joseph Sang (ICRAF) and Philip Ogwen (KIRDI) for their support in data preparation and analysis using SWAT model. I acknowledge support from members of staff at the department of Environmental and Biosystems Engineering (EBE) for their support during this research. I wish to thank Dr. C. T. Omuto (EBE) for his support in acquisition of soils data from the international soils data base and in development of pedotransfer functions. I thank Januarius Agullo (EBE) for his support with statistical software for analysis and programming aspects. I wish acknowledge support from institutions that provided data used in the study including Kenya Soil Survey (KSS), Ministries of Lands and Settlement and of Water and Irrigation, Kenya, Kenya Meteorological Department, and the Natural Resource Monitoring, Management and Modeling (NRM³), Nanyuki, Kenya, Regional Centre for Resource Mapping Research and Development (RCMRD), Kasarani, Kenya. I also wish to acknowledge the useful comments and contributions from the anonymous reviewers through the journal of water resource and protection who played a key role in improving the manuscript that was published from part of this research thereby improving the quality of the thesis writing.

PEDOTRANSFER FUNCTIONS FOR SATURATED HYDRAULIC CONDUCTIVITY FOR SURFACE RUNOFF MODELING

**By John P. O. Obiero, Department of Environmental and Biosystems Engineering,
University of Nairobi, P. O. Box 30197, Nairobi, Kenya.**

Abstract

The study involved development of pedotransfer functions (PTFs) for determining saturated hydraulic conductivity (K_s) used in surface flow prediction. This preceded evaluation of existing PTFs for K_s in flow simulation. The pedotransfer functions were developed to predict parameters used in the determination of K_s using selected basic soil properties. The Soil Water Assessment Tool (SWAT) model was used in flow prediction in the Naro Moru river catchment of the Ewaso Ng'iro river basin, Kenya. The developed pedotransfer functions were then used in the simulation of surface runoff on the catchment and their performance in surface flow prediction compared with that of existing pedotransfer functions.

Initial model runs during flow simulation yielded poor daily flow simulations compared to monthly simulations. This was attributed to differences in the timing of peak discharges for the observed and simulated hydrographs. The model was calibrated for a three year period followed by a three year validation period based on monthly flows. Calibration results yielded acceptable, but modest agreement between observed and simulated monthly stream flows. The modest model performance was associated with input data deficiencies and model limitations. The results indicated that the model could be adapted to the local conditions. Manual flow calibration was performed to improve simulation results initially based on average annual conditions followed by monthly calibration. There was significant improvement in the model performance based on monthly flow simulations. The model simulation of surface flow registered better performance compared to base flow and total flow indicating the model to be a better simulator of surface flow than baseflow.

Observed and predicted surface runoff was compared to evaluate performance of existing PTFs. Model performance was similar for the existing PTFs selected. There was diversity

in performance of PTFs when used for surface runoff prediction. It was felt there is the need for continued development of PTFs for predicting K_s . The developed PTFs were evaluated for accuracy and reliability. The PTFs developed for saturated soil moisture content (θ_s) produced better performance in reliability compared to the remaining parameters in the van Genuchten moisture retention equation. The developed pedotransfer functions were then used in predicting K_s for surface flow simulation. The model performance in surface runoff simulation using developed PTFs was found acceptable. The study provides insight in developing equations for predicting K_s from basic soil properties being an input parameter in hydrological models. Hydrologic modeling plays a significant role in enabling policy makers, watershed planners and managers make appropriate decisions consistent with sustainable management of watershed resources.

TABLE OF CONTENTS

DECLARATION.....	ERROR! BOOKMARK NOT DEFINED.
DEDICATION.....	II
ACKNOWLEDGEMENTS	III
ABSTRACT.....	IV
LIST OF FIGURES.....	IX
LIST OF TABLES	XI
LIST OF ABBREVIATIONS	XIV
1 INTRODUCTION	1
1.1 BACKGROUND.....	1
1.2 STATEMENT OF THE PROBLEM.....	2
1.3 JUSTIFICATION	5
1.4 OBJECTIVES	6
2. LITERATURE REVIEW	7
2.1 EVALUATION OF HYDROLOGIC SYSTEM MODELS PERFORMANCE	7
2.1.1 <i>Calibration and validation of hydrologic system models.....</i>	<i>10</i>
2.2 HYDROLOGIC MODEL APPLICATIONS	12
2.2.1 <i>Modeling effects of Land-Use Change on Watershed Hydrology</i>	<i>12</i>
2.2.2 <i>Modeling the Impact of Climate Change on Stream flow and surface runoff..</i>	<i>13</i>
2.3 IMPROVEMENT OF A HYDROLOGICAL MODEL PERFORMANCE.....	14
2.4 MATHEMATICAL MODELING OF HYDRAULIC CONDUCTIVITY FOR SURFACE FLOW PREDICTION.....	16
2.4.1 <i>Saturated hydraulic conductivity prediction.....</i>	<i>16</i>
2.4.2 <i>Effective hydraulic conductivity (Ke).....</i>	<i>21</i>
2.4.3 <i>Development of pedotransfer functions for saturated hydraulic conductivity.....</i>	<i>22</i>
2.5. THEORETICAL EQUATIONS USED IN DEVELOPING PEDOTRANSFER FUNCTIONS FOR Ks	27
2.6 THE SWAT MODEL REVIEW	30
2.6.1 <i>Brief Description of the SWAT Model.....</i>	<i>30</i>
2.6.2 <i>Swat Model Origin and Applications</i>	<i>31</i>
3. METHODOLOGY	32
3.1 THE STUDY CATCHMENT	32
3.1.1 <i>Description of the study area.....</i>	<i>32</i>
3.1.2 <i>Available Data.....</i>	<i>36</i>
3.2 DATA ACQUISITION AND PRELIMINARY ANALYSIS.	37
3.2.1 <i>Catchment Characterization (Topographical maps acquisition and processing) </i>	<i>37</i>
3.2.2 <i>Photogrammetric Mapping (Aerial photo acquisition and processing)</i>	<i>38</i>
3.2.3 <i>Digitizing of the contours and drainage networks.....</i>	<i>38</i>
3.2.4 <i>Hydro meteorological data acquisition and processing</i>	<i>40</i>

3.3	DATA ANALYSIS	44
3.3.1	<i>Model Selection</i>	44
3.3.2	<i>Data Preparation for Hydrological Modelling Using SWAT</i>	45
3.3.3	<i>Model Evaluation</i>	55
3.3.4	<i>Manual calibration</i>	56
3.3.5	<i>Selection and evaluation of pedotransfer functions for estimating K_s</i>	58
3.3.6	<i>Development of pedotransfer functions</i>	59
3.4	VERIFICATION OF DEVELOPED PEDOTRANSFER FUNCTIONS FOR SATURATED HYDRAULIC CONDUCTIVITY IN SURFACE FLOW PREDICTION.....	62
4	RESULTS AND DISCUSSION.....	64
4.1	PRELIMINARY ASSESSMENT OF MODEL PERFORMANCE	64
4.2	SENSITIVITY ANALYSIS AND MODEL CALIBRATION	68
4.3	MODEL VALIDATION	70
4.4	GENERAL ASSESSMENT OF MODEL PERFORMANCE	72
4.5	MANUAL CALIBRATION AND MODEL IMPROVEMENT	73
4.5.1	<i>Separation of base flow and surface runoff</i>	73
4.5.2	<i>Manual Calibration Results</i>	75
4.5.2.1	<i>Water balance and total flow calibration</i>	75
4.5.2.2	<i>Model performance in surface flow and base flow simulation after annual flow calibration</i>	83
4.5.2.3	87
4.5.3	<i>Summary of model performance assessment during manual calibration</i>	89
4.6	EVALUATION OF SELECTED PEDOTRANSFER FUNCTIONS IN ESTIMATING SATURATED HYDRAULIC CONDUCTIVITY (K_s) BASED ON MODEL PERFORMANCE IN SURFACE RUNOFF PREDICTION.....	91
4.7	DEVELOPMENT OF PEDOTRANSFER FUNCTIONS	96
4.7.1	<i>Evaluation of the distribution of the moisture retention parameters</i>	96
4.7.2	<i>Statistical linear regression analysis</i>	99
4.7.3	<i>Multiple regression analysis and accuracy of the pedotransfer functions developed</i>	99
4.7.4	<i>Analysis of cross correlations and correlation matrix</i>	105
4.7.5	<i>Mutiple linear regression results</i>	106
4.7.6	<i>Validation of the developed pedotransfer functions to evaluate reliability</i>	107
4.8	SURFACE RUNOFF SIMULATION USING SATURATED HYDRAULIC CONDUCTIVITY VALUES PREDICTED USING DEVELOPED PEDOTRANSFER FUNCTIONS.....	113
5	CONCLUSIONS AND RECOMMENDATIONS	117
6.	REFERENCES	119
	APPENDIX 1: PHOTOGRAPHIC MAPPING PROCEDURE FOR NARO MORU CATCHMENT	128
	APPENDIX 2: PROCEDURE USED IN THE DIGITIZATION OF NARO MORU CATCHMENT	132
	APPENDIX 3: RUNOFF CURVE NUMBERS FOR VARIOUS LAND COVER TYPES	136

APPENDIX 4: PLOTS OF NORMALITY CHECK FOR RESPONSE VARIABLES
..... 138

APPENDIX 5: PERFORMANCE RATINGS FOR STATISTICAL MEASURES
..... 141

**APPENDIX 6: QUALITIES OF FIT BETWEEN VAN GENUCHTEN MOISTURE
RETENTION PARAMETERS AND BASIC SOIL PROPERTIES IN LINEAR
REGRESSION** 142

**APPENDIX 7: ANALYSIS OF CROSS CORRELATION RESULTS FOR
PREDICTOR VARIABLES IN PEDOTRANSFER FUNCTION DEVELOPMENT**
..... 147

LIST OF FIGURES

Figure 3.1 Location of the study area	33
Figure 3.2. Drainage network and gauging stations in the Naro Moru catchment	34
Figure 3.3. Land use types covering the study area	35
Figure 3.4 Soil types in the Naro Moru catchment	36
Figure 3.5. Flow diagram showing sequence of events followed to obtain digital map of the test catchment	39
Figure 3.6: Location of meteorological stations in the vicinity of the catchment.	41
Figure 3.7. contour network derived for the study area Figure 3.6. Contour network derived for the study area	46
Figure 3.8. Digital elevation model derived from contour network	47
Figure 3.9 Major soil units in the study area	50
Figure 4.1. Hydrographs of observed and simulated daily flows during the period 1/7/92 to 31/6/95	64
Figure 4.2. A comparison of simulated and observed daily flow during the period 1/7/92 to 31/6/95	65
Figure 4.3. Hydrographs of simulated and observed mean daily flows in month for the period July 1992 to June 1995	67
Figure 4.4. Comparison of observed and predicted mean daily flows in the month for the period 1/7/1992 to 30/6/1995.....	67
Figure 4.5 Hydrographs of simulated and observed mean daily flows in month for the validation period 1/1/98 to 31/12/2000	71
Figure 4.6. Comparison of observed and predicted mean daily flows in the month for the period 1/1/1998 to 31/12/2000.....	71
Figure 4.7. Hydrograph of observed and simulated monthly total flows after annual flow calibration	82
Figure 4.8 Comparison of simulated and observed average daily flow in month after calibration of average annual water yields	82
Figure 4.9. Hydrograph of observed and simulated monthly surface flows after annual flow calibration	84

Figure 4.10. Comparison of simulated and observed average daily surface flow in month after calibration of average annual water yields	85
4.11 Hydrograph of observed and simulated monthly base flows after annual flow calibration	85
4.12 Comparison of simulated and observed daily base flow in month after calibration of average annual water yields.....	86
4.13 Hydrographs of simulated monthly total flow, base flow and surface flow after annual flow calibration	86
Figure 4.14. Monthly time series of simulated and observed surface flows in the Naro Moru river catchment in the period 1/1/1998 to 31/12/2000.....	88
Figure 4.15. Regression of simulated average daily surface flow in month and observed surface flows during validation.....	88
Figure 4.16. Check for normality of distribution of the saturated soil moisture content (θ_s) showing the histogram and normality plot.	98
Figure 4.17: Response versus predictor variables correlations with corresponding best possible quality fit.....	102
Figure 4.18 : Predicted value of transformed variable $\frac{1}{\theta_s}$ versus measured by correlation for a developed equation that yielded the best possible quality of fit.....	109
Figure 4.19 : Predicted value of transformed variable e^{θ_r} versus measured by correlation for a developed equation that yielded the best possible quality of fit.....	110
Figure 4.20 : Predicted value of transformed variable $\sqrt{\alpha}$ versus measured by correlation for a developed equation that yielded the best possible quality of fit.....	111
Figure 4.21 : Predicted value of transformed variable \sqrt{n} versus measured by correlation for a developed equation that yielded the best possible quality of fit.....	112
Figure 4.22: Hydrographs of simulated and observed surface runoff based on K_s values determined using developed pedotransfer functions.....	113
Figure 4.23: Comparison of simulated and observed surface runoff using linear regression based on the value of K_s calculated using developed pedotransfer functions.....	114
Figure 4.24: Comparison of simulated and observed surface flow for the developed and some existing PTFs.....	116
Figure A2.1 Outline of map sheets covering study area.....	132

LIST OF TABLES

Table 2.1: Pedotransfer functions for estimating saturated hydraulic conductivity.....	20
Table 2.2: Soil parameters used in pedotransfer functions.....	29
Table 3.1: Summary of raw data available for the catchment under investigation.....	37
Table 3.2: List of rainfall stations inside and in the vicinity of the test catchment	42
Table 3.3: Reclassification of observed land uses to SWAT land types	48
Table 3.4: Land use look up table	48
Table 3.5 Soil properties for the major soil units in the study catchment	51
Table 3.6: Soil look up table	51
Table 3.7: Statistical analysis of daily precipitation data (1992-2000).....	54
Table 4.1: Sensitivity analysis results showing the order of sensitivity of input parameters	69
Table 4.2 Evaluation of model results for first flow simulation runs, calibration and validation	72
Table 4.3 An illustration of the daily output file for the month of January 1992 showing daily baseflow filter values	74
Table 4.4 Summary of the fraction of stream flow contributed by base flow for each of the 3 passes made by the base flow filter programme for the period 1992 to 1995	74
Table 4.5: Estimation of observed and simulated annual water yield, base flow and surface flow (mm) for annual calibration	75
Table: 4.6 Summary of annual observed and simulated water yield (mm).....	75
Table 4.7: Curve number values for various land use types in the sub basins used in the calibration	78
Table 4.8: Summary of annual observed and simulated flows following parameter adjustments during manual calibration.....	79
Table: 4.9 Summary of observed and simulated average annual yield before and after annual flow calibration	81
Table 4.10 Comparison of model performance for monthly simulation based on auto calibration and annual flow calibration processes.	83

Table 4.11: Comparison of model performance in monthly surface flow simulation after total flow calibration, monthly surface flow calibration and validation	89
Table 4.12 Calculated values of saturated hydraulic conductivity (K_s) based on pedotransfer functions evaluated in the study.....	92
Table 4.13 Values of selected performance measures in model surface flow simulation based on various pedotransfer functions used in estimating K_s	94
Table 4.14 Rating of selected pedotransfer functions in predicting K_s for surface runoff prediction using SWAT model during calibration and validation	94
Table 4.15 Normality of distribution check for parameters in the van Genuchten water retention characteristic and their transformations:	97
Table 4.16. Statistical parameters obtained on normality check for the saturated hydraulic conductivity data	99
Table 4.17: Quality of fit based between response variable transformations and selected predictor variable transformations in the multiple linear regression.....	104
Table 4.18. Summary of equations for transformations of van Genuchten parameters and their performance in reliability.....	108
Table 4.19: Comparison of performance of developed pedotransfer functions for K_s and existing PTFs in surface flow simulation.....	115
Table A1.1 selected control points for mapping using aerial photos and toposheets	129
Table A2.1 Areal extents of map sheets covering the test catchment.....	132
Table A3.1 Runoff curve numbers for other agricultural watersheds.....	136
Table A3.2 Runoff curve numbers for urban areas.....	137
Table A5.1: Performance ratings for recommended statistics in flow prediction for monthly time step.....	141
Table A6.1. Quality of fit between transformations of moisture retention parameters and sand transformations in the linear regression.	142
Table A6.2. Quality of fit between the moisture retention parameters and silt transformations in the linear regression.....	143

Table A6.3. Quality of fit between the moisture retention parameters and silt transformations in the linear regression.....	144
Table A6.4 Quality of fit between the moisture retention parameters and clay transformations in the linear regression analysis.....	145
Table A6.5 Quality of fit between the moisture retention parameters and organic carbon transformations in the linear regression analysis.....	146
Table A7.1 Cross correlation tests for transformed predictor variables for $\frac{1}{\theta_s}$	147
Table A7.2 Cross correlation tests for transformed predictor variables in e^{0r}	147
Table A7.3 Cross correlation tests for transformed predictor variables in $\sqrt{\alpha}$	148
Table A7.4 Cross correlation tests for transformed predictor variables in \sqrt{n}	148

LIST OF ABBREVIATIONS

AGNPS – Agricultural Non Point pollution Source
AI – Artificial Intelligence
API – Antecedent Precipitation Index
ARC – Antecedent Runoff Condition
AVSWAT – Arc-View SWAT
AWC – Available Water Capacity
BD – Bulk Density
C – Carbon content
CEC – Cation Exchange Capacity
CN – Curve Number
DEM – Digital Elevation Model
 D_v – Deviation volume
EBE – Environmental and Biosystems Engineering
EPIC – Environmental Impact Policy Climate
FAO – Food and Agriculture Organization
GIS – Geographical Information Systems
GLEAMS – Ground Water Loading Effects on Agricultural management Systems
CREAMS – Chemical, Runoff and Erosion from Agricultural Management Systems
HRU – Hydrologic Response Unit
IA – Index of Agreement
ICRAF – International Centre for Research in Agroforestry
ILWIS – Integrated Land and Water Information System
ISRIC – International Soil Reference and Information Centre
KENSOTER – Kenya Soil Terrain
KIRDI – Kenya Industrial Research and Development Institute
KSS – Kenya Soil Survey
 K_s – Saturated hydraulic conductivity
MARS – Multiple Regression Splines
ME – Mean Error
NRM³ – Natural Resource Monitoring, Management and Modeling
NRCS – Natural Resources Conservation Service
NSE – Nash-Sutcliffe Efficiency

OM – Organic Matter
PBIAS – Percentage Bias
PE – Percentage Error
PTF – Pedotransfer Function
RCMRD – Regional Centre for Mapping of Resources for Development
RMSE – Root Mean Square Error
SNOWMOD – Snowmelt Model
RSR – Observation Standard Deviation Ratio
SCS – Soil Conservation Service
SMC – Soil Moisture Characteristic
SNOWMOD – Snowmelt Model
SSE – Sum of Squared Errors
SWAT – Soil Water Assessment Tool
SWRRB – Simulator for Water Resources in Rural Basins
USDA-ARS – United States Department of Agriculture – Agricultural Research Service
UNESCO – United Nations Educational, Scientific and Cultural Organization
USLE – Universal Soil Loss Equation
UTM – Universal Transverse Mercator

1 INTRODUCTION

1.1 Background

Information on surface runoff is needed for several purposes in soil and water management (Pathak, et al. 1989). For example, the prediction of both volume and rate of runoff from a watershed is vital in the good design of hydraulic structures (Van Mullem, 1991). Such structures include those used in soil and water conservation, rainwater harvesting, flood control, hydro electric power generation, etc. In order to simulate surface runoff in catchments, a number of models have been developed with rainfall being a major input in such runoff simulation models. The transformation of rainfall into runoff involves many other hydrological processes that include infiltration, evapotranspiration, deep percolation, lateral sub surface flow, etc, each of which has factors that influence them. Factors that influence the hydrological processes and which determine the surface runoff volumes include land use, soil characteristics, topography, vegetation, management practices, etc.

A number of runoff simulation models have been developed in an attempt to estimate surface runoff taking into account the effects of factors that affect the runoff generation process. These models use different approaches and techniques in predicting the runoff process. Studies involving prediction of surface runoff process based on runoff simulation models have been useful in examining the effects of various land use and management practices on water flow behaviour for the purposes of Natural Resources Management in watersheds. Hence development of knowledge on surface runoff prediction would be useful in water resources planning especially through simulation of the effects of management strategies on the surface water resources in the watershed. Scientists have therefore made attempts to study the process of the runoff generation for water resources management based on developed models. Singh et al. (2008) reviewed the understanding of the runoff generation process and simulation of daily stream flow for a 516km² basin in the Himalayas using the SNOWMOD model. The study was notably useful for the planning and management of water resources in high altitude areas and for designing hydropower projects. Sharda, et al. (2006) employed the use of artificial intelligence based techniques to predict surface runoff, base flow and total flow

as affected by rainfall and morphological features of micro-watersheds. The study was part of an attempt to quantify the effect of environmental and morphological factors on flow behaviour in the micro-watersheds for the purposes of efficient planning and execution of water management practices in a sustainable manner in the Himalayas mountains. Surface runoff is closely related to soil erosion and sedimentation in a watershed. The runoff rate therefore gives an indication of how much soil is being lost and the resulting sedimentation of reservoirs used in water supply or for hydro electric power generation. Potential effects of changes in climate on surface water hydrology may be studied using runoff simulation models. Such models have been developed for selected watersheds to analyze hydrologic sensitivity for selected watersheds to climate change scenarios (Bekele and Knapp, 2008) for the purposes of regional water supply planning efforts.

In surface runoff prediction, hydraulic conductivity serves as an important parameter that influences hydrological process that affect flow in streams and rivers. For instance, ground water flow is determined by saturated hydraulic conductivity. A number of formulations have been developed over the years in an attempt to predict soil hydraulic conductivity from readily measurable soil properties. The need for mathematical modeling of saturated hydraulic conductivity, K_s , arises from consideration of the fact that insitu or laboratory measurements of hydraulic conductivity are time consuming, labour intensive and expensive as noted by Wells et al. (2006), making it practically unlikely in reality to collect permeability data. Besides, direct measurements are also considered unreliable for site specific applications and accuracy or replicability of the measurements cannot be ascertained. Use of pedotransfer functions is an alternative to determination of hydraulic conductivity and involves relationships that enable the soil property to be predicted based on measurable basic soil properties available in most soil data bases.

1.2 Statement of the problem

Runoff modeling is notably important in the management of soil and water resources of a watershed. A number of models have been developed to simulate the runoff process. The

models use different input data, are applicable in different climatic and soil conditions and vary in their level of complexity and accuracy in predicting surface runoff on selected catchments. Hydrologic models are not universal in their applicability and their reliability of runoff prediction is catchment specific. For a model to be used in estimating surface runoff at a selected site, its performance at the chosen site should be evaluated to examine its efficiency in the runoff prediction and necessary adjustments made on the model inputs to adapt it to the selected site and conditions. This may be partly achieved through improved model calibration and further validation to establish its applicability for runoff prediction while identifying possible reasons for poor performance. This research project aims to test the applicability of a selected watershed model in predicting stream flow on a chosen catchment located in Kenya, propose possible ways of improving its performance on flow simulation and establish the suitability of the model in estimating the surface runoff on the selected catchment in Kenya. The study further involves development of PTFs in predicting saturated hydraulic conductivity using an input parameter to the SWAT model for which it is sensitive in runoff prediction.

In most developing countries, the data availability poses major challenge and significantly influences the choice of models and modeling approach. It is indeed challenging to identify a model whose data requirement fits with the data available for the chosen catchment. The models in most cases have been developed for environments other than those for which one may wish to apply them for catchment management. It would of major interest to identify a model that takes into account most of the runoff generation factors in predicting surface runoff and also able to be run using readily available data on local catchments without putting too much demand on data input requirements. A number of catchments in developing countries like Kenya are not gauged and most of those that are gauged do not have continuous daily data extending over decades owing to breaks resulting from missing data, hence obtaining continuous readily available historical data on stream flows is uncommon. Through adaptation of runoff simulation models to specific sites, it would be possible to make estimates of stream flows through prediction for simulation of watershed hydrologic responses. This would offset problems of costs associated with setting up stream flow measuring devices which are not only

expensive to set up but also require maintenance, an expensive process as well. Sarangi et al. (2008) points out that continuous recording of rainfall and runoff is not only cumbersome, but also a costly affair. Runoff simulation models employ different techniques of surface runoff prediction. The suitability of these methods differ in various environments and requires assessment. The suitability of Curve Number (CN) procedure for continuous runoff prediction has been examined by Lamont et al. (2008) who noted some significant error associated with the numerical procedure.

An input parameter in many hydrological models for predicting surface runoff is the hydraulic conductivity (K_s). Determination of the K_s value in a watershed is time consuming, tedious and costly. Besides, there is no specific method that can be considered as being most accurate in determining this variable soil hydraulic property as the method used depends on the soil and environmental conditions. Furthermore, data on K_s is not readily available in most local soil survey data and even where they are available, reliability of their measured values is unknown. There is uncertainty in the prediction of K_s using existing pedotransfer functions some of which have been developed in Europe and America and which may not be applicable worldwide. Methods to develop these functions are also unknown if not well understood especially in Africa. Well known moisture characteristics functions are associated with parameters that can be determined using moisture retention characteristic. Data on moisture retention is not readily available in most soil data bases. It becomes essential to estimate such parameters using equations that relate them to the basic and easily available soil properties like bulk density, texture etc. Equations that relate moisture retention parameters to the basic soil properties are not easily available and attempts to develop them is rare. It is noted by Saxton and Rawls (2006) that modern simulation and analysis of hydrologic processes relies heavily on the accurate and reliable description of soil water holding and transmission characteristics yet hydrologists lack the capacity or time to perform field or laboratory determination of such properties which include the saturated hydraulic conductivity. The author further points out that estimated values of such hydraulic characteristics can be determined from local soil data bases like soil maps and published water retention and saturated hydraulic conductivity estimates. This involves predicting

saturated hydraulic conductivity from basic soil properties like texture and bulk density using developed pedotransfer functions based on analysis of the existing data sets, an approach that has been applied in agricultural hydrology and watershed management.

1.3 Justification

Watershed models that have been developed for simulation of surface runoff are not universal in terms of their accuracy in surface runoff prediction. They may over predict or under predict runoff depending on the location and the characteristics of the catchment in question. The model may be adapted, through calibration and validation processes, to an environment for which it was not developed. However, the model may have deficiencies associated with its inability to simulate some hydrological processes or require input parameters influencing the runoff process, but which may not be readily available or easily determined reliably. This may affect its performance significantly. Some input parameters for a watershed model may require improvement in their prediction techniques. Various models also differ in their approaches used to predict surface runoff. The prediction techniques vary in terms of the input parameters that are used, the variables considered and assumptions made. The methods used in estimating the input parameters or processes also differ in their accuracy and reliability. This partly explains the under-prediction or over-prediction of surface runoff by the model. There is need to assess some methodologies for estimating some input parameters. One such parameter is the saturated hydraulic conductivity.

Saturated hydraulic conductivity is among the parameters to which runoff simulated using SWAT hydrologic model is sensitive. The need for prediction of saturated hydraulic conductivity based on basic soil data is necessary to offset problems of cost and time in measuring it. Measurement of K_s is also in doubt in terms of accuracy and reliability. Prediction of K_s involves the use of pedotransfer functions. In hydrological modeling using such models as SWAT, determining the input parameter K_s (saturated hydraulic conductivity) is based on the use of pedotransfer functions that have been developed from elsewhere in Europe and America. The use of these pedotransfer functions

have not been evaluated for surface flow simulation using the model and hence their role in surface flow simulation is not known. It would also be of interest to show how pedotransfer functions can be developed from basic soils data. K_s can also be predicted using moisture retention curve described using moisture retention functions. Relationships between parameters of such functions and basic soil properties is rare in practice. In this study, several pedotransfer functions are evaluated to assess their performance in predicting K_s for surface runoff prediction using a watershed model. The research further demonstrates how pedotransfer functions can be developed from a data base of limited soils information in predicting moisture retention parameters, a step in predicting saturated hydraulic conductivity from basic, easily measurable or readily available data on basic soil properties that can be obtained from most available survey data.

1.4 Objectives

The broad objective of this study is to develop pedotransfer functions for determining saturated hydraulic conductivity for flow prediction on a catchment.

The specific objectives of the study are:

1. To identify soil parameters which are pertinent to saturated hydraulic conductivity.
2. Using statistical regression, to develop mathematical equations for predicting saturated hydraulic conductivity (K_s) using parameters in 1 above.
3. To verify mathematical equations developed in 2 above, using empirical data, in surface runoff simulation on a Kenyan catchment.

2. LITERATURE REVIEW

2.1 Evaluation of hydrologic system models performance

A number of hydrologic models have been developed for estimation of surface runoff or other water balance components in a catchment. In order for these models to be applied for surface runoff prediction on a catchment there is need for its performance to be tested and evaluated so as to establish the possibility of using it for a specified application on the watershed in question and where possible be improved for better performance. The process involves initial observation of model performance, calibration and validation of the simulation models. The evaluation process usually involves the use of observed data on stream flow, sediment discharge etc which is compared to the model predictions to establish the goodness of fit. The performance indicators of a model are established through the use of a number of measures of goodness of fit criteria. Such measures include the coefficient of determination (r^2), Nash Sutcliffe Coefficient (NSE), etc. One or more of these measures may be used in model evaluation. The measures help establish how well the model predictions compare with the observed data. Such measures have been used in a number of research studies to describe model performance. Jain and Sudheer (2008) points out that the quantitative assessment of the degree to which the modeled behaviour of system matches with the observations provide means of evaluating a model's predictive abilities, and further observes that the Nash Sutcliffe efficiency index is widely used in water resources, but cautions that the index alone should not be considered adequate in describing model performance and so proposes that other statistical measures be used alongside it to reach a definite conclusion about the hydrologic model performance.

Statistical parameters that have been used in evaluating hydrologic model performance are presented as follows:

Model efficiency factor presented as:

$$E = \frac{\sum_{i=1}^n (P_i - O_i)^2}{\sum_{i=1}^n (O_i - \bar{O})^2} \quad (2.1)$$

Where n is the total number of observations, O_i is the *i*th observed value, \bar{O} the mean observed values and P_i the predicted value. Sarangi, et al. (2008) used the model efficiency to select the best model for predicting surface runoff based on the E-value approaching one (1) while comparing performances of the curve number and geomorphological instantaneous unit hydrograph methods of surface runoff prediction. The coefficient of determination was used alongside it in the study.

The Nash and Sutcliffe efficiency (NSE), another statistical parameter for evaluating model performance is expressed as,

$$NSE = 1 - \frac{SSE}{\sum(\bar{Q} - Q)^2} \quad (2.2)$$

Where SSE is the sum of squared errors given by,

$$SSE = \sum(\hat{Q} - Q)^2 \quad (2.3)$$

\hat{Q} = The simulated value of stream flow (m^3/s).

Q = observed value of stream flow (m^3/s).

\bar{Q} = The average observed value of stream flow (m^3/s).

The NSE parameter takes values between $-\infty$ and 100%, in which the latter corresponds to a perfect fit between the observed and simulated values. NSE indicates the predictive power of hydrological models and determines the relative magnitude of the residual variance as noted by Dawadi and Ahmad (2012).

Levesque et al. (2008) notes that the use of NSE is recommended mainly because of its extensive use and adds further that an assessment of underestimation or overestimation of the total observed stream flow may be done using the deviation volume, D_v presented as,

$$D_v = 100 \frac{(\sum \hat{Q} - \sum Q)}{\sum Q} \quad (2.4)$$

\hat{Q} = simulated stream flow (m^3/s).

Q = observed stream flow (m^3/s).

The coefficient of determination, r^2 has also been used severally in evaluating model performance in various modeling studies. The correlation coefficient (r) has been used by Schmalz et al. (2008) in establishing correlation between measured and modeled discharges on selected catchments while modeling water balances and hydrological processes in lowland river basins. The method was used together with the Nash-Sutcliffe coefficient in the study. A good agreement is associated with the value of r approaching unity with a value of 1 indicating a perfect correlation.

Other measures of performance that may be used in evaluation of model performance include:

Percent bias (PBIAS) which measures the average tendency of simulated data to be larger or smaller than the corresponding observed values. An optimal value of 0.0 indicates accurate model simulation where as positive values indicate the extent to which the model underestimates the observed values and the negative values indicating the extent of model overestimation. It is expressed as:

$$PBIAS = \left[\frac{\sum_{i=1}^n (Y_i^{obs} - Y_i^{sim}) * 100}{\sum_{i=1}^n Y_i^{obs}} \right] \quad (2.5)$$

Where PBIAS expresses the percentage deviation of data being evaluated. Y_i^{obs} is the i th observation, Y_i^{sim} is the i th simulated value for the constituent being evaluated and n the number of observations.

Root Mean Square Error (RMSE) is an error index statistic expressed as

$$\mathbf{RMSE} = \left[\sqrt{\sum_{i=1}^n (Y_i^{obs} - Y_i^{sim})^2} \right] \quad (2.6)$$

A lower value of RMSE implies better model performance. Closely related to RMSE is the RSR calculated as the ratio of RMSE and the standard deviation of measured data. It varies from the optimal value of 0 for perfect model simulation to a larger positive value. It is expressed as;

$$\mathbf{RSR} = \frac{\left[\sqrt{\sum_{i=1}^n (Y_i^{obs} - Y_i^{sim})^2} \right]}{\left[\sqrt{\sum_{i=1}^n (Y_i^{obs} - Y_i^{mean})^2} \right]} \quad (2.7)$$

A lower RSR implies lower RMSE and hence better model simulation performance.

The above indicated statistical measures of model performance are further documented in Moriasi et al. (2007) who also indicates the ranges of values of some of the parameters and associated levels of performance in calibration and validation of hydrologic system models.

2.1.1 Calibration and validation of hydrologic system models

In order to test the applicability of a mathematical hydrologic model, on a selected catchment, it has to taken through a rigorous calibration and validation process. The data is initially prepared according to the prescribed format required for the data input to the measured model followed by test running of the model to examine whether it is able to successfully load the data. Based on the preliminary outputs from the model after it has

been run e.g. surface runoff, sediment discharge, ground water flow, etc, a comparison is made of the observed and predicted parameter values of the selected output parameter. The initial performance of the model is then assessed using performance indicators discussed earlier. Initial observations may show poor comparisons between the observed and predicted parameter values. In the subsequent calibration process, a combination of input parameters are determined for which the observed and predicted values would show significant improvement in the goodness of fit as measured by performance indicators like correlation coefficient, Nash Sutcliffe coefficient, etc. A proportion of data set acquired is usually used for the validation.

The need for calibration arises from the fact that models and their associated parameters are approximations to reality (Maidment, 1993). The process involves checking with observed data the parameter values estimated from physical considerations. In stream flow prediction, for instance, where flow records are available, one way of calibration involves the adjustment of one or more model parameters to give the best possible fit between the predicted and observed hydrographs. Adjustment of parameters may be carried out by manual trial and error adjustment of parameters or automatic optimization programmes. Chinnarasi et al. (2008) made a comparative study of two mathematical models for analyzing lateral erosion in which the models were validated by means of comparison with observations. The model calibration process involved comparing mean daily water levels with computed results for both the models. The average percentage error (PE) between the computed and measured values for a selected period was used as a basis for comparisons. The Soil Water Assessment Tool (SWAT) model has been calibrated for daily stream flows using manual and automatic calibration methods (Bekele and Knapp, 2008). This was done to assess the performance of a developed version of the model intended to improve the model's capability of simulating low flows. Comparison of observed and predicted flows has therefore become an acceptable basis for evaluation of hydrologic models' performance in predicting hydrologic behavior in catchments. In this study, a preliminary performance of the model performance in predicting stream flow is evaluated using automatic calibration method after a sensitivity analysis has been carried out to establish the order of sensitivity of the input parameters.

Thereafter, a detailed manual calibration is performed to improve model performance in flow simulation. The correlation coefficient, Coefficient of Determination, Nash-Sutcliffe Efficiency, percentage bias, and other statistical performance measures were used in the evaluation of model performance during calibration and validation periods.

2.2 Hydrologic Model Applications

2.2.1 Modeling effects of Land-Use Change on Watershed Hydrology

Land use practices in a watershed is a factor that significantly influences stream flow and hence surface runoff in a catchment. Changes in land use alter the hydrologic response of a catchment through its effects on various hydrologic processes like infiltration, interception, evapotranspiration, erosion and sedimentation, subsurface flow among others. As a result, water availability for various purposes that include irrigation, crop production, hydroelectric power generation, ground water exploration, etc are determined by the kind of land use/land cover prevailing in a watershed. For the purposes of planning and management of future water supply capabilities in a watershed and to evaluate catchment water resources, especially under changing land-use scenarios, it is important that potential effects of land use on the water resources is predicted with certainty. This may be done through watershed hydrologic modeling using various approaches. The Soil Water Assessment Tool (SWAT) has been, in various ways, used to study the effect of land use on catchment hydrology in a number of watersheds. In Technical Brief 2 (2007), a modeling exercise, based on SWAT, was carried out to estimate runoff from various land use types in upper Malalprabha catchment, India. Various land use scenarios were built into the model to study the impact of water availability in irrigated agriculture. Based on an existing trend in irrigation water demand at a downstream section, an ideal land use to guarantee the required river flow is modeled. Such a study presents a modeling approach in which the predicted impact of land use on stream flow may be used for planning land use for irrigation water management in agriculture. Land use scenario analysis has also been done by Heuvelmans et al. (2005), to predict the impact of land allocation on the hydrology and erosion on selected watersheds in which the rate variables used to describe the land use impacts included evapotranspiration, surface

runoff, discharge, ground water recharge, and soil loss through erosion. Such variables can be simulated using the SWAT model. Hydrologic modeling has also been used to understand the effects of land use/land cover changes on the hydrological behavior of a watershed. A study to investigate land use/land cover dynamics and impacts on stream flow has been conducted by Tadele and Förch (2007) on Hare river watershed, Ethiopia which drains a land area of 167.3km². This was in recognition of the fact that knowledge of the influence of land use/land cover changes would serve as an important tool for use by local governments and policy makers in formulating and implementing effective and appropriate response strategies intended to minimize undesirable effects of future land use/land cover changes. The study provided insight to understanding the upstream-downstream linkages with respect to irrigation water use by relating seasonal stream variability to land use/land cover dynamics. The study demonstrates a typical application of modeling land use/land cover for the purposes of water resources management in irrigated agriculture. Runoff simulation was based on the use of SWAT model and land cover maps used to analyze land use/land cover dynamics were chosen for years 1967, 1975 and 2004. An assessment of the fact that land use influences stream flow was done by performing simulations for a chosen period using different maps.

2.2.2 Modeling the Impact of Climate Change on Stream flow and surface runoff

Flow in streams/rivers is significantly affected by variations in weather parameters associated with climate change especially temperatures and rainfall. Such changes may be modeled to study the resulting effects on surface flow in streams and rivers. Such studies are important in simulating variations in the river flows to establish possibility of floods or low flows, droughts, etc, when they are likely to occur and the expected magnitudes, an important approach in water resources planning. Climate change and extreme weather related events like droughts, floods, etc significantly affect such important sectors as agriculture, energy, water resources, among others. A modeling approach to investigate climate change impacts on stream flow has been attempted by Githui et al. (2008) using the SWAT model on the Nzoia catchment in the Lake Victoria basin, Kenya. This was in view of the fact that the impact on the ability of the basin to support community livelihoods in the region resulting from stream flow variations

associated with climate change is of primary concern. The approach involved setting up the model based on available historical spatial and temporal data followed by calibration using observed stream flow to establish the model performance in stream flow prediction. Based on predicted future climate change scenarios obtained from the Global Circulation Models, it was possible to simulate future changes in stream flow and its relationship to changes in temperature and rainfall. Similarly, Odira et al. (2010) demonstrated the use of SWAT model in evaluating the effect of land use/land cover change in the hydrology of Nzoia catchment for use in flood management. Noting that evaluation of water resources in light of future climate change is important for sustainable planning and management of the resource, Obuobie and Bernd (2008) applied the Soil Water Assessment Tool (SWAT), to simulate the present considered to be 1990-2000 and the future (2030-2039) water resources. A calibration period of 1981-1991 was used and the model validated for the period 1992-1999, with the model being evaluated using performance indicators that included Nash-Sutcliff Model Efficiency (NSE), coefficient of determination (R^2), and Index of Agreement (IA). Comparison of observed and simulated flow formed the basis of model evaluation, in which good correlations were noted between simulated and observed flows. The study demonstrates how a calibrated model may be used in the prediction of future changes of flow in the rivers resulting from rainfall increases associated with climate change. Based on this study, reliability of the water resources can be assessed and occurrence of extreme events like floods and droughts predicted to help in watershed planning.

2.3 Improvement of a Hydrological model performance

In order to improve the performance of a hydrological model in predicting surface runoff, several approaches may be used as has been attempted by scientists. Lamont et al. (2008), in an attempt to improve the simulation of surface runoff using curve number technique, developed a soil physics model intended to calculate the soil moisture content that would be equivalent to an Antecedent Runoff Condition (ARC), an input into the NRCS curve number. The developed model was to provide estimates of soil moisture and infiltration parameters based on soil type and soil depth. Bekele and Knapp (2008), in developing the

SWAT model to improve flow simulations, introduced a level pool routing algorithm to simulate the reservoir storage routing, which was incorporated into the SWAT model. Sharda et al. (2006) developed artificial intelligence (AI) models based on Multiple Regression Splines (MARS) techniques for use in predicting surface runoff, base flow and total flow in the Himalayan mountains. This technique involved the use of data from two (2) watersheds to develop the MARS models and exploring their applicability on an ungauged watershed. Daily rainfall, runoff, base flow and total flow recorded over two years in three watersheds were used. The adoption of the use of AI models was based on the argument that they require fewer easily available data and measurable input parameters which can be used to simulate complex phenomenon. The author points out that process based models (e.g. SWAT, AGNPS, etc.) require a large number of input parameters that cannot be easily and accurately measured under field conditions. The AI models, however, are site specific, but can be applied under analogous climatic conditions. This approach to model development required that the watershed used in the study be gauged to monitor daily rainfall, runoff at their respective outlets through construction of broad based weirs equipped with automatic water level recorders and recording as well as non-recording gauges. This requirement is expensive in terms of investment and maintenance. The model performance was evaluated through computation of a number of statistical parameters that included; correlation coefficient, absolute deviations, average deviations etc to compare observed and predicted values. The variables identified as being most important for simulating runoff in hilly watersheds were API5, day of the year, runoff estimated by curve number method and watershed area. Observed and predicted values were found to be in reasonable agreement using this method. In this study, prediction of saturated hydraulic conductivity was chosen as a way of making a contribution to development of an important input parameter in hydrological model. The approach involved development of pedotransfer functions for predicting saturated hydraulic conductivity being an input parameter in the SWAT model that is not readily available in the existing data bases and also cannot be easily and accurately measured. It involved prediction of K_s from easily measurable basic soil properties.

2.4 Mathematical modeling of hydraulic conductivity for surface flow prediction

Methods used to determine hydraulic conductivity has been categorized (Zayani et al., 1992), among them predictive methods involving mathematical relationships. Theoretical methods for the estimation of hydraulic conductivity from basic physical and chemical soil properties like % sand, % clay, % organic matter etc or soil moisture characteristic have been developed. Some of the proposed relations may be exemplified as follows;

2.4.1 Saturated hydraulic conductivity prediction

Young and Gowing (1996) proposed the following relationship for estimating saturated hydraulic conductivity, K_s (mm/h).

$$K_s = 0.00035 \left[\frac{\phi_e - \theta_r}{1 - \phi_e} \left[\frac{BD}{\theta_R} \right] \right]^2 C^2 \quad (2.8)$$

Where ϕ_e = Effective soil porosity (vol/vol).
 θ_r = residual soil moisture content (vol/vol).
 BD = Soil bulk density (g/cm^3)
 C = soil carbon content (%).
 $\phi_e = \phi - \theta_r$

$$\phi = \frac{2.65 - \rho}{2.65} \quad \text{in which } \rho = \text{soil bulk density.} \quad (2.9)$$

$$\theta_R = 0.2 + 0.1 \left[(OM) + 0.25(CI)(CEC)^{0.45} \left(\frac{BD}{100} \right) \right] \quad (2.10)$$

$$\theta_r = (0.2 + 0.1(OM) + 0.25(CI)(CEC)^{0.45}) \frac{BD}{100} \quad (2.11)$$

Where OM = % organic matter content
 CI = % clay content
 CEC = Fractional Cation Exchange Capacity of Clay (cm cm^{-1})
 BD = Bulk density (gm/cm^3)

$$\text{Entrap air} = \frac{1.0 - 3.8 + 0.0019 \left[CI^2 - 0.337(Sa) + 0.126(CEC)(CI) + OM \right]}{100} \quad (2.12)$$

Where CI = % clay content
 Sa = % sand content
 OM = % organic matter content.

$$C = 0.17 + 0.181(CI) - 0.00000069(Sa^2)(CI^2) - 0.00000041(Sa^2)(100-Sa-CI)^2 + 0.000118(Sa^2)(BD^2) + 0.000049(Sa^2)(CI) - 0.000085(100-Sa-CI)(CI^2). \quad (2.13)$$

The soil water content at field saturation (θ_s) is given by:

$$\theta = \phi - \text{EntrapAir}$$

Lorentz et al. (2001) presented the following equations for estimating hydraulic conductivity. An equation reportedly developed by Lorentz in 1995 for estimating saturated hydraulic conductivity based on Brooks-Corey hydraulic characteristic function is presented as:

$$K_s = \phi \frac{\rho_g \sigma^2}{k_o \mu P_b^2} \left[1 + 0.173 \left[\frac{h_d}{\theta_e} \right]^{\frac{\lambda}{2+\lambda}} \lambda^{0.185} \right] (\theta_s - \theta_r)^2 \left[\frac{\lambda}{2+\lambda} \right] \quad (2.14)$$

Where ϕ =porosity
 γ = pore size distribution index
 ρ = density of water
 g = gravitational acceleration
 σ = surface tension of water
 k_o = pore shape factor, equal to 2.5 and
 P_b = air entry pressure = $h_d \rho g$

The author further notes that three hydraulic characteristic functions have been used in the determination of hydraulic conductivity. These are Van Genuchten, Brookes Corey and Campbel functions presented as follows:

$$s_e = \left[\frac{1}{(1 + \alpha h)^n} \right]^m \quad (\text{Van Genuchten relationship}) \quad (2.15)$$

Where

$$S_e = \text{effective saturation} = \left[\frac{\theta - \theta_r}{\theta_s - \theta_r} \right] \quad (2.16)$$

θ = volumetric water content ($\text{mm}^3 \text{mm}^{-3}$)
 θ_r = residual volumetric water content ($\text{mm}^3 \text{mm}^{-3}$)
 θ_s = saturated volumetric water content ($\text{mm}^3 \text{mm}^{-3}$)
 α = air entry parameter (mm^{-1})
 h = matric pressure head (mm)
 n = pore size distribution parameter (dimensionless) and
 m = pore connectivity parameter taken as $1 - \frac{1}{n}$

The Brooks Corey retention characteristic is written as;

$$S_e = \left(\frac{h_d}{h}\right)^\lambda \quad \text{for } h > h_d \quad (2.17)$$

$$S_e = 1 \quad \text{for } 0 \leq h \leq h_d$$

Where h_d = air entry pressure (mm) and
 λ = pore size distribution parameter (dimensionless).

The Campbell function for the retention characteristic comprises two functions presented as;

$$\left(\frac{\theta}{\theta_s}\right) = \left(\frac{h_e}{h}\right)^{1/b} \quad \text{for } h \geq h_i \quad (2.18)$$

Where h_e = Campbell air entry parameter (mm)
 b = Campbell pore size distribution parameter and
 h_i = inflexion point where the equation changes from exponential to quadratic.

$$\theta = \theta_s (1 - ch^2) \quad \text{for } 0 \leq h \leq h_i \quad (2.19)$$

h_i is reportedly related to the pore size distribution parameter b , and air entry parameter h_e , as,

$$h_i = \frac{h_e}{a} \quad (2.20)$$

$$c = \frac{(1-a)}{2h_i} \quad (2.21)$$

$$a = \frac{2b}{1+2b} \quad (2.22)$$

Parameters in the Van Genuchten and Brooks-Corey functions can be obtained by fitting the soil hydraulic characteristic functions to the Van Genuchten and Brooks-Corey functions using optimization models. Similarly parameters in the Brooks Corey model is fitted to data to determine its parameters.

The hydraulic conductivity function is reportedly derived as follows;

$$K(h) = K_s \left[\frac{(1 - (\alpha h)^{n-1} (1 + (\alpha h)^n)^{-m})^2}{(1 + (\alpha h)^n)^{m/2}} \right] \quad (2.23)$$

Where $K(h)$ = unsaturated hydraulic conductivity (mmh^{-1}) and
 K_s = saturated hydraulic conductivity (mmh^{-1})

The above equations use the Van Genuchten retention characteristic substituted into a conductivity model.

Based on the Brooks-Corey retention characteristic applied to a model of hydraulic conductivity, a hydraulic conductivity function is yielded as,

$$K(h) = K_s [S_e]^{2+3\lambda} \quad (2.24)$$

The hydraulic conductivity function used with the Campbell function equation is expressed as

$$K(h) = K_s \left(\frac{\theta}{\theta_s} \right)^{b.n} \quad (2.25)$$

n is fitted directly to the hydraulic conductivity data.

It is reported by Obiero et al. (2003) that the Campbell function may be used to express saturated hydraulic conductivity as:

$$K_s = M \left(\frac{\theta_s}{h_e} \right)^2 \left[\frac{S_i}{b+1} + 2(1-S_i) \right]^2 \quad (2.26)$$

$$S_i = 2b(1+2b) \quad (2.27)$$

M = Constant dependent on liquid properties = 9.81×10^8 to yield K_s in mmh^{-1}

The equation is reported to have been developed by Hutson, 1983 (Lorentz et al., 2001)

Nandagari and Prasad (1995) points out that the Brooks-Corey and Van Genuchten models are the most suitable provided a reliable insitu soil moisture content is available.

The soil moisture characteristic (SMC), which expresses the relationship between matric potential (h) and moisture content (θ) is important in estimating unsaturated hydraulic conductivity. Hydraulic conductivity can be predicted from water retention curve (Touma & Albergel, 1992). Two common methods for measuring SMC are the laboratory pressure plate extraction and the use of insitu paired neutron probe tensiometer measurements as indicated by Nandagari & Prasad (1997) who observed that insitu measurements are more realistic as they represent field conditions better, but involves more time and expenses necessitating the use of texture based models of SMC that rely

on regression approach. The authors further noted that model predictions compare more favourably with laboratory measured SMC than with insitu measured SMC.

Several PTFs used to estimate saturated hydraulic conductivity from basic soil data such as texture, bulk density, porosity etc are summarized by Sobieraj et al. (2001; Table 3), who also used the indicated PTFs for estimating K_s in modeling storm flow. Some of the PTFs reported by the author are tabulated (**Table 2.1**). The tabulated PTF's would yield different values of K_s since they use different sets of data obtained from different sources and regions.

Table 2.1: Pedotransfer functions for estimating saturated hydraulic conductivity

Brakensiek	$K_s = 10\text{Exp}((19.52348\phi) - 8.96847 - (0.028212(CI)) + (0.00018107(Sa)^2) - (0.0094125(CI)^2) - (8.395215\phi^2) + (0.077718(Sa)\phi) - (0.00298(Sa)^2\phi^2) - (0.019492(CI)^2\phi^2) + (0.0000173(Sa)^2(CI)) + (0.02733(CI)^2\phi) + (0.001434(Sa)^2\phi) - (0.0000035(CI)^2(Sa)))$
Campbell and Shiozawa	$K_s = 54\text{exp}(-0.07Si - 0.16CI)$
Jabro	$\text{Log}(K_s)(\text{cm/h}) = 9.56 - 0.81\text{log}Si - 1.09\text{log}(CI) - 4.64(BD)$
Puckett	$K_s(\text{mm/h}) = 156.96\text{exp}(-0.1975CI)$
Dane and Puckett	$K_s(\text{mm/h}) = 303.84\text{exp}(-0.144CI)$
Saxton	$K_s(\text{mm/h}) = 10\text{exp}[12.012 - 0.0755Sa + (-3.895 + 0.03671Sa - 0.1103CI + 0.00087546CI^2/\theta_s)]$ <p>Where $\theta_s = 0.332 - 0.0007251Sa + 0.1276\text{log}_{10}(CI)$</p>

(Adapted from Sobieraj et. al., 2001).

In this study, some of the aforementioned pedotransfer functions were evaluated in establishing the effect of their use in estimating saturated hydraulic conductivity (K_s) as an input parameter in surface flow prediction using the SWAT model. Measured K_s value

was not available from the data base used in obtaining the soil related input parameter to the model. The PTFs selected were those that could be determined using data available from the secondary data base used in the research.

2.4.2 Effective hydraulic conductivity (K_e)

Closely related to saturated hydraulic conductivity is the effective hydraulic conductivity which is a key parameter important as an input to hydrologic models e.g. Soil Water Assessment Tool (SWAT). Its prediction is based on accurate determination of saturated hydraulic conductivity. The effective hydraulic conductivity (K_e) may be approximated as half the saturated hydraulic conductivity (Neistch, 2002 (b)) i.e.

$$K_e = \frac{1}{2} K_s \quad (2.28)$$

Neistch et al. (2002)(b) reports that an equation had been developed to calculate the effective hydraulic conductivity which can be expressed as a function of saturated hydraulic conductivity and curve number. This equation incorporates the effect of land cover impacts into the calculated effective hydraulic conductivity. The equation is expressed as,

$$K_e = \frac{56.82 K_{sat}^{0.286}}{1 + 0.05 \exp(0.62 CN)} - 2 \quad (2.29)$$

where K_{sat} = Saturated hydraulic conductivity (mm/hr)
 CN = Curve Number .

The effective conductivity is evidently a function of saturated hydraulic conductivity whose determination is key to the prediction of this parameter. This study evaluated the role of selected pedotransfer functions in estimating the hydraulic conductivity as an input parameter in the SWAT model. The equations selected were those that could be determined using the soil properties available from the data source used in the study which did not have measured values of K_s and which therefore had to be estimated using a program known as “soil water characteristics” that is based on the use of pedotransfer functions indicated as Saxton2006. The purpose of the evaluation was to help establish if

the output from the calibration and validation of surface flow simulation using the SWAT model is dependent on the pedotransfer function used.

2.4.3 Development of pedotransfer functions for saturated hydraulic conductivity

Pedotransfer functions are predictor functions that relate soil hydraulic characteristics to the basic soil properties. The main reason for developing pedotransfer functions arises from the fact that soil hydraulic characteristics like saturated hydraulic conductivity are difficult to measure accurately and are also tedious and time consuming and expensive to measure making it an expensive venture. In this study, pedotransfer functions were developed to predict saturated hydraulic conductivity (K_s), being a significant input parameter in hydrologic modeling and yet measured data in this property are not usually readily available in many local data bases.

Approaches to predicting saturated hydraulic conductivity, K_s

Two techniques can be used to predict saturated hydraulic conductivity from readily available data on soil properties using pedotransfer functions. Where data on saturated soil hydraulic conductivity is available, K_s can be directly related to the basic soil properties as has been illustrated in the previous section (**Table 2.1**). Saturated hydraulic conductivity can also be indirectly predicted from soil properties using moisture retention characteristics equations like Van Genuchten, Brooks Corey etc. where data on moisture characteristic curves is available. In this case, the parameters associated with moisture retention equations are related to the basic soil properties using pedotransfer functions. Observed values of the parameters are determined by fitting the measured moisture retention characteristics curve to the prediction equation based on the aforesaid moisture retention equation. The latter technique was used in this study owing to the availability of the moisture retention characteristics data and corresponding basic soil data from the international soils data base that was accessible. Readily available data on K_s is rare and characterized with doubt on their accuracy as different methods in different environments yield varied results as observed by Rasoulzadeh (2011) who pointed out that the relative accuracy varies significantly among soil types and field conditions. The author further

states that no single method has ever been developed that can be regarded as being able to perform well in a wide range of circumstances and for all soils, hence the need to use indirect methods from easy to measure soil properties from texture, carbon content etc.

Types and methods of developing pedotransfer functions for K_s

Three types of pedotransfer functions are recognized (Wösten et al., 2001) that be developed. One category predicts soil moisture retention from particle size distribution, bulk density and particle density and requires at least one point of measured hydraulic conductivity characteristic. The other is based on point prediction of water retention characteristic in which regression equations that predict specific points of interest on the moisture retention curves are developed. The method has been used by Fooladmand (2011) to predict points along the moisture retention curve on Iranian soil based on pedotransfer functions using readily available soil properties. Regression equations were developed for nine points along the moisture retention curve. This approach is considered accurate and predicts specific points along the moisture retention curve e.g. $\theta(h)$, at 1500kPa and $\theta(h)$ at -10kPa. The method reportedly has disadvantage in that it requires a large number of regression equations required to quantify complete moisture retention characteristics. Another category involves prediction of parameters used to describe hydraulic characteristics and relies on models that give sufficiently accurate description of K_s . This method predicts parameters in models describing the θ -h-K relationship. The latter method is credited as being more straight forward than point prediction procedure due to the fact that the results are directly applied in simulation models. Hence the approach was used in the first instance in this study. The study however involves prediction of the moisture retention model parameters using the pedotransfer functions that relates simple and easy to measure soil properties to these parameters, a notable gap in the science of development of pedotransfer functions.

In developing pedotransfer functions, statistical regression is commonly used. Most of the available and well established pedotransfer functions for predicting soil hydraulic characteristics from continuous soil properties are based on statistical regression (Vereecken and Herbst, 2004), which is concerned with analysis and construction of

dependence (response) variables like parameters describing moisture retention curve and independent (predictor) variables like soil texture and bulk density. In regression analysis, an attempt is made to obtain the best possible relationship that can be used to estimate the response variable. In predicting the response variable from a number of n predictor variables, x_i the statistical multiple linear regression tool is expressed as (Vereecken and Herbst, 2004):

$$y = a + \sum_{i=1}^n b_i + \varepsilon, i = 1, \dots \dots \dots n \quad (2.30)$$

In which “a” is the intercept, b_i is the regression coefficient and ε is the error. The multiple linear regression technique was used in this study to develop pedotransfer functions that relate parameters in the soil moisture retention equation and basic soil properties.

The three main groups of pedotransfer functions are; class pedotransfer functions, continuous pedotransfer functions, and neural networks some of which have been incorporated into stand alone computer programmes that are able to fit the moisture retention curves to predict moisture retention functions. The purpose is to determine the parameters of the moisture retention characteristic equation. This study involved the use of one such programme to fit observed moisture retention curve to Van Genuchten moisture retention equation to estimate the Van Genuchten model parameters which were then related to basic properties using developed pedotransfer functions.

The procedure to develop a pedotransfer function involves some steps which include;

Assessing normality of distribution of selected response variable

This first step of statistical analysis is intended to test if each parameter distribution may be considered as normal distribution (Walczak et al., 2006). It involves examining distribution of the sample data for a given response variable with a view to obtaining information about transformations that yield distribution more similar to the normal distribution, a requirement in statistical regression. The normality of a distribution is

tested by certain statistics like the Shapiro-Wilk (W-value), Skewness coefficient, Kurtosis, etc. The shape of the distribution is determined by a measure of its Skewness given by the equation,

$$S = \left[\frac{n}{(n-1)(n-2)} \right] \sum_{i=1}^n \frac{(x_i - \bar{x})^3}{s^3} \quad (2.31)$$

Where

S = Skewness coefficient

x_i = observation value

\bar{x} = mean value

n = number of observations

s = standard deviation

For a normally distributed data, the skewness = 0, hence the closer the skewness coefficient is to zero (0), the better is the normality of the distribution. The level of Kurtosis also provides information about the shape of the distribution and the extent of the normality of the distribution and is given by,

$$K = \left[\frac{n(n+1)}{(n-1)(n-2)(n-3)} \right] \sum_{i=1}^n \left[\frac{(x_i - \bar{x})^4}{s^4 - (3n-1)^2 / (n-2)(n-3)} \right] \quad (2.32)$$

The measure of Kurtosis is zero (0) for a normally distributed population. Graphical aids are also used as alternative to check the normality of the distribution of response variables and includes the normality probability plot.

Selection of predictor variables using correlation matrix

It is important to check the mathematical relationship between potential predictor variables with respect to the response variables and by detecting exponential logarithms,

or square root tendencies. This is to enable response variables or their transformations to be related to predictor variable transformations for which the best possible fit can be determined. It is also of utmost importance to eliminate the problem of redundant information in the predictor set brought about by linear dependence between predictor variables (Vereecken and Herbst, 2004). This problem of multicollinearity can be examined by means of correlation matrix to help establish possible correlation between any pair of selected predictor variables. The correlation matrix helps choose predictor variables to be used in the multiple linear regression equation.

Multiple linear regression

After selecting the response variables and corresponding predictor variables, a multiple linear regression analysis is done to determine the appropriate relationship.

Evaluation of the pedotransfer functions

A number of statistical parameters may be used to evaluate the pedotransfer functions. The purpose is to assess correspondence between the predicted value of the response variable estimated using the pedotransfer function with the measured value obtained from a data base. Evaluation of the pedotransfer functions involve testing them for accuracy and reliability in which a response variable estimated using the pedotransfer function is compared with measured value obtained from a data base. The accuracy of an equation is evaluated if the measured values are used to develop the equation. The reliability is evaluated when measured values are different from the ones used to develop the equation. The evaluation statistic used is the same in assessing accuracy or reliability. Hence a dataset is usually split into two parts with one portion being used for testing accuracy while the other for evaluating reliability.

The most common statistics used in evaluating pedotransfer functions include the following (Wösten, et al. 2001):

- (i) Multiple determination coefficient given by

$$R^2 = 1 - \frac{\sum_1^N (y_i - \hat{y}_i)^2}{\sum_1^N (y_i - \bar{y})^2} \quad (2.33)$$

(ii) Root mean square error given by

$$RMSE = \sqrt{\frac{\sum_1^N (y_i - \hat{y})^2}{N}} \quad (2.34)$$

(iii) Mean Error given by

$$ME = \sum_1^N \frac{(y_i - \hat{y})}{N} \quad (2.35)$$

(iv) Mean Absolute Error given by

$$ME = \sum_1^N \frac{|y_i - \hat{y}_i|}{N} \quad (2.36)$$

Where y_i denotes the actual value, \hat{y}_i the predicted value, and \bar{y} the average of the actual value and N is the total number of observations.

Tomasellah and Hodnet (2004), evaluated performance of different pedotransfer functions for estimating certain soil hydraulic properties that include available water capacity (AWC) using R^2 , RMSE and ME. Schaap and Leij (1998) used only RMSE in assessing pedotransfer functions developed for predicting hydraulic conductivity and water retention based on the use of neural networks.

2.5. Theoretical equations used in developing pedotransfer functions for Ks

The pedotransfer functions for saturated hydraulic conductivity developed in this study were based on the use of Van Genuchten water retention equation. The equation relates moisture content (θ) to the soil matric suction (h) (section 2.4.1). Based on this equation, the matric pressure head expressed as a function of soil moisture content is written as follows when h is made subject of the formula:

$$h(\theta) = \frac{1}{\alpha} [S_e^{n-1} - 1]^{\frac{1}{n}} \quad (2.37)$$

where $S_e = \left[\frac{\theta - \theta_r}{\theta_s - \theta_r} \right]$

The above equation enables one to calculate h for a given value of θ . To determine saturated hydraulic conductivity, K_s , the hydraulic conductivity function is used and is expressed as (section 2.4.1),

$$K(h) = K_s \left[\frac{(1 - (\alpha h)^{n-1} (1 + (\alpha h)^n)^{-m})^2}{(1 + (\alpha h)^n)^{m/2}} \right] \quad (2.38)$$

In order to determine K_s using the above equation, a relationship between hydraulic conductivity and matric suction, h is required. Such an equation can be obtained through equations that relate hydraulic conductivity to permeability (Obiero, 1996). The equation is expressed as:

$$K = \frac{\rho g k}{\eta} \quad (2.39)$$

Where K = hydraulic conductivity (m/s)

ρ = density of water (kg/m^3)

g = acceleration due to gravity (m/s^2)

k = permeability (m^2)

η = viscosity of water ($\text{kg m}^{-1} \text{s}^{-1}$)

Intrinsic permeability k is related to porosity according to the equation

$$k = \frac{\varepsilon r^2}{8} \quad (2.40)$$

in which ε is the porosity and r the pore radius (Marshall and Holmes, 1988). The effective pore radius can be estimated from a simplified equation relating it to the soil matric suction expressed as;

$$s = \frac{2\gamma}{\rho g r_t} \quad (2.41)$$

Where s is the suction, γ is the surface tension of water and r_t the effective radius of the pores. Based on the above equations, it is thus possible to calculate hydraulic conductivity from the matric suction hence $K(h)$ can be computed. Knowing the values of α , n , θ_s , θ_r , it is possible to determine $h(\theta)$ from the moisture retention equation of Van

Genuchten if the soil moisture content is known and by extension, terms in the aforesaid hydraulic conductivity function can be determined, thereby enabling the saturated hydraulic conductivity K_s to be determined. Data on soil moisture characteristic curve is not usually readily available in many national soil data bases to enable determination of the Van Genuchten moisture retention parameters. Its direct measurement is also costly and time consuming (Fooladmand, 2011). There is need to predict these parameters from basic soil properties like bulk density, organic carbon and texture is essential, an area of study that has not received much attention in the recent past. Development of pedotransfer functions that relate these parameters to basic soil properties would ease the determination of K_s based on moisture retention equations thereby contributing to the knowledge base in determining soil hydraulic properties. The study used this approach in determination of the moisture retention equation parameters to determine K_s as an input parameter in a watershed scale model for surface flow simulation.

Based on the PTFs indicated earlier (**Table 2.1**), the list of basic soil parameters used in the pedotransfer functions as summarized by Sobieraj et al., (2001; Table 3) are highlighted below (**Table 2.2**).

Table 2.2 : Soil parameters used in pedotransfer functions (Based on Sobieraj et al., 2001).

Pedotransfer Function	Soil Parameters
Brakensiek	% sand, % silt, % clay, bulk density (g/cm^3)
Campbell and Shiozawa	% silt, % clay
Jabro	% silt, % clay, bulk density (g/cm^3)
Puckett	% clay
Dane and Puckett	% clay
Saxton	% sand, % clay,
Young and Gowing (1996)	% sand, % clay, bulk density(g/cm^3), %CEC, %Organic matter content .

From **Table 2.2**, the basic soil parameters (soil properties) that affect the saturated hydraulic conductivity include sand, silt, clay, organic carbon, cation exchange capacity and bulk density. The properties used in this study to develop pedotransfer functions were

bulk density, sand content, silt content, clay content and organic carbon being physical soil properties common in most soil data bases. These properties are also most common for use in the development of pedotransfer functions. The actual properties, used in predicting the moisture retention parameters in this study depended on results from statistical regression that identified the strength of their correlations with the selected transformations of the response variable and also in minimizing multicollinearity among the predictor variables.

2.6 The SWAT Model Review

2.6.1 Brief Description of the SWAT Model

SWAT model is a process based, continuous physically based distributed parameter river basin model that simulates water, sediment and pollutant yields developed in the early 1990's to assist water resources managers assess impact of land use management on water, and diffuse pollution for large ungauged catchments with different soil types, land use and management practices (Levesque et al., 2008). Model components include weather, hydrology, erosion, soil, temperature, plant growth, nutrients, pesticides, land management, channel and reservoir routing (Rostamian et al., 2008). The first step in creating a SWAT model involves delineation of the sub-watersheds in the basin each of which is treated as an individual unit. The sub basins are further divided into hydrologic response units (HRU's). These units are composed of homogeneous land use, soil characteristics and management practices. Relevant hydrologic components like surface runoff, ground water flow and sediment yield are estimated for each HRU unit. Two methods are used for surface runoff estimation in SWAT *i.e.* the SCS curve number and Green-Ampt infiltration. This study is based on the use of curve number for surface runoff and hence stream flow simulation. A SWAT model can be built using the Arc-View interface called AVSWAT which provides suitable means to enter data into the SWAT code.

2.6.2 SWAT Model Origin and Applications

The historical development and applications of the SWAT model is well documented in Gassman et al.(2007) in which it is reported that early origins of SWAT is traced to models previously developed by the United States Department of Agriculture, Agricultural Research Service (USDA-ARS) models that included the Chemicals, Runoff and Erosion from Agricultural Management Systems (CREAMS) model, Ground Water Loading Effects on Agricultural Management Systems (GLEAMS) model and the Environmental Impact Policy Climate (EPIC) model originally called Erosion Productivity Impact Calculator. The authors further note that the current SWAT model evolved from the Simulator for Water Resources in Rural Basins (SWRRB) model whose development commenced in the early 1980's and through modifications that incorporated inputs from other models, the SWAT model finally developed when SWRRB was eventually merged with Routing Outputs Outlet (ROTO) model to overcome their limitations. Since its creation in the early 1990's, the model has undergone continuous review and expansion of its capabilities. The model has been applied worldwide for purposes that include simulation of sediment flow (Ndomba, 2010), modeling hydrologic balance (Setegen et al., 2008), Evaluation of the impact of land use and land cover changes on the hydrology of catchments (Odira et al., 2010). It has also been used to assess the effect of certain interventions on river and sediment flows (Tripathi et al., 2005). The model has registered good performance and also limited success. Limited success has been reported in SWAT simulation for stream flow in South African catchments (Govender and Everson, 2005). Over estimation of flows between 1 and 3 mm was reported while flows between 4 and 7 mm were overestimated. The model performance was notably better in the dry than in the wet years. Discrepancies between the observed and predicted flow for the two catchments considered was attributed to their small drainage basins. The model was developed to simulate large catchments, a limitation which may affect model performance.

3. METHODOLOGY

3.1 The Study Catchment

3.1.1 Description of the study area.

The study area is Naro Moru river catchment. The catchment covers an area of 172 km². The catchment lies at the North Western slopes of Mt. Kenya. The river originates from the peak of Mount Kenya and is tributary to the Ewaso Ng'iro River. The catchment lies between latitudes 0° 03' and 0° 11' South and longitudes 36° 55' and 37° 15' East. The altitude of the Naro Moru catchment ranges from 5200m at the peak of the mountain to 1800m above mean sea level at its confluence with Ewaso Ng'iro river. The catchment lies on the leeward side of Mt Kenya and therefore is characterized by low amount of rainfall as presented by Ngigi (2006) who also reported that the mean annual rainfall within the catchment increases from 650mm at the outlet to 1500mm at 3300m altitude and drops to 500mm in the moorland. On average the annual potential evaporation is above 2500mm. The climatic conditions that prevail in the catchment and Agro-ecological zones are documented by Thomas et al. (1993) varying from the glaciated peaks of Mount Kenya (5200m) to the semi- arid Laikipia plateau (1800m) above mean sea level. The catchment has five different ecological zones being peak, moorland, forest, foot zone and savannah and so has diversity of vegetation/land use and soil types. Location of the study area in Kenya is shown in **Figure 3.1**. The drainage basin has several river gauging stations from the top of Mount Kenya to the point where the river joins the Ewaso N'giro river. It is reported by Gathenya et al. (1993) that these stations were installed in 1982 and had been maintained by the Laikipia Research Programme since then. Some of these are shown in **Figure 3.2** alongside the river drainage network. The Kenya Meteorological Department and the Ministry of Water and Irrigation, Kenya also has collected weather data and river flow data respectively on some gauging stations in the catchment. The land-use types and major soil types based on the FAO-UNESCO system of classification of 1990 and predominant in the catchment are shown on **Figures 3.3 and 3.4** respectively.

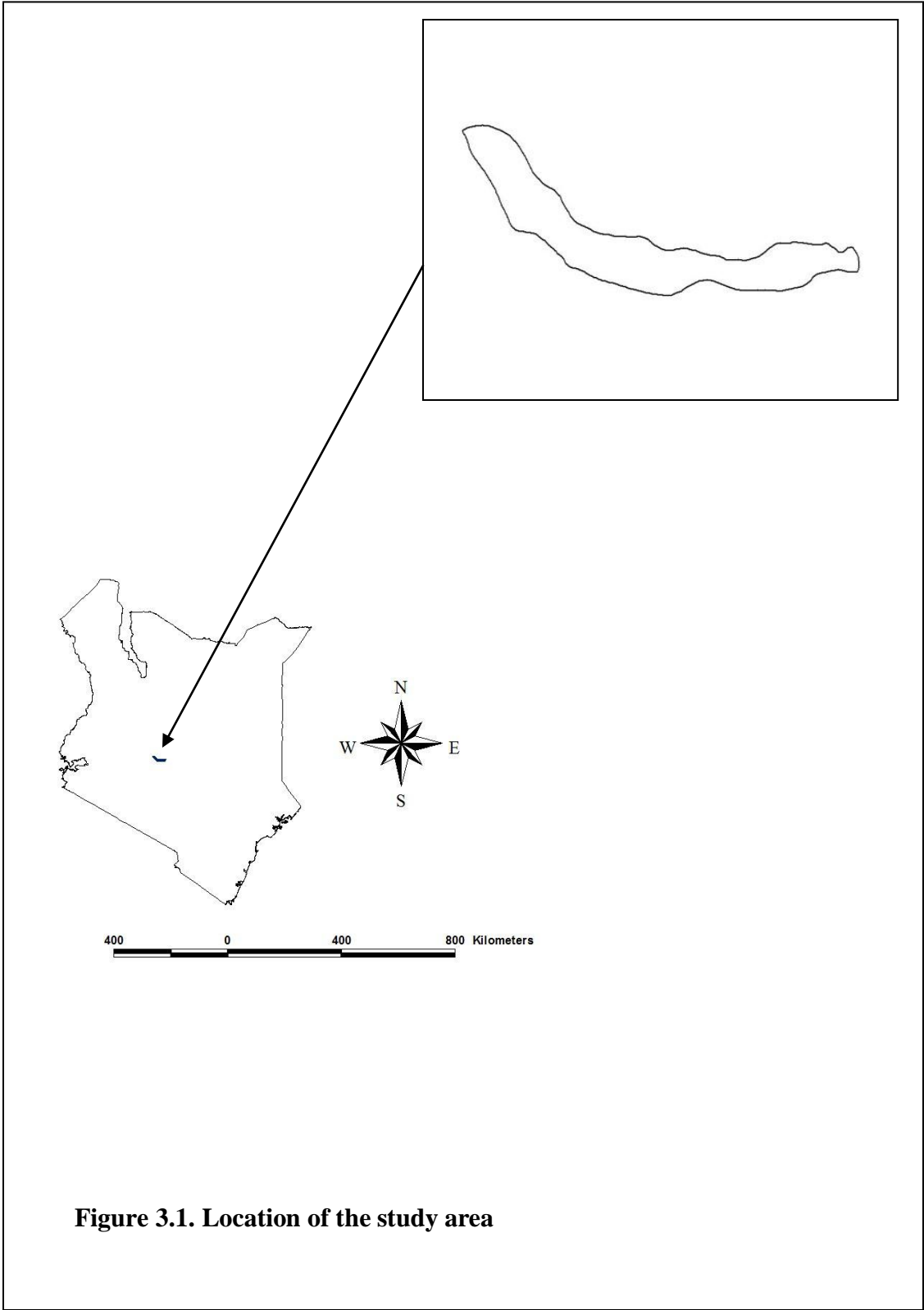


Figure 3.1. Location of the study area

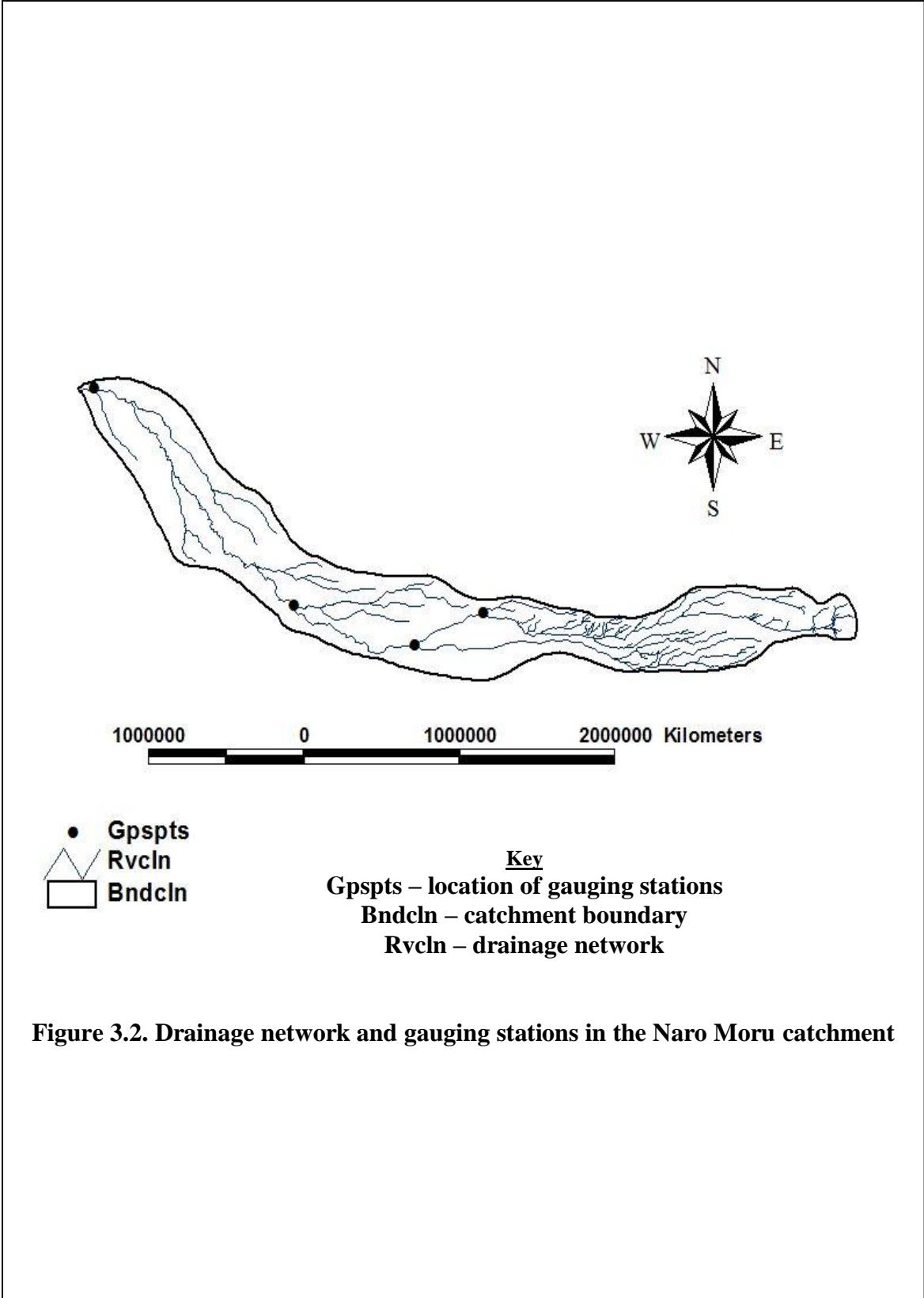
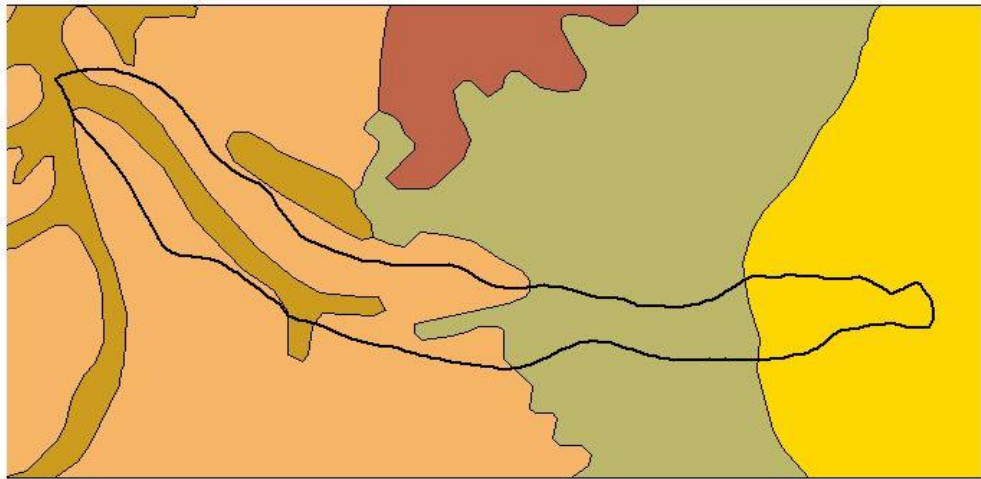








Figure 3.2. Drainage network and gauging stations in the Naro Moru catchment

1000000 0 1000000 2000000 Kilometers



 Catchment boundary

Land use types

-  Agriculture sparse
-  Barren land
-  Forest
-  Plantation
-  Town
-  Woodland

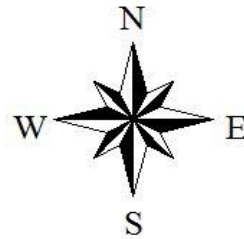
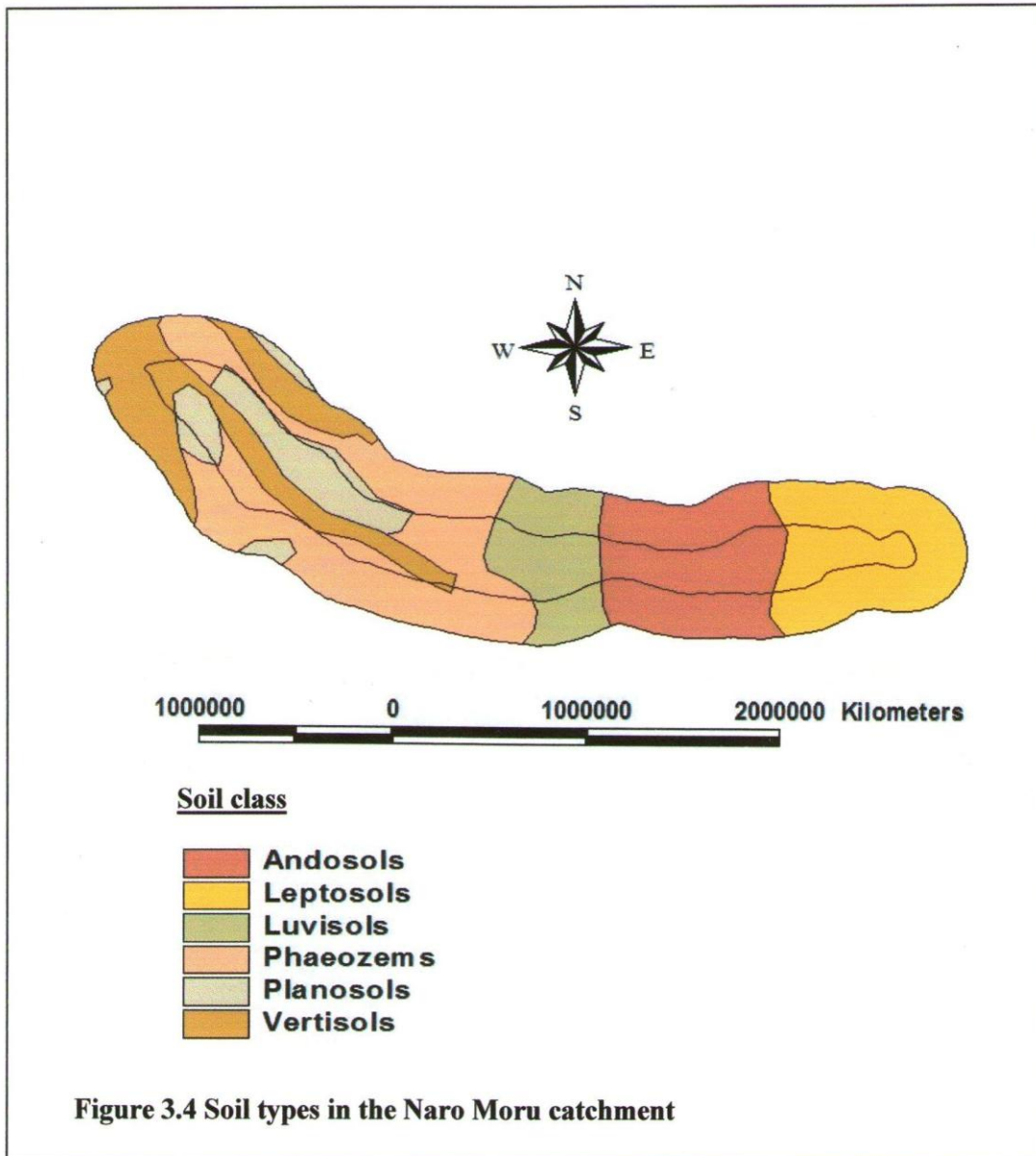


Figure 3.3. Land use types covering the study area



3.1.2 Available Data

A summary of raw data found to be available for the above described test catchment is presented as shown in **Table 3.1**. This data base consisted of hydro meteorological data and other information relevant for the purposes of this study.

Table 3.1: Summary of raw data available for the catchment under investigation

<u>Data Type</u>	<u>Period</u>	<u>Comment</u>
<i>1. Rainfall</i>		
(i) Hourly rainfall	1995-2001	Available on selected dates
(ii) Daily rainfall	1992-2000	Continuous daily data from stations in catchment vicinity.
(iii) Weather data	1992-2001	Continuous daily data on temp., wind speed, etc
<i>2. Stream flow</i>		
(i) Daily flow	1985-2000	Available from station 5BC2 from Ministry of water development.
(ii) Gauge heights	1995-2001	Hourly records of 4-9hrs each day.
<i>2. Topographic maps</i>		
	-	Scale 1:50,000
<i>3. Aerial photographs</i>		
	1988	scale 1:70,000

3.2 Data Acquisition and Preliminary Analysis.

3.2.1 Catchment Characterization (Topographical maps acquisition and processing)

Topographic maps at the scale of 1:50,000 were acquired from Survey of Kenya, Ruaraka, Nairobi, Kenya. Approval from the Department of Defence, Kenya was a required to collect the data. Three topographical sheets adequately captured the area of research interest. This data was utilized in digitizing and clipping of the various essential features and characteristics of the study area. Using the topographical sheets, the Kenya

Soil Survey department, Nairobi, Kenya, attempted to manually delineate the catchment boundary and thereafter digitized such features as contours, river networks, roads, sub catchment boundaries, markets centres, schools etc. and these were presented as themes/layers in the Arc-View GIS software. Similarly, important catchment characteristics that influence the rainfall-runoff processes were digitized and clipped to provide these essential characteristics for the Naro Moru river catchment. These included soil types, rooting depths, Agro ecological zones, land-use types, slope classes, etc. These were likewise made available in soft copy in Arc-View so as to feature as themes/layers that can be shown separately or overlaid as required in an analysis. Data on land cover types, an important parameter required as an input to relevant watershed models was downloaded from the FAO Africover data base after permission was granted from the relevant authorities. This data, available for the whole country, was considered together with other data sources for the study area in preparation for relevant analysis.

3.2.2 Photogrammetric Mapping (Aerial photo acquisition and processing)

Another set of data acquired for this research were aerial photographs that adequately covered the study area. The relevant air photographs were identified with the assistance of the Survey of Kenya personnel using the transparent diapositive overlaid on a topographical map and showing the flight paths. The identified photographs were purchased from the Survey of Kenya. Processing of the aerial photographs through photogrammetric mapping was carried out at the Department of Surveying, University of Nairobi. Processing of the photographs involved an elaborate procedure of mapping to plot the contours at a shorter interval (higher resolution) for use in catchment delineation after digitization. The river networks were also plotted and so were other features e.g. roads, town centres, railway lines etc. The detailed mapping procedure is appended (Appendix 1). The Digital Elevation Model (DEM) was then derived from the digitized contours since readily available data on DEM was not available.

3.2.3 Digitizing of the contours and drainage networks.

The process of digitizing began initially with registration in which the topographical sheet is mounted onto the digitizing plate and four georeferenced points lying on a

rectangular grid marked. These are designated 1, 2, 3, and 4 in a clockwise mode. Using Arc-view software with the appropriate extension, these points with their coordinates are registered on the screen. Other details were also recorded and checks made accordingly until registration is achieved within the limits of acceptable error. The process of digitizing is then set to commence. Alternatively the mapped topographic sheets are scanned to obtain a raster image which is then retrieved on the screen in Arc-view or other appropriate software and then followed by screen digitizing of the contour and other desired features and river networks. The ultimate procedure chosen for digitizing the test catchment as discussed by Siriba (2004) is appended (Appendix 2) and summarized in **Figure 3.5**

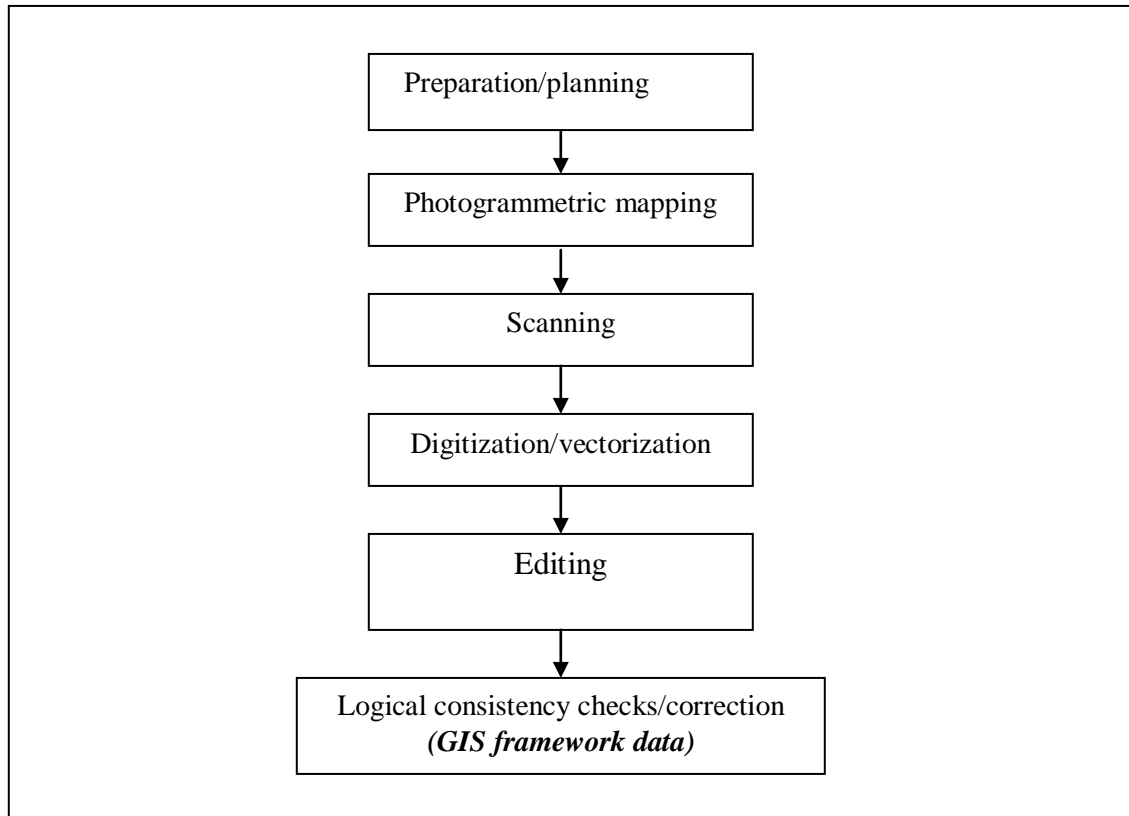


Figure 3.5. Flow diagram showing sequence of events followed to obtain digital map of the test catchment

The main aim of digitization of the catchment was to produce a digital map of contours and stream networks for use in preparing input data for hydrological modeling. The

procedure for input data preparation is described in more detail in the subsequent sections.

3.2.4 Hydro meteorological data acquisition and processing

Rainfall data acquisition.

A list of all the rainfall stations reportedly located in the Ewaso Ng'iro River Basin and within which the study area lies and their geographic coordinates was provided by the Meteorological department. Details of the rainfall station names, geographic coordinates, elevations, etc, were recorded in excel software and retrieved into arc-view using an appropriate format so that a theme for the stations was created to feature as a layer in arc-view. The list of rainfall stations identified to be inside or in the vicinity of the test catchment are shown in **Table 3.2**. Only two stations were observed to lie within the catchment. These were the Naro Moru gate and Meteorological Station Lodge. The data reportedly available for these stations is daily rainfall. A third station used together with the above mentioned two stations to obtain daily rainfall data was the Sirimon Gate station also near to the catchment. Out of the stations provided by the meteorological department, only two stations were reported to be full meteorological stations from which all the meteorological variables could be obtained apart from daily rainfall. These are the Laikipia Air Base (Nanyuki) and the Nyeri Meteorological Station. These stations are the ones that were used for obtaining such information as solar radiation, temperature, wind speed, etc. required as input in the model intended for this research work. Only the Nyeri meteorological station had continuous data over the nine year period during when the analysis was done and so was used in the analysis. The stations are show in **Figure 3.6**.

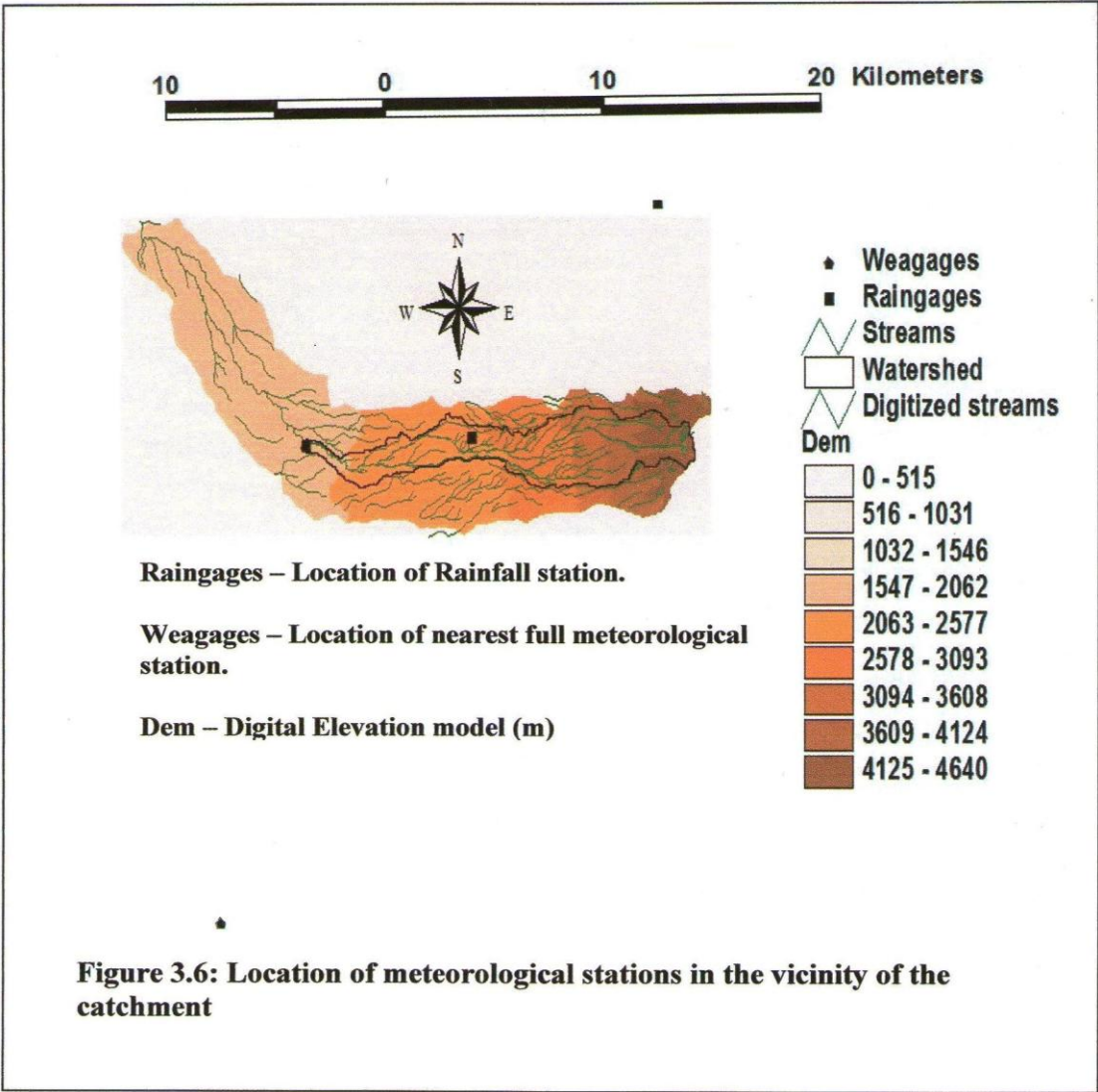


Table 3.2: List of rainfall stations inside and in the vicinity of the test catchment

Station Name	Latitude	Longitude	Altitude
Laikipia Air base (Nanyuki)	0 03N	37 02E	6200FT
Sweet Waters Tented Camp	0 00S	36 56E	1780FT
Karemeno School of Agriculture	0 08S	36 46E	6800FT
Ngobit Suguroi Estate	0 2S	36 39E	6666FT
Lamuria Met Site	0 08S	36 52E	6100FT
UasoNyiro Lawrence Mwangasi Farm	0 17S	36 51E	6800FT
Riunge Hill	0 24S	36 43E	10400FT
Kurase Hill Aberdare Park	0 20S	36 40E	11200FT
Ark Gate Aberdare Park	0 20S	36 51E	7000FT
Ngobit Police Post	0 3S	36 47E	6500FT
Ol Pejeta Ranch Ltd Loideni	0 2S	36 52E	6000FT
Mugunda Primary School	0 10S	36 42E	7500FT
Ndunyu Gwathi Market	0 16S	36 51E	6800FT
Rhino Gate National Park	0 20S	36 36E	8600FT
Matanya LRP Station	0 02S	36 04E	1800FT
Watuka Met Station	0 16S	36 45E	2400FT
Kangaita Forest Station	0 0S	37 9E	6800FT
Gathiru Forest Station	0 5S	37 7E	7500FT
Burguret Forest Guard Post	0 06S	37 02E	6400FT
Naro Moru Gate	0 10S	37 9E	7200FT
Sirimon Gate, Mt Kenya Park	0 2S	37 17E	8500FT
Meteorological Station Lodge	0 10S	37 2E	10,000FT
Ol Pejeta Sweet Water Dip	0 2N	36 58E	6000FT
Ontulili Forest Station	0 01N	37 11E	7400FT
Nanyuki WDD	0 01N	37 05E	6300FT
Nanyuki Ministry of Water Development	0 01N	37 05E	6500E

Acquisition of River flow data

The above data were acquired from the NRM³ data base at Nanyuki, Kenya. The data was in the form of daily stream gauge heights recorded for the period 1995-2000. This provided the stage graphs for the aforementioned period. The stage graph data was collected for each of the gauging stations at the main outlet and of the sub catchments within the Naro-Moru river basin. These were designated as A3, A4 and A5 and A6. Daily river flow data recorded in m³/s were also collected for the four gauging stations. The river flow data was also acquired for the station at the outlet of the catchment designated 5BC2 from the Ministry of Water Development, Nairobi, Kenya for the period 1985-1998 for which the data is available. This station corresponded to that of A5 and was the one used in the analysis of model evaluation. Together with the river flow data, rating equations to convert the stage graphs to discharge were also acquired from the NRM³ data base. These were as follows for the Naro Moru river basin.

<u>Stage</u>	<u>Station</u>	<u>Rating Eqn</u>	<u>Error</u>	<u>Period</u>	<u>Application</u>
H>0	A1	$Q=0.8239(H-0.058)^{1.0181}$	0.004698	27/7/82-11/1/91	1981 to date
H>0	A2	$Q=1.6304(H-0.078)^{4.863}$	0.03965	23/11/81-24/11/94	1981 to date
H>0	A3	$Q=18.1176(H+0.079)^{3.6485}$	0.07838	25/5/81-18/10/95	1981 to date
H>0	A4	$Q=6.1291(H-0.05)^{2.6446}$	0.05105	8/11/84-18/10/95	1981 to date
H>0	A5	$Q=15.973(H+0.012)^{1.762}$ $Q=36.4200(H-0.0)^{2.374}$	0.06869 -	30/4/48-12/11/82 18/5/83-8/4/94	30/4/48-18/4/83 26/4/83-17/5/95
0.1≤H≤0.3		$Q=19.8976(H-0.033)^{1.8001}$	-		
H>0.3		$Q=14.3615(H-0.07)^{1.4035}$			
H≥0		$Q=8.1040(H-0.039)^{2.5616}$	0.03125	5/10/95-28/10/95	18/5/95 to date
H≥0	A6	$Q=2.3292(H+0.076)^{1.6767}$	0.15707	3/3/82-3/11/95	3/3/82 to date

3.3 Data Analysis

3.3.1. Model Selection

The Soil Water Assessment Tool (SWAT) was chosen for hydrological modeling in the watershed under study using the Arc-view SWAT (AVSWAT2003). SWAT is a watershed scale model developed to predict the impact of land management practices on water, sediment and agricultural chemical yields with varying soils, land use and management conditions over long periods of time (Neistch et al., 2005). One basis for model selection was due to its worldwide use for variety of applications. The model has in the recent past gained significant publicity having been used widely for various applications world over with notable success (Ndomba and Birhanu, 2008) with recent applications in the Nilotic catchments that include Kenya, Tanzania, Ethiopia, Uganda, among others. SWAT has gained international acceptance as a robust interdisciplinary watershed modeling tool as evidenced by international SWAT conferences, hundreds of SWAT related papers presented at numerous scientific meetings, and many articles published in peer reviewed journals. The model has been used for a wide range of applications for reasons that include its computational efficiency and flexibility on input data requirements (Stehr et al., 2008). The available data for the catchment under study could be used in hydrological modeling using SWAT. SWAT is capable of modeling changes in land use and management practices, can model variety of catchment areas ranging from a few hectares to thousands of square kilometers and performs long term simulations. Besides, the model is freely available and can be downloaded from the internet. The model website has a well developed system for support to model users. The model is in the public domain and therefore available without many restrictions. The model has options for daily, monthly and yearly time step simulations that can be carried out without altering the input data. Model predictions are spatially distributed thereby providing spatial information regarding upstream sources of modeled quantities (Andualem and Yonas, 2008).

3.3.2 Data Preparation for Hydrological Modelling Using SWAT

Summary of model input data.

The input parameters required to run the model included;

- Daily precipitation
- Digital elevation Model (DEM)
- Weather data (Solar Radiation).
- Soils information.
- Land use data.
- Drainage data (optional).

Digital elevation model (DEM)

The DEM was created from the contours previously digitized and converted to a shape file and covering the area under study. The DEM created, based on the digitized contours had a resolution of 55 in metres. **Figures 3.7 and 3.8** shows the digitized contour network and the corresponding DEM derived from it.

Land use input data preparation

The land use data was extracted for the rectangular grid covering the selected study area. The shape file land use map was compiled from the Kenya Soil Survey. The SWAT model requires that these land use types be re-classified to the corresponding SWAT land uses for the model to successfully load the land use classes that the model recognizes. Hence, for each of the land uses observed in the study area, a corresponding SWAT land use was identified and used in the reclassification. A land use look up table (dbf) was thus prepared showing the SWAT land use types and corresponding abbreviations. The land use look up table (Table 3.4) is also an essential input to the SWAT model which links the table to the land use map. The land uses and abbreviations are shown in **Table 3.3**. Then numbers in the value column in **Table 3.4** above correspond to the numbers in the field named LUNUM of the land use attribute table that correspond to the indicated land use types that have been re-classified into SWAT land use.

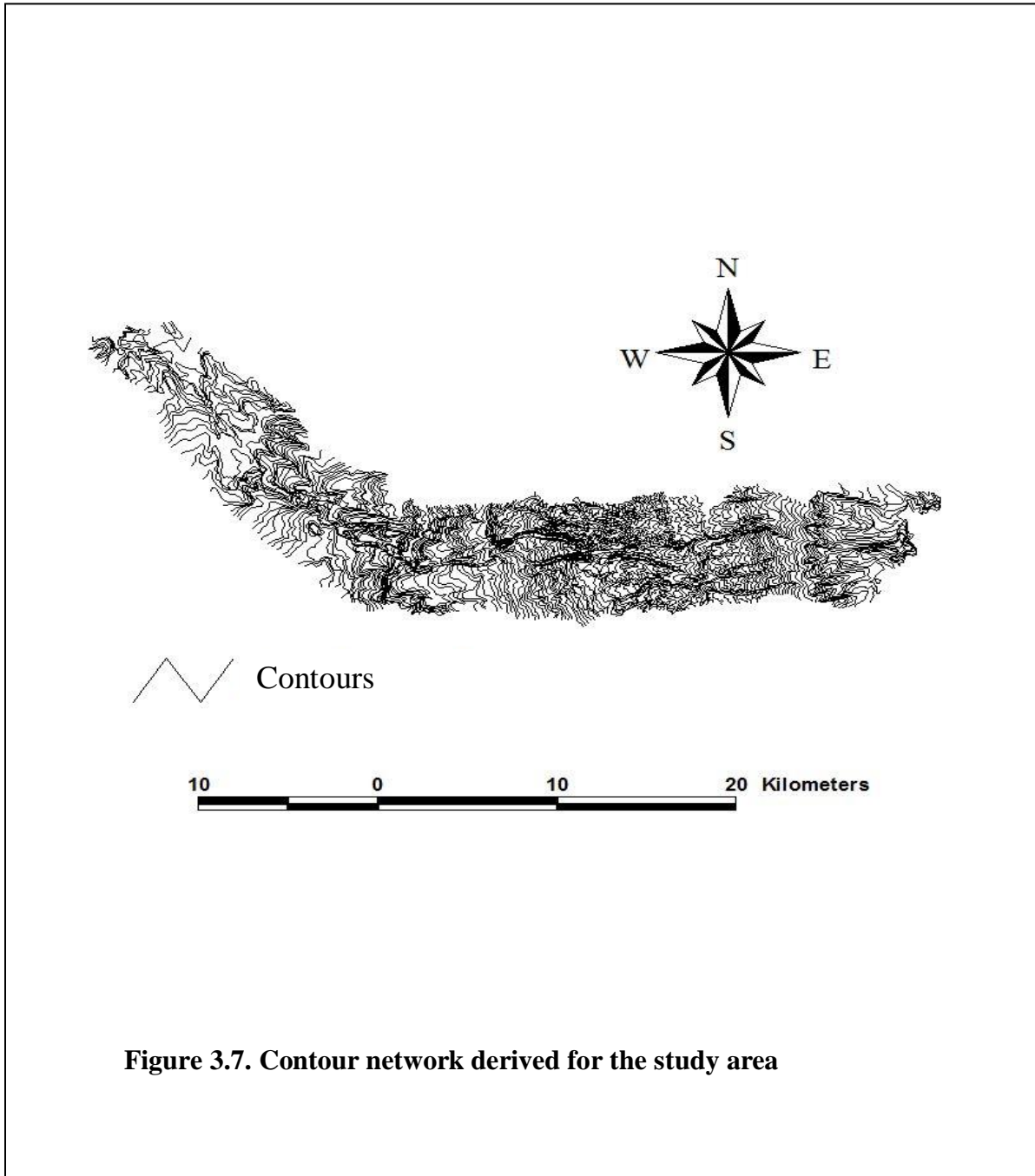
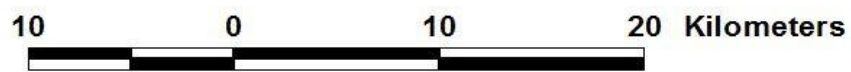
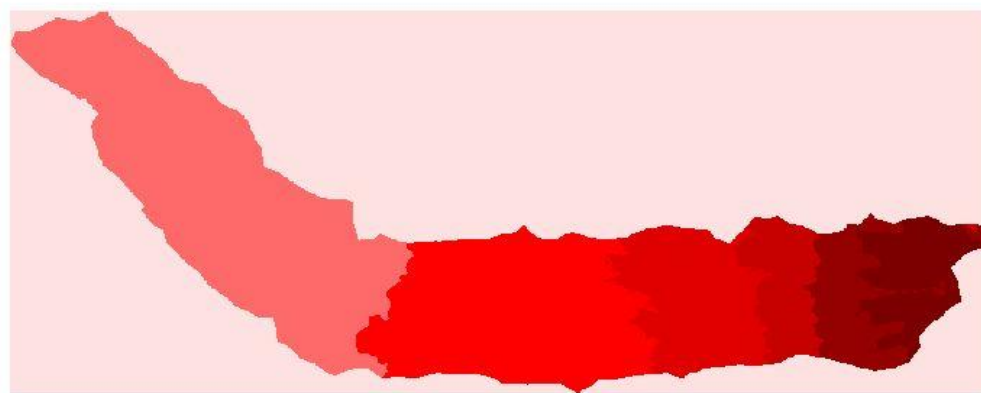


Figure 3.7. Contour network derived for the study area



Digital Elevation Model (m)

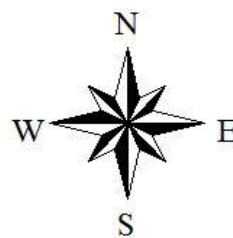
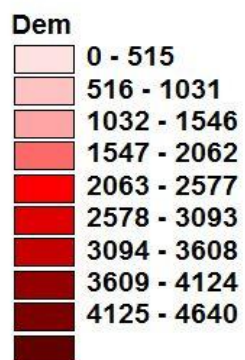


Figure 3.8 Digital elevation model derived from contour network

Table 3.3: Reclassification of observed land uses to SWAT land types

Kenya Land Use	SWAT land use	SWAT land use code
Plantation	Mixed forest Land	FRST
Agriculture (sparse)	Cropland and pasture	AGRL
Woodland	Evergreen Forest Land	FRSE
Forest	Deciduous Forest Land	FRSD
Barren Land	Strip Mines	SWRN

Table 3.4: Land use look up table

Value	Land use (SWAT)
1	FRST
2	AGRL
3	FRSE
4	FRSD
5	SWRN

Soils input data preparation

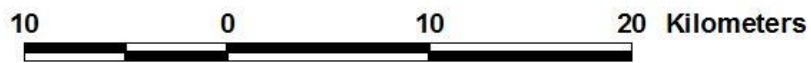
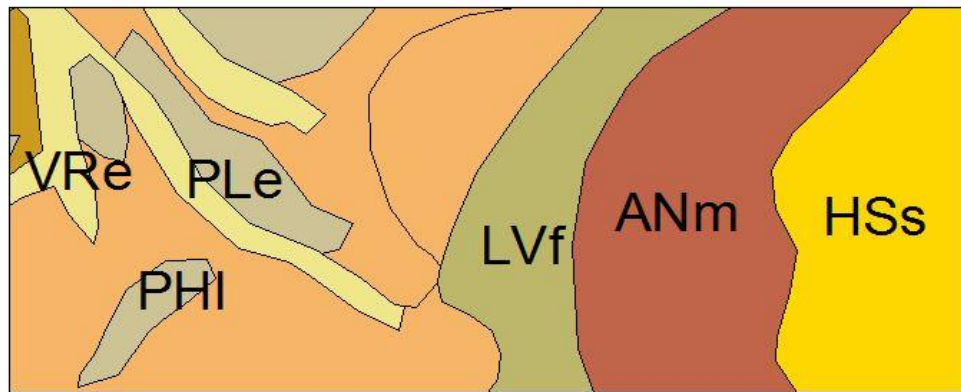
The input data for soil included the shape file soil map extracted from the soil of Kenya obtained from the data base of the Kenya Soil Survey (KSS). The map provided information on the major soil types classified based on the FAO classification and some of the soil attributes e.g. proportion of main soil type in the soil unit. For each of the soil mapping units occurring in the study area, the soil physical and chemical properties relevant as input data to SWAT were determined from the corresponding soil unit identified from the table of the soil properties (KENSOTER table). These properties included the proportions of sand, silt, clay and coarse fragments (i.e. % sand, %clay, %silt), bulk density, Cation exchange Capacity (CEC), Electrical Conductivity, Total Carbon, etc. The soil properties were then used in editing the SWAT data base accessed through the arc-view interface AVSWATX. The soil units were added as new soil unit in the data base. Some of the soil properties required as input to the soils database could not be obtained directly from the table of soil properties and included saturated hydraulic conductivity (K_{sat}) and Soil Erodibility (USLE_K). To determine K_{sat} , the programme known as Soil Water Characteristics was used. This software predicts K_{sat} and other soil properties using information on soil texture and organic matter content. Soil texture information was obtained from the table of properties in the KENSOTER soils data base. The organic matter content (O.M) was determined from its relationship with Total Carbon (C) shown in the equation below (Neistch, et al., 2002(a)) .

$$O.M=1.72C$$

Where O. M. = organic matter content (%)

C = organic carbon content (%).

The level of compaction was assumed normal needed also as input to the Soil Water Characteristics programme. **Figure 3.9** shows the major soil units found in the study area. The soil properties of the units are shown in **Table 3.5** indicating the relevant soil input information for SWAT. The USDA texture based nomographs were then used to estimate the soil erodibility (USLE_K). The nomographs required information on texture, O.M. soil structure and permeability. The structure and permeability were predicted from the soil texture. The use of KENSOTER soils data base to estimate the soil properties is well documented in Bartes and Gicheru (2004). A user soil data base in the form of a dbf table



Soil units

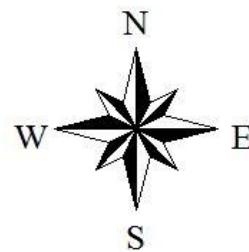
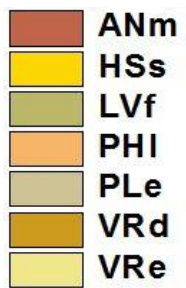


Figure 3.9 Major soil units in the study area

(excel) was also prepared to assist in re-classifying the user soils to be consistent with the SWAT classification. The soils look up table is shown in **Table 3.6**. The table is linked to the arc-view shape file of soil map through the attribute table of the shape file in which the fields (columns) similar to those in the soils look up table is added.

Table 3.5 Soil properties for the major soil units in the study catchment.

soil	%sand	%silt	%clay	Bulk density g/cm ³	CEC %	Ksat mm/h	ELCO mmhos/cm	TAWC %	TOTC g/kg	USLE_K
HSs	30	56	14	0.36	15	7.8	0	35	80.0	0.28
LVf	26	22	52	1.48	18	0.03	0	12.3	9.3	0.15
VRe	30	30	40	1.49	40	0.11	0	12	15	0.17
PHI	24	17	59	1.10	14	0.02	0	11	14.8	0.19
ANm	59	20	21	1.13	33	0.84	0	17	23.5	0.05

Table 3.6: Soil look up table

VALUE	NAME
1	VRE
2	PHI
3	ANM
4	LVF
5	HSS
6	PLE

Weather data input preparation

The location of the meteorological station with the weather data based on its UTM co-ordinates and elevation was required. The X-coordinate (Easting) and Y-coordinate (Northing) based on the Universal Transfer Mercator (UTM) co-ordinate system for the Nyeri meteorological station which was chosen to provide weather input data were 273695 and 992078 respectively. The station elevation was 1817m above sea level. The weather input variables required included; Solar radiation, wind speed, relative humidity, precipitation, dew point temperature, minimum and maximum temperatures. Based on the data available at the Kenya Meteorological Station, it was possible to acquire the data for all the above weather variables for 9 years daily data during the period 1992-2000 based on three rainfall stations in the vicinity of the study catchment. Data for precipitation, however was obtained from three (3) rainfall stations in the vicinity of the study catchment. Two of the stations lie inside the catchment while one is outside. The co-ordinates and elevation of the stations were then included in the precipitation station location and input data tables.

The weather generator

The weather generator provides input data to SWAT being parameters derived from the weather information. The weather generator data was derived from the nine years daily data (1992-2000) on rainfall, minimum temperature, maximum temperature, relative humidity, windspeed, solar radiation and dewpoint temperature. The data output included the following for precipitation:

- Average monthly precipitation (PCP_MM).
- Standard deviation for precipitation (PCPSTD).
- Skew coefficient (PCPSKW).
- Probability of a wet day following a dry day (PR_W1).
- Probability of a wet day following a wet day (PR_W2)
- Average number of days of precipitation in month (PCPD).

The data was provided on a monthly basis for each month. The same information as indicated in the aforementioned list for precipitation was also determined for the other variables i.e temperature, solar radiation, etc. To determine the indicated weather input

data, a programme (pcpSTAT) was used to compute the weather input data from the daily data of the weather variables. The programme required the following input information to determine the weather generator data.

- Input file name i.e. file containing all the daily data for the period in question (text file) arranged as a column.
- The number of years of record.
- The year of commencement of data which has to be 1st January in this case 1/1/1992.
- Output file name.

Once the above information is provided as input to the programme, as required for each weather variable, the output will be in excel file format containing all the weather generator input data required for running the SWAT model. **Table 3.7** shows the output data in the tabular form showing the input data for SWAT weather generator for each of the months of the year computed for precipitation. A similar table was generated from the programme for solar radiation, wind speed, etc. From the SWAT arc-view interface, the SWAT data base was then edited through addition of weather station containing the weather generator parameters i.e. Nyeri Meteorological Station. Other information required in the template for weather input data included the co-ordinates and elevation of the chosen weather station and then the weather generator data (**Table 3.7**).

Drainage Data

Drainage data input into SWAT was provided in the form of digitized stream network. The digitized stream network was made available as shape file. The stream network used as input to SWAT together with the DEM was used in the catchment delineation using a selected watershed outlet. The location of gauging stations also available as shape file were used in the location of the watershed outlets.

Table 3.7: Statistical analysis of daily precipitation data (1992-2000)

Statistical Analysis of Daily Precipitation Data (1992 - 2000)

Number of Years = 9

Number of Leap Years = 3

Number of Records = 3288

Number of No Data values = 488

Month	PCP_MM	PCPSTD	PCPSKW	PR_W1	PR_W2	PCPD
Jan.	81.53	6.1107	4.0802	0.1867	0.6991	12.56
Feb.	41.86	3.9803	4.2765	0.1326	0.6486	8.22
Mar.	88.84	5.9178	4.0596	0.3262	0.6304	15.33
Apr.	182.99	9.7280	2.8104	0.4828	0.7486	20.33
May.	106.46	6.2821	3.2343	0.2444	0.7083	16.00
Jun.	55.50	4.2508	4.4372	0.2484	0.6055	12.11
Jul.	36.07	2.8174	4.1589	0.1244	0.6410	8.67
Aug.	65.67	4.1635	3.6610	0.1783	0.7541	13.56
Sep.	50.42	3.0005	3.3155	0.2053	0.6891	13.22
Oct.	16.93	6.2111	3.6169	0.4634	0.7766	21.89
Nov.	194.61	8.3931	2.5351	0.5682	0.8540	25.11
Dec.	101.79	6.7759	4.4150	0.2741	0.7014	16.00

PCP_MM = average monthly precipitation [mm]

PCPSTD = standard deviation

PCPSKW = skew coefficient

PR_W1 = probability of a wet day following a dry day

PR_W2 = probability of a wet day following a wet day

PCPD = average number of days of precipitation in month

Loading the SWAT input data.

The SWAT model was able to successfully load and process the input data described earlier. In loading the DEM, the DEM properties had to be defined based on the selected projection (Transverse Mercator) and other parameters required for the projection were specified. Having successfully loaded the DEM, the digitized stream network was next input and processed together with the DEM to yield the drainage outlets of the various possible sub-watersheds. On the selection of the main watershed outlet and further processing, the watershed was delineated and likewise the sub-watersheds within the selected drainage basin and sub-basin parameters calculated. The next input data to be loaded was the land-use data which successfully loaded as shape file and also the land

use look up table. These were then re-classified into SWAT land-uses. The soil data, was next loaded as shape file earlier mentioned and followed by the soils look up table and again re-classified. The land use and soils data were then overlaid to produce the hydrologic response units. The weather data was next input. This involved the selection and loading of the location tables, and input data for precipitation, solar radiation, wind speed, relative humidity, weather generator data and dewpoint temperatures etc. Having successfully loaded the indicated data and provided estimations on other data (not major inputs data) as required, the model was able to accept and load all the input data acquired for the catchment in question and hence ready to run and produce the necessary output information on stream flow, soil moisture, chemical pollutants levels, on a daily, monthly or yearly basis during the selected period.

3.3.3 Model Evaluation

Graphical and statistical techniques were used for evaluating model performance. One of the evaluation statistics used was the Nash-Sutcliffe efficiency (NSE) being the most widely used evaluation criterion for testing the goodness of fit between the observed and simulated values. It indicates how well the plots of observed and simulated data fits the 1:1 line and is expressed as

$$NSE = 1 - \left[\frac{\sum_{i=1}^n (Y_i^{obs} - Y_i^{sim})^2}{\sum_{i=1}^n (Y_i^{obs} - Y^{mean})^2} \right] \quad (3.1)$$

Where Y_i^{obs} is the i th observation for the constituent being evaluated, Y_i^{sim} is the i th simulated value for the constituent being evaluated, Y^{mean} is the mean of observed data for the constituent being evaluated, and n is the total number of observations.

The Pearson's correlation coefficient (r) and coefficient of determination (r^2) is another goodness of fit criterion used in this model evaluation to describe the degree of co linearity between simulated and observed data. The Pearson's correlation coefficient is expressed as:

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (3.2)$$

Where x_i is the i th observed value, \bar{x} is the mean of the observed values, y_i is the i th predicted value, \bar{y} is the mean of the predicted values and n is the number of observations.

The deviation of volume (D_v) was also used in the model evaluation to assess over estimation or underestimation of the stream flow. The equation that represents this method of evaluation is indicated as:

$$D_v = 100 \frac{(\sum \hat{Q} - \sum Q)}{\sum Q} \quad (3.3)$$

Where:

\hat{Q} = simulated stream flow.

Q = observed value of stream flow.

3.3.4 Manual calibration

Detailed manual calibration was carried in order to improve model performance. The process of manual calibration was carried out for average annual conditions in which calibration was done for base flow, surface runoff and total stream flow. To obtain the observed annual base flow and surface flow, the base flow filter program (Arnold, 1995) was used to separate the base flow and surface runoff using available daily stream flow data

Separation of base flow and surface runoff

The stream flow was separated into base flow and surface runoff using the base flow filter program. The programme uses the daily flow data to perform the separation. Hence the base flow separation was carried out for the period 1/1/1992 to 31/12/2000 using the available stream flow records for the gauging station at the outlet of the sub catchment in question. The base flow is considered the ground water contribution to stream flow (Arnold et al. 1995). The programme receives as input the name of the file containing the

daily stream flow data for the selected period in a prescribed format. Several streamflow data files can be processed simultaneously. As an output, the types of information obtained for each stream flow input are as follows:

1. Name of stream flow data file (gauge file)
2. Fraction of stream flow contributed by base flow estimated in the first pass (Baseflow Fr1)
3. Fraction of stream flow contributed by base flow estimated in the second pass (Baseflow Fr2)
4. Fraction of stream flow contributed by base flow estimated in the third pass (Baseflow Fr3)
5. Number of individual base flow recessions used to calculate the master recession curve (NPR)
6. Base flow recession constant (Alpha Factor).
7. The number of days for base flow recession to decline through one log cycle (Base flow days).

Process of manual calibration

The process was begun by determining the average annual observed and predicted water yields. The initial simulated flow was based on the default values of the input parameters in the SWAT model prediction. The average daily flow for each year was estimated by first calculating the average daily flow for each month of the year, using the program pcpSTAT with observed daily flows as input to the program. The average daily flow for each month is computed automatically by the program. The annual daily average is then obtained by determining the mean of the average monthly values. The annual flows are then converted to depth (mm) to obtainable the annual water yields. The simulated water yield is determined from an output file in SWAT simulation that gives the simulated flow. The variable FLOW_OUT which gives the simulated flow is used. In SWAT simulation, options are provided for daily, monthly and annual simulation. When simulation is carried on an annual basis, the FLOW_OUT variable will give the annual average daily flows for each year covered in the simulation and the average values for the

years in question. The contributions made by base flow and surface runoff are estimated using the average annual values for the SWAT simulation outputs GWQ, SURQ, and WLD representing ground water flow, surface runoff and total water yield respectively. The average annual values for surface flow and base flow are converted into fractions by dividing by the total water yield. The fractions are then multiplied by the total water yield obtained from FLOW_OUT variable to obtain the actual average annual values for base flow, surface flow and Total Water Yield. These values of base flow, surface flow and total water yield are converted to depth in mm using the catchment area. If d is the annual average daily flow in m^3/s , then the annual water yield, W (mm) is determined as follows:

$$W = \frac{dx3600x24x365x10^9}{Ax10^9} \quad (3.4)$$

Where A is the catchment area in km^2 .

Adjustments were made appropriately on the input parameters while observing the changes in the predicted surface flow, baseflow and total flow to examine how they compare with the observed values. This is continued until reasonable correspondence between observed and simulated annual values is achieved.

3.3.5 Selection and evaluation of pedotransfer functions for estimating K_s

The saturated hydraulic conductivity is one of the input parameters for surface flow simulation using the SWAT model. This soil property is highly variable spatially as well as over time and therefore very unpredictable. Its determination has remained a challenge to scientists. K_s is usually predicted by relating it to other measurable soil properties using pedotransfer functions. Measured values of saturated hydraulic conductivity was not available from the soils data base used in this study. The parameter was therefore estimated using the a program known as “Soil Water Characteristics” that estimates K_s from soil texture and other properties such as organic matter content, percentage coarse fragments and density of compaction. The pedotransfer function option used in the Soil Water Characteristics program is indicated as Saxton2006 in the programme menu and is bases on PTFs developed by Saxton and Rawls (2006). An option of indicated as Saxton1986 in the menu is also available. Measured data on saturated soil hydraulic conductivity was not available from the data base used in this research. Other soil

properties were, however available and include texture, soil bulk density, total carbon, porosity among others. The pedotransfer functions used were those that could be used to calculate the saturated hydraulic conductivity using the available soil physical inputs indicated. The computations were done for each soil layer of 20cm thickness to a depth ranging from 0.6m to 1.0m depending on data available for each soil unit from the data base. The pedotransfer functions used are well documented (Lorentz et al., 2001, Gowing and Young, 1996 and Sobieraj et al., 2001).

3.3.6 Development of pedotransfer functions

Saturated hydraulic conductivity can be predicted by relating it to the basic soil properties using pedotransfer functions. To develop these functions, measured data on saturated hydraulic conductivity obtained alongside corresponding measured basic soil properties like texture, bulk density, organic carbon content etc may be used. In existing soil data bases, it is rare to find measured values of saturated hydraulic conductivity. Besides, the methods of measurement differ from region to region. Different methods also differ in their accuracy and reliability thereby rendering the measured values uncertain. An alternative approach to predicting hydraulic conductivity is by the use of moisture retention and hydraulic conductivity functions. The moisture retention equations have parameters that are obtained by fitting the observed data on moisture retention curve to the moisture retention equations. This technique was used in this study and based on the Van Genuchten moisture retention equation. Wösten et al. (2001) noted that data from existing international soil data bases having measured soil data can be analysed to enable prediction of hydraulic characteristics from measured soil data and points out that good approximations may be accurate enough for many applications that include being used as inputs to hydrologic models. Availability of measured soil hydraulic characteristics for wide range of soils and from a large and reliable international data bases are considered prerequisite for development of pedotransfer functions.

In this research, the International Soil Reference and Information Centre (ISRIC) soils database was used to obtain measured data on soil hydraulic characteristics and basic soil properties. It was possible to obtain all the required measured data on 457 soil samples

from various parts of the world. The data contained measured moisture retention curves as well as data on texture (percent sand, percent clay, and percent silt). The observed moisture retention curve obtained from measured data on moisture retention characteristics was fitted to the Van Genuchten moisture retention equation using a computer programme to determine parameters of the Van Genuchten equation. The parameters were the saturated soil moisture content (θ_s), residual soil moisture content (θ_r), air entry parameter, α and pore size distribution, n . The pedotransfer functions were then developed to estimate these parameters from readily available measured basic soil properties. Relating the moisture retention parameters to the basic soil properties then makes it possible to predict the moisture retention characteristics from the basic soil properties and hence enabling prediction of the saturated hydraulic conductivity. It would therefore not be necessary to have measured data on moisture retention characteristics to determine saturated hydraulic conductivity but instead, the basic soil properties would be used in determining the equations for moisture retention which can then be used to estimate K_s . Measurement of moisture retention characteristics is quite elaborate, tedious and time consuming and also expensive especially if a large number of sites is involved. From the sample data set of 457 samples, a sub dataset consisting of 342 (75%) samples was randomly selected for use in calibration of the pedotransfer functions to be developed while the remainder portion of 115 samples (25%) of the data was used in the validation process. In development of the pedotransfer functions, the following processes were undertaken:

Evaluation of the distribution of the moisture retention parameters

A requirement in developing pedotransfer functions is that the response (dependent) variables should be normally distributed. As a result, the moisture retention parameter of the Van Genuchten equation and their possible transformations were evaluated to establish which would best reflect the normal distribution. To check the extent to which the response variables would be normally distributed, statistical measures of Shapiro Wilk (W-value), Skewness Coefficient, and measures of Kurtosis were used in the assessment.

Statistical regression analysis

Statistical regression was performed between each transformed response variable and each of the basic properties and their transformations so as to establish the goodness of fit in each case based on the measure of the coefficient of determination (r^2). The purpose of this was to determine, for each response variable, the predictor variable or its transformation form that gives the best quality of fit between the two when a regression is performed. This would give an indication of which predictor variable or its transformation correlates best with the response variable or its transformation.

Analysis of cross correlations

To establish the level of dependence between the selected predictor variables for each response variable best correlated to them, a cross correlation was performed between the independent variables by determining the correlation coefficient (r) between each pair of the variables. This kind of analysis is performed to determine if there is any correlation between any two predictor variables among the set chosen for determining the multiple regression equations. It is preferable that the independent variables in a multiple linear regression equation be independent among themselves.

Multiple linear regression and validation of developed equations

This was carried out between each selected transformed response variables and selected combination of predictor variables/transformation consisting of the independent variables for which the response variable best correlated as determined by the measure of coefficients of determination. The multiple coefficient of determination was determined for each equation developed to assess the accuracy of the equation. The multiple regression was carried out between response variables and the predictor variables or transformations consisting of independent variables that are not themselves significantly correlated ($r < 0.5$). After development of the pedotransfer functions using the given data set, their reliability is then tested by comparing the values of predicted parameters using the developed equations with the observed data using an independent data set. An

independent data set consisting of 115 samples was used in the validation process. The evaluation statistic used included was the correlation coefficient among others.

3.4 Verification of developed pedotransfer functions for saturated hydraulic conductivity in surface flow prediction.

The developed equations relating parameters in the Van Genuchten moisture retention equation were used to determine saturated hydraulic conductivity for the catchment under study in the prediction of surface runoff. The relevant basic soil properties for each soil type in the study area was used to calculate the parameters. The calculated parameters were then used to determine the moisture retention characteristics for moisture contents within the range 0 to 1. i.e for each moisture content selected, say 10% (0.1), the corresponding suction was calculated based on the Van Genuchten moisture retention equation using the relationship indicated below:

$$h = \frac{1}{\alpha} [S_e^{n-1} - 1]^{\frac{1}{n}} \quad (3.5)$$

$$\text{where } S_e = \left[\frac{\theta - \theta_r}{\theta_s - \theta_r} \right] \quad (3.6)$$

The above indicated equation was used to determine the matric suction at various moisture contents in each of the soil types covering the catchment area.

To determine saturated hydraulic conductivity, the equation relating $k(h)$ to the Van Genuchten parameters is used i.e.

$$K(h) = K_s \left[\frac{(1 - (\alpha h)^{n-1} (1 + (\alpha h)^n)^{-m})^2}{(1 + (\alpha h)^n)^{m/2}} \right] \quad (3.7)$$

To determine $K(h)$ from moisture retention characteristic, an equation that relates hydraulic conductivity to suction head is required so as to obtain $K(h)$ from $h(\theta)$. The equations relating hydraulic conductivity to permeability (Obiero, 1996; Marshal and Holmes, 1958) used to determine $K(h)$ from h when combined may be expressed as:

$$K(h) = \frac{(\rho_s - \rho_b)(\gamma^2)}{2\eta\rho_s g h^2} \quad (3.8)$$

Where ρ_b is the soil bulk density and ρ_s the soil particle density.

Using typical values of γ , η , ρ_s , and ρ , $K(h)$ values were computed for each value of moisture suction. Also for each moisture suction, the expression adjacent to K_s on the RHS of the hydraulic conductivity equation was computed. Since K_s is constant, $K(h)$ was plotted against the expression adjacent to K_s on right hand side of the hydraulic conductivity equation (eqn. 3.7) and a straight line passing through the origin obtained. The slope of this line would then give the saturated hydraulic conductivity. Hence for each hydrologic response unit in the catchment, the saturated hydraulic conductivity was determined depending on the dormant soil type. These values of saturated hydraulic conductivity then served as input to the SWAT model for simulation of surface runoff. The surface runoff was thus determined at the main catchment outlet. The surface runoff simulated using the developed pedotransfer functions was evaluated based on comparison with the observed surface runoff to assess its performance in surface flow simulation. The performance was also compared to that of other selected existing pedotransfer functions previously assessed in surface flow simulation to examine its performance relative to existing pedotransfer functions.

4 RESULTS AND DISCUSSION

4.1 Preliminary assessment of Model Performance

Daily simulations

The model was initially run for a warm-up period of six months in which a daily simulation was carried out during the period 1/1/92 to 30/6/92. The warm up period was chosen to fall before the calibration period which commenced from 1/7/92 and being a period when continuous data on observed flow was available without gaps (missing data). Daily and monthly stream flow simulations were then performed in the period 1/7/92 to 30/6/95 representing three years. Model evaluation was begun with simulation based on a daily time step. **Figure 4.1** shows the trend of stream flow hydrographs for the observed and simulated flows (m^3/s) during the three year period based on the first modeling run.

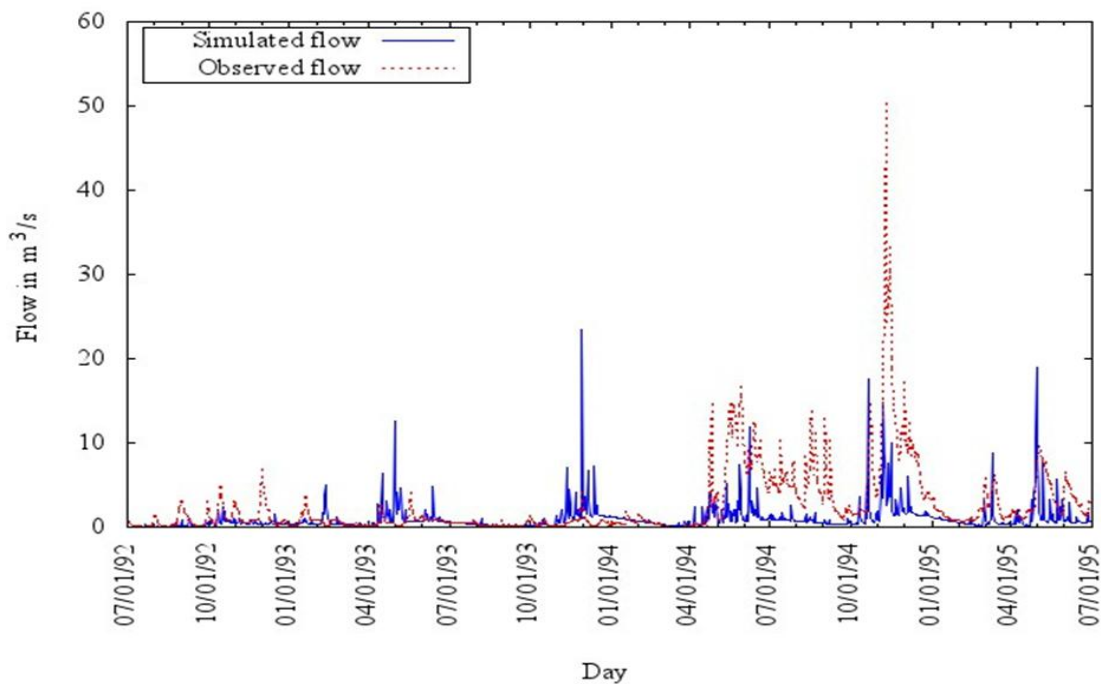


Figure 4.1. Hydrographs of observed and simulated daily flows during the period 1/7/92 to 31/6/95

The model over predicted flow during certain periods and under predicted in others while in some periods, the observed and simulated flows were in agreement. In general, the model under predicted high flows while simulating low flows fairly.

Comparisons based on daily time step are likely to be misleading due to the manner in which the model computes the daily flows which differ from that used in recording observed flows and this affects the values of peak flows. The observed flows are based on instantaneous readings taken at a certain time of day (e.g. 9.00am) in the morning while the simulated flows are based on the daily average. If heavy rainfall occurs close to the time when the observation is about to be read say 7am in the morning the resulting peak flow is likely to be reflected in the observed record. However if the rainfall occurs much earlier e.g. the previous day, then it is likely that the resulting runoff will have passed the catchment outlet before a reading is taken so the peak flow would not be reflected in the daily flow reading. The surface runoff may, however, be captured by the flow simulation especially if the daily flow is reasonably high so that the daily average of the runoff will have a high value. For storm events that occur closer to the time of observation, the peak flows are captured by both the observed and simulated flows. **Figure 4.2** shows a comparison of observed and simulated flows based on linear regression with values of the y-intercept and coefficient of determination (r^2) also indicated.

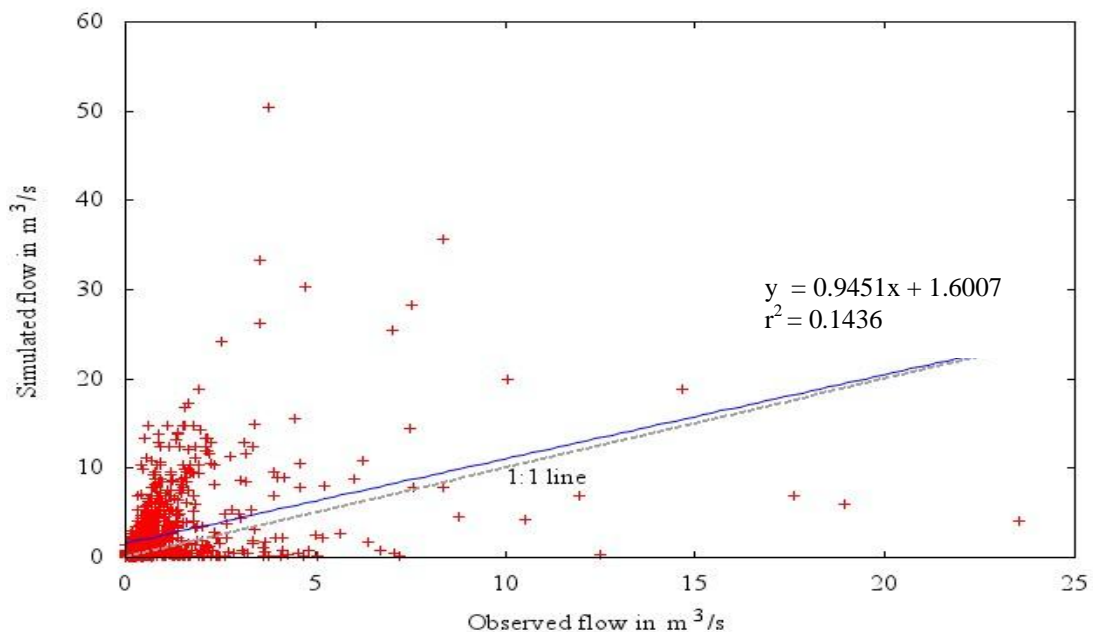


Figure 4.2. A comparison of simulated and observed daily flow during the period 1/7/92 to 31/6/95

The value of r^2 was found to be 0.144 while Nash-Sutcliffe efficiency (NSE)=0.01 reflecting a poor linear relationship between the observed and predicted values based on daily simulation. This poor performance, which is often misleading, may be attributed to differences in the timing of observed and simulated hydrographs likely to occur when using daily rainfall data. Daily simulations do not provide values that are expected to compare reasonably well with the predicted ones. This is partly due to the poor prediction procedure for peak flows. As a result therefore, detailed evaluation of the model performance including model calibration was done based on monthly time step. This is in consideration of the fact that monthly values are likely to be more representative than the daily values since with daily values cumulated over the month, the daily errors are likely to be cancelled out. Hence the subsequent calibration process was considered on the basis of monthly simulations.

Monthly simulations

Simulation was done based on monthly basis to observe the performance of the model based on a monthly time step. **Figure 4.3** shows the hydrographs of the average daily flows for each month for observed and predicted flows during the period 1/7/92 to 30/6/95 during the first modeling run. From the figure, it can be observed that the model performance has improved compared to that based on a daily time step. A similar observation was made by Githui et al. (2009) while evaluating performance of SWAT model in the Nzoia catchment in western Kenya in which the author noted that the agreement between observed and simulated flows was stronger with monthly than with daily flows during calibration. During the period July 1992 to around April 1994, the model predictions of stream flow seem to agree with the observed values except for a few instances where there was over prediction of flow like in December 1994. For the period April 1994 to June 1995, the model generally under predicted the flows. The model is therefore a poor simulator of high flows but fairly simulates low flows. **Figure 4.4** shows a plot of observed and predicted flows based on the linear regression. The values of the coefficient of determination and regression equation expressing the linear relationship are as well indicated.

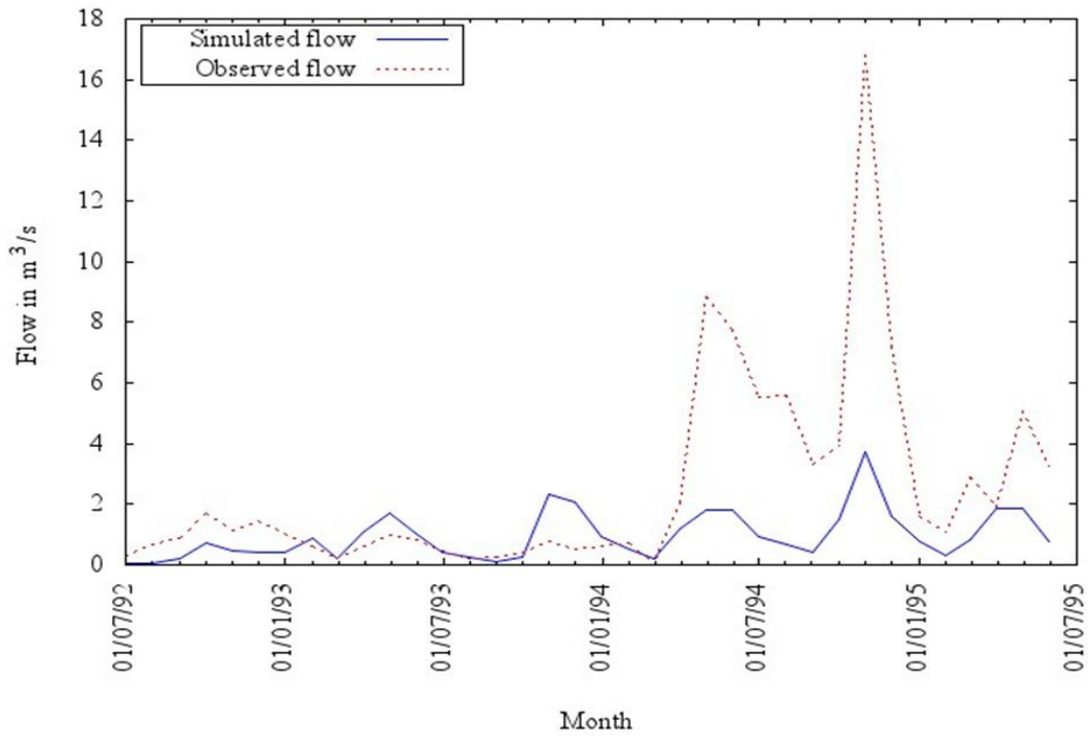


Figure 4.3. Hydrographs of simulated and observed mean daily flows in month for the period July 1992 to June 1995

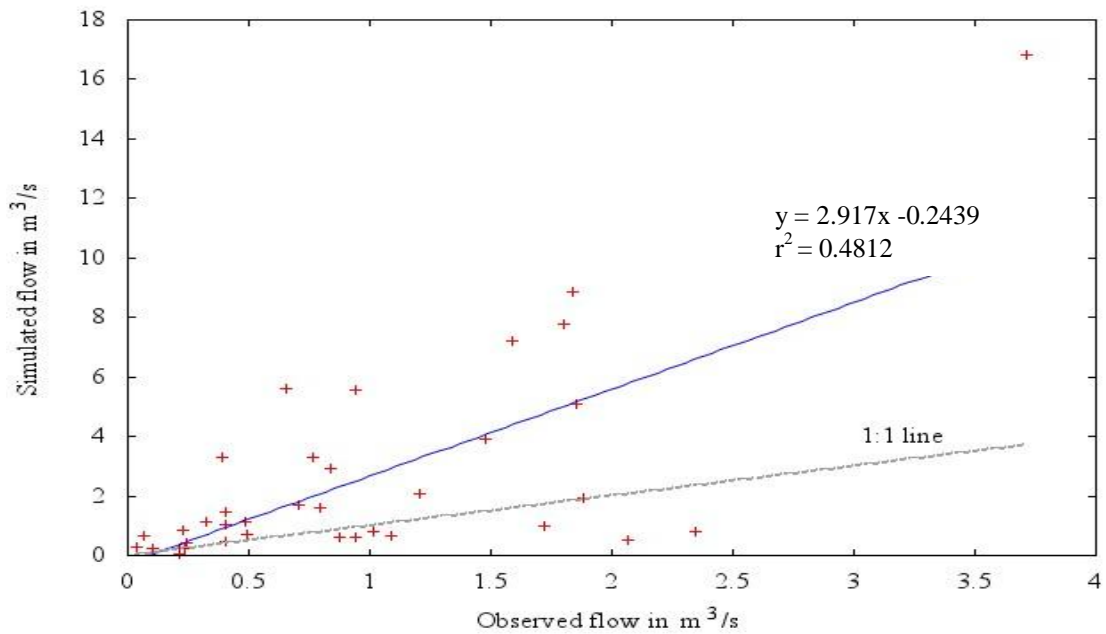


Figure 4.4. Comparison of observed and predicted mean daily flows in the month for the period 1/7/1992 to 30/6/1995

The value of the coefficient of determination (r^2) of about 0.5 indicates acceptable performance. The correlation coefficient $r = 0.7$ showing that the predicted and observed flows exhibit linear relationship. The Nash-Sutcliffe efficiency, however registered a low value of 5% indicating poor simulation performance. The low value of NSE can be attributed to a strong deviation volume (D_v) of 61.7% as noted by Levesque et al. (2008). The high positive value of D_v indicates the average tendency of the simulated flows to under estimate the flows. The process of model calibration was carried out to further assess model performance and to examine possibilities of domesticating the model for application in the local catchment.

4.2 Sensitivity Analysis and Model Calibration

Sensitivity analysis was carried out to find the order of sensitivity of stream flow to the input parameters. The sensitivity analysis process was carried out automatically by the model. The model produced the output indicating the order of sensitivity of the model input parameters. **Table 4.1** shows an extract of the summary of the output format from sensitivity analysis. Interpretation of the results indicate that the curve number (CN2) is the most sensitive parameter. The ranking of the parameters is indicated in the row labeled “out”.

An auto calibration process was the carried out based on a selection of the three most sensitive parameters. These were the Curve number (CN2), Soil Evaporation compensation factor (ESCO), and the Threshold water depth in the shallow aquifer for “revap” (QWQMN). After the auto calibration process, the best values of the selected parameters for which the model prediction closely agreed with the observed were determined and produced in the auto calibration results. The model was again rerun based on these values. Not much change was noted in the hydrographs of observed and simulated flows after the auto calibration process.

Table 4.1: Sensitivity analysis results showing the order of sensitivity of input parameters

	SFMX	SMFMN	ALPHA_BF	GWQMN	GW_REAP		
Of 1	28	28	3	1	28		
Out 1	28	28	12	3	28		

	REVAPMN	ESCO	SLOPE	SLSUBBSN	TLAPS	CH_K2	
Of 1	28	5	8	28	12	2	
Out 1	28	2	6	19	28	15	

	CN2	SOL_AW	surlag	SFTMP	SMTMP	TIMP	
Of 1	2	4	11	28	28	28	
Out 1	1	4	14	28	28	28	

	GW_DELAY	Rchrg_dp	canmx	sol_k	sol_z	
Of 1	13	6	28	7	9	
Out 1	18	18	8	7	5	

	Sol_alb	epco	ch_n	blai	BIOMIX	
Of 1	28	10	28	28	14	
Out 1	16	13	17	11	19	

A comparison of observed and predicted monthly flows yield a coefficient of determination slightly above 0.50 with minimal change in the value of NSE which still remained low at 6% and a small reduction in deviation volume to 61.3%. This indicated acceptable but modest performance of the model for the catchment in question. An attempt was made to perform a manual calibration of the model. This was done by varying each of the three most sensitive parameters by 10% from their default values, but

within the allowable range, and selecting the value that provides the best possible agreement between observed and simulated flows. The parameters were varied one at a time while keeping the others constant until an optimal value is obtained. A slight, but insignificant improvement was observed in the values of the evaluation statistics with $r=0.72$ ($r^2=0.51$) and NSE of 5%, with the deviation volume rising to 64%. **Table 4.2** shows the calibration results. The seemingly poor performance of the model could be associated with input data deficiencies also observed by Jayakrishnan et al. (2005). Daily rainfall data in the vicinity of the catchment was available from three rainfall stations in which only two were located within the catchment and near the outlet. The third station was located outside the catchment near the upstream end. Hence the rainfall may not have been representative. Only one full meteorological station was available with adequate weather data for use in the modeling but was located well outside the catchment. The weather data may therefore, also not have been adequately representative. Besides, there were also cases of missing data during certain periods for the stations used.

4.3 Model Validation

Based on the optimized parameters obtained during the calibration period, a further simulation was carried out to assess the model performance during the period 1/1/98 to 31/12/2000 which is outside the period when the model was calibrated. **Figure 4.5** shows the graphical representation of the observed and simulated flows during this validation period. Visual observation of the hydrographs shows a fairly close fit, an indication of improved model performance. **Figure 4.6** shows a plot of validation results of simulated against observed monthly flows.

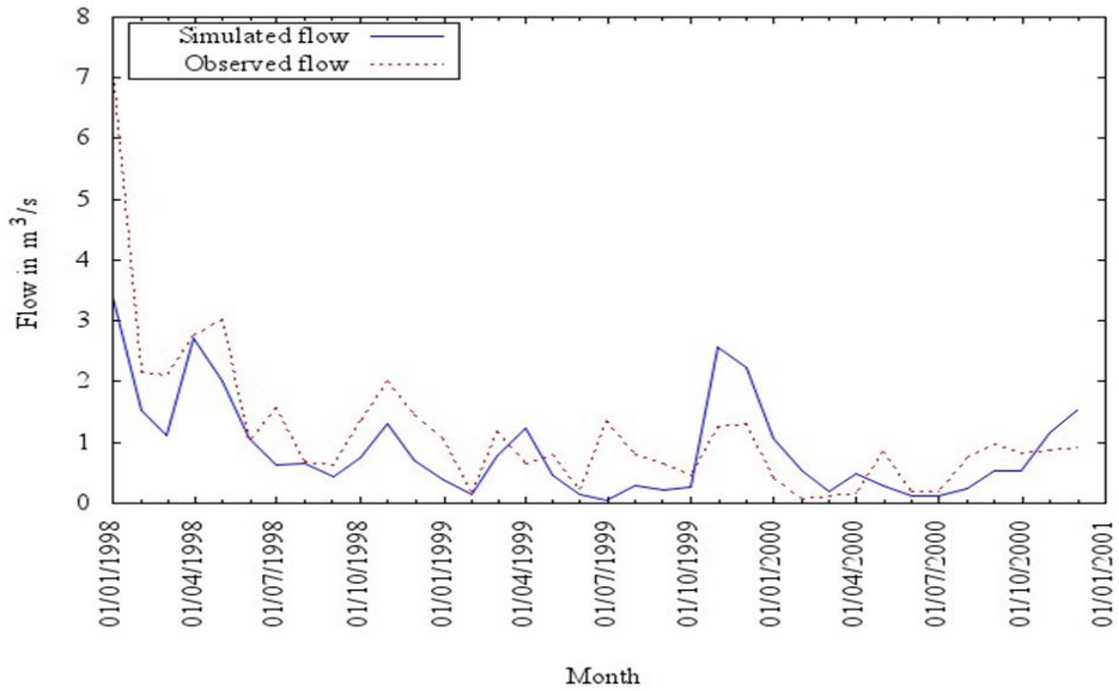


Figure 4.5 Hydrographs of simulated and observed mean daily flows in month for the validation period 1/1/98 to 31/12/2000

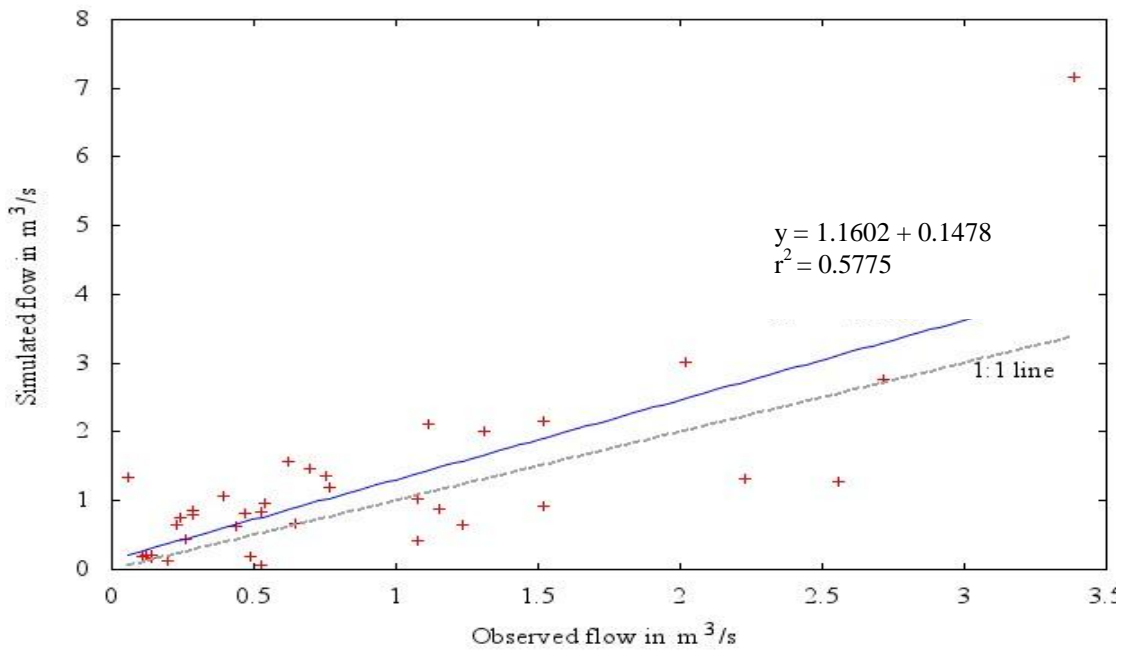


Figure 4.6. Comparison of observed and predicted mean daily flows in the month for the period 1/1/1998 to 31/12/2000

There is evidently improved performance of the model with the coefficient of determination, $r^2=0.58$ ($r=0.76$). The NSE value significantly improved to 0.51 while the deviation volume reduced to 24.7%. This reflects acceptable model performance (Moriasi et al., 2007) which can be considered satisfactory and therefore promising for applicability in the catchment. A summary of calibration and validation results are indicated in **Table 4.2**

Table 4.2 Evaluation of model results for first flow simulation runs, calibration and validation

Evaluation Statistic	First Simulation Run		calibration	Validation
	Daily	monthly	Monthly	Monthly
Nash-Sutcliffe efficiency (NSE)	0.01	0.05	0.05	0.51
Correlation Coefficient (r)	0.30	0.69	0.72	0.76
Coefficient of Determination (R^2)	0.09	0.48	0.51	0.58
Deviation Volume (D_v)	22.57	61.7	64.06	24.66

4.4 General Assessment of Model Performance

The results obtained in this study are not unique. Applications of SWAT worldwide has yielded diverse results some encouraging while in some instances, the success in the use of the model has been limited. This may be attributed to strengths and weaknesses associated with the use of SWAT across the world spectrum. A few studies involving SWAT supports this observation. Good model performance has been reported by Schmalz et al. (2008) while modeling water balance with SWAT on three catchment areas in northern Germany. However, the first model runs failed to represent the stream flow correctly (also observed in this study) showing underestimation of observed high winter discharge peak and over estimation of base flow. Success in model prediction was

attributed to repeated attempts to improve the model performance. The challenges of modeling with SWAT in lowland areas is mentioned and which had to be taken into account in model parameterization. Stehr et al. (2008) evaluated the performance of the SWAT model in several sub basins in Chile. The model performance was not uniform in all the sub catchments studied. The NSE index ranged from satisfactory to good for the calibration period depending on the sub basin. The model was observed to have underestimated peak flows, a similar observation to this study. An explanation for this was given as inadequate description of rainfall input field due to the limited number of available meteorological stations and poor representation in higher areas due to orographic effects. This scenario is similar to this study. Only one rainfall station was available to represent the rainfall near the higher elevations of Naro Moru catchment. The station was also outside of the catchment but was the nearest available for use. The standard interpolation method used in AVSWAT for estimating rainfall (Thiesen polygons) was notably a limitation as its reliability is yet to be tested if discrete improvements in model performance is to be expected.

4.5 Manual calibration and Model improvement

Preliminary assessment indicated a modest SWAT model performance for this catchment under study (Obiero et al. 2011). An elaborate manual model calibration was carried out to improve model performance yielded results discussed as follows:

4.5.1 Separation of base flow and surface runoff

Separation of daily stream flow into base flow and surface runoff yielded as output a daily file containing the data, stream flow and Bflow pass 1, Bflow pass2, and Bflow Pass3 representing the Baseflow value that is estimated in the first , second and third passes respectively. The daily output data file for the gauging station and period in question is illustrated in **Table 4.3**. The base flow was determined as the average value of the average value of BaseFlow Fr1 and Baseflow Fr2 since the fraction of water yield contributed by baseflow is expected to fall somewhere between the two passes. The primary output file is illustrated in **Table 4.4** showing a summary of the output.

Table 4.3 An illustration of the daily output file for the month of January 1992 showing daily baseflow filter values

YEAR	MNDY	Streamflow	Bflow Pass1	Bflow Pass2	Bflow Pass3
1992	1 1	1.14E-01	5.70E-02	0.00E+00	5.70E-02
1992	1 2	5.40E-02	5.40E-02	0.00E+00	0.00E+00
1992	1 3	7.70E-02	5.49E-02	0.00E+00	0.00E+00
1992	1 4	1.09E-01	5.77E-02	0.00E+00	0.00E+00
1992	1 5	1.42E-01	6.28E-02	0.00E+00	0.00E+00
1992	1 6	1.74E-01	6.99E-02	0.00E+00	0.00E+00
1992	1 7	4.39E-01	8.77E-02	0.00E+00	0.00E+00
1992	1 8	4.39E-01	1.14E-01	0.00E+00	0.00E+00
1992	1 9	3.07E-01	1.33E-01	0.00E+00	0.00E+00
1992	110	1.74E-01	1.41E-01	0.00E+00	0.00E+00
1992	111	1.68E-01	1.44E-01	0.00E+00	0.00E+00
1992	112	1.61E-01	1.45E-01	0.00E+00	0.00E+00
1992	113	1.55E-01	1.46E-01	0.00E+00	0.00E+00
1992	114	1.05E-01	1.05E-01	0.00E+00	0.00E+00
1992	115	1.21E-01	1.06E-01	0.00E+00	0.00E+00
1992	116	1.55E-01	1.08E-01	0.00E+00	0.00E+00
1992	117	1.74E-01	1.12E-01	0.00E+00	0.00E+00
1992	118	1.74E-01	1.17E-01	0.00E+00	0.00E+00
1992	119	1.74E-01	1.21E-01	0.00E+00	0.00E+00
1992	120	1.74E-01	1.25E-01	0.00E+00	0.00E+00
1992	121	1.74E-01	1.29E-01	0.00E+00	0.00E+00
1992	122	1.05E-01	1.05E-01	0.00E+00	0.00E+00
1992	123	7.10E-02	7.10E-02	0.00E+00	0.00E+00
1992	124	3.60E-02	3.60E-02	0.00E+00	0.00E+00
1992	125	7.60E-02	3.75E-02	0.00E+00	0.00E+00
1992	126	1.15E-01	4.19E-02	0.00E+00	0.00E+00
1992	127	1.55E-01	4.88E-02	0.00E+00	0.00E+00
1992	128	1.39E-01	5.62E-02	0.00E+00	0.00E+00
1992	129	1.37E-01	6.23E-02	0.00E+00	0.00E+00
1992	130	1.05E-01	6.67E-02	0.00E+00	0.00E+00
1992	131	2.72E-01	7.59E-02	0.00E+00	0.00E+00

Table 4.4 Summary of the fraction of stream flow contributed by base flow for each of the 3 passes made by the base flow filter programme for the period 1992 to 1995

Gage file	BaseflowFr1	Baseflow Fr2	Baseflow Fr3	NPR	Alpha Factor	Baseflow Days
0929500.prn	0.68	0.52	0.43	3	0.0458	50.2652.

4.5.2 Manual Calibration Results

Table 4.5 illustrates the computations of simulated and observed water yields for the calibration period 1/7/1992 to 30/6/1995 based on default values of the SWAT input parameters after first annual simulation run. Based on the above indicated computations, the initial average annual observed and simulated results are summarized as shown in **Table 4.6** for the indicated period.

Table 4.5: Estimation of observed and simulated annual water yield, base flow and surface flow (mm) for annual calibration

DATE	Simulated Surface flow (SURQ)	Simulated Ground water flow (GW_Q)	Simulated water yield (WYLD)	$\frac{SURQ}{WYLD}$	$\frac{GW_Q}{WYLD}$	Observed annual water yield(mm)	Observed annual base flow (mm)	Observed Annual surface flow (mm)	Simulated total flow (mm)	Simulated surface flow (mm)	Simulated base flow (mm)
Col 1	Col 2	Col 3	Col4	Col5	Col 6	Col 7	Col 8	Col 9	Col 10	Col 11 =col 5x col10	Col 12 =col10-col 11
1992	26.353	263.560	291.437	0.090	0.904	375.06	165.098	209.958	121.231	10.962	110.269
1993	77.865	623.419	704.703	0.110	0.885	213.27	127.593	85.675	334.462	36.956	297.506
1994	81.214	758.152	843.003	0.096	0.899	1919.41	1169.29	750.113	476.175	45.874	430.300
1995	12.313	192.465	206.146	0.060	0.934	985.44	620.68	364.761	403.737	24.115	379.622
Ave	65.975	613.092	682.386	0.097	0.898	873.29	520.666	352.627	445.655	43.087	402.568

Table: 4.6 Summary of annual observed and simulated water yield (mm)

	Total Water Yield (mm)	Baseflow (mm)	Surface Flow (mm)
Actual	873	521	352
SWAT(simulated)	446	403	43

Evidently from **Table 4.6**, the observed annual total water yield, base flow and surface flow are all higher than the simulated flow values hence the need for appropriate adjustment of the appropriate input parameters through the manual calibration process.

4.5.2.1 Water balance and Total Flow Calibration

As observed by Neistch et al. (2002)(c) calibration for the water balance and stream flow

is to be carried out based on annual conditions before shifting to monthly records for fine tuning. Therefore adjustments were made on selected SWAT input parameters with a view to bringing the simulated flow values closer to the observed stream flow values. The process went as follows:

Calibration of annual surface, subsurface and total stream flows

In an attempt to raise the surface flow to an acceptable level, the curve number was adjusted for each of the 27 sub basins based on vegetation cover and hydrologic soil group initially assuming fair hydrologic cover conditions. Values of curve numbers for various land uses and hydrologic soil groups are tabulated for agricultural land and urban areas (see **Tables A 3.1** and **A3.2** appended). The curve number values selected for each sub basin was based on the cover type that closely described the land cover in the sub basin.

The area around the peak of the mountain consisting mainly of rocks was assumed to be equivalent to pavements in the urban areas with a curve number of 98. The curve number was adjusted initially assuming good hydrologic condition, as a result of which there was no significant change in the value of average annual surface runoff which changed from 43mm to 40mm. The value of base flow was however significantly increased from a value of 403mm to 454mm. Further adjustment of the curve number was carried out based on fair conditions. In this case the annual average surface flow increased to a value of 96mm while the base flow increased further to a value of 564mm. The resulting value of surface flow still fell far below the observed which is 352mm. In order to increase the surface flow further, the curve number values were adjusted to those associated with the poor conditions. In this case there was a significant increase in the simulated surface flow to a value of 427mm up from the value of 96mm associated with fair conditions. The base flow was however drastically reduced from 564mm to 245mm. It therefore emerged that an appropriate value would fall somewhere between the poor and fair conditions.

The curve numbers based on average values for the fair and poor conditions was then used in adjusting the curve numbers further. An annual simulation was again performed.

The average value of surface runoff was reduced to 219mm nearly reflecting half the previous value. The base flow was reduced to 446mm. The surface flow value obtained was still not acceptable compared to the observed surface flow. The base flow was however falling within 15% of the observed base flow. In order to increase the surface flow further, the mean value was determined between the poor conditions and the previously computed average values between the fair and poor conditions. This yielded an average annual surface flow of 308mm while the base flow was further reduced to 360mm when an annual simulation was done. This resulting value of surface flow is acceptable falling within 15% of the observed. **Table 4.7** Shows the estimated curve number values assigned to each sub basin during the adjustment process based on the predominant land use type in the basin and the hydrologic soil group associated with the soil predominant in the basin. The value of base flow was, however, still very low compared to the observed.

In an attempt to raise both the surface flow and base flow, the value of soil evaporation compensation factor (ESCO) was adjusted to the maximum possible value of 1.0. on running the annual simulation, the value of surface runoff was increased to 324mm while the base flow increased to 418 reflecting a significant increase bringing the simulated base flow closer to the observed. The SWAT model provides various options of estimating the evapotranspiration. The default method of estimating the evapotranspiration is the Priestly Taylor. When the method of evapotranspiration was then changed to Penman in an attempt to observe the response of surface and subsurface flow, it was observed that the surface flow increased further to a value of 350mm per annum drawing much closer to the observed value of 352. However the base flow was slightly reduced to a value of 406mm per annum. Further adjustments were made on various input parameters including the threshold depth of water in the shallow aquifer for “revap” to occur (REVAPMN), Deep aquifer percolation fraction (RCHRG_DP), among others in an attempt to bring the observed and predicted average total surface and subsurface flows closer. **Table 4.8** illustrates the adjustments with the adjusted values of the parameters indicated in italics enclosed in the brackets.

Table 4.7: Curve number values for various land use types in the sub basins used in the calibration.

Reach (sub basin)	Land use type	Hydrologic soil group	Curve Number					
			Default	Poor Conditions	Fair conditions	Good cover	Average (fair & poor)	Ave (Poor & Fair&Poor)
1	Barren Land	D	80	98	98	98	98	98
2	Barren Land	D	80	98	98	98	98	98
3	Barren Land	D	80	98	98	98	98	98
4	Barren Land	D	61	98	98	98	98	98
5	Forest	D	77	83	79	77	81	82
6	Forest	D	77	83	79	77	81	82
7	Barren Land	D	80	98	98	98	98	98
8	Barren Land	D	80	98	98	98	98	98
9	Forest	C	66	77	73	70	75	76
10	Barren Land	D	61	98	98	98	98	98
11	Forest	C	66	77	73	70	75	76
12	Forest	C	66	77	73	70	75	76
13	Forest	C	66	77	73	70	75	76
14	Forest	C	66	77	73	70	75	76
15	Forest	D	77	83	79	77	81	82
16	Forest	C	66	77	73	70	75	76
17	Woodland	D	73	86	82	79	84	85
18	Forest	C	77	83	73	70	78	81
19	Forest	C	66	77	73	70	75	76
20	Forest	C	77	83	79	70	81	82
21	Woodland	D	73	83	79	79	81	82
22	Forest	C	77	77	73	70	75	76
23	Woodland	C	60	86	76	72	81	84
24	Woodland	D	79	86	82	79	84	85
25	Forest	C	66	83	73	70	78	81
26	Barren Land	D	61	98	98	98	98	98
27	Woodland	B	60	86	65	58	76	81

Table 4.8: Summary of annual observed and simulated flows following parameter adjustments during manual calibration.

	Input parameter values	Total Water Yield	Baseflow	Surface Flow	r	NSE
Actual		873	521	352		
SWAT(simulated)	(DEFAULT SWAT Input values)	446	403	43	0.69	0.05
CN-GOOD CONDITIONS		494	454	40	0.65	0.06
CN-POOR CONDITIONS	DEFAULT VALUES	672	245	427	0.56	0.15
CN-FAIR CONDITIONS	DEFAULT VALUES	660	564	96	0.58	0.14
CN-AVE-FAIR&POOR	DEFAULT VALUES	664	446	219	0.57	0.14
CN-[AVE-(FAIR&POOR)&(POOR)]	DEFAULT VALUES	668	360	308		
	(<i>esco=1</i>)	742	418	324	0.58	0.19
	(<i>esco=1, PENMAN</i>)	756	406	350	0.58	0.20
	(<i>esco=1, PENMAN, REVAPMN=50</i>)	767	423	345	0.58	0.20
	(<i>esco=1, PENMAN, REVAPMN=100</i>)	767	423	345	0.58	0.20
	(<i>esco=1, PENMAN, REVAPMN=100, RCHRG_DP=0</i>)	774	436	339	0.58	0.20
	(<i>esco=1, PENMAN, REVAPMN=100, RCHRG_DP=0, GWDELAY = 100</i>)	767	422	346	0.55	0.18
	(<i>esco=1, REVAPMN=100, RCHRG_DP=0, GWDELAY=31, HARGREAVES</i>)	752	432	320	0.62	0.22
	(<i>esco=1, , REVAPMN =100, RCHRG_DP=0, GWDELAY = 31, HARGREAVES, surlag=2</i>)	751	432	319	0.62	0.22
	(<i>esco=1, , REVAPMN =100, RCHRG_DP=0, GWDELAY = 31, HARGREAVES, surlag=2, k100%increase</i>)	762	454	309	0.60	0.22

DEFAULT VALUES: REVAPMN =1.0, ESCO=0, Priestly Taylor, GWQMN=0, RCHRG_DP=0.05, ALPHA_BF=0.048, CH_K1=0.500, GW_DELAY =31.

Acceptable values of base flow and surface runoff were obtained when the following adjusted values of input parameters were used: $Esco=1$, $REVAPMN=50$, $RCHRG_DP=0$ and using the penman method of evapotranspiration. The difference between the observed and simulated values were within 15%. Any attempt to improve the simulation values did not yield further improvement e.g. when the GW_DELAY was adjusted to 100, the surface flow increased from 339 to 346 improving the surface flow simulation, however the simulated base flow was slightly reduced from 436mm to 422mm. After each parameter adjustment, monthly simulations were also done and the observed and predicted monthly flow compared. The correlation coefficient changed from a value of 0.67 to 0.58 while the coefficient of efficiency increased from 0.05 to 0.20 signifying improved model performance. To examine the effect of change in the method of evapotranspiration estimation from Penman to Hargreaves, the annual simulation was carried out using the Hargreaves option for estimating evapotranspiration while GW_DELAY was reverted to the default value of 31. This resulted into an increase in the value of base flow to 432 mm, surface flow was reduced to 320mm while the total flow was 752mm which could still be regarded as acceptable. The difference between the observed and simulated baseflow was however increased slightly to 17% of the observed flow. The correlation coefficient between observed and simulated monthly flows however improved significantly to 0.62 while the coefficient of efficiency increased slightly to 0.22 indicating that the use of Hargreaves improved model performance for monthly simulations. To smoothen the hydrographs, the value of surlag was reduced from the default value of 4 to 2. This resulted into a slight decrease in the average annual surface flow from 320mm to 319mm. The value of simulated base flow did not change but was still considerably lower than the observed one. A parameter that was used to increase the baseflow was the hydraulic conductivity. The saturated hydraulic conductivity value was then raised by 100% in each layer for every sub basin. This raised the value of simulated base flow significantly to 454mm up from 432 hence the simulated base flow fell within 10-15% of the observed base flow. The surface runoff was however reduced to 309mm from the previous value of 319mm. This value was still falling within 10-15% of the observed value of 352mm. The total flow became 762mm also falling

within 10-15% of the observed total flow; hence the annual calibration process was concluded at this stage. **Table 4.9** illustrates the change in simulated surface flow, base flow and total flow after the annual flow was concluded when comparison of the flows indicated that the surface, base flow and total flow fell within 10-15% of the observed values.

Table: 4.9 Summary of observed and simulated average annual yield before and after annual flow calibration

		Total Water Yield (mm)	Base flow (mm)	Surface Flow (mm)
Actual		873	521	352
SWAT (simulated)	Before calibration	446	403	43
	After calibration	762	454	309

After the annual calibration of water yields, a monthly simulation was then carried out to examine the model performance based on monthly simulations. **Figure 4.7** shows the hydrograph of observed and simulated total stream flows after the completion of the annual calibration. The low flows are favourably predicted with flow peaks being slightly overpredicted. However high flows are generally underpredicted with the peak flows grossly underpredicted. **Figure 4.8** shows a comparison of observed and simulated average daily total flows in month after the calibration. The correlation coefficient is 0.6 considered favourable still while the coefficient of efficiency (Nash-Sutcliffe) is 0.22, a significant improvement from a value of 0.05 before the annual calibration exercise. **Table 4.10** shows model performance after manual calibration compared to previous auto calibration. It is evident that the manual calibration based on annual flows resulted into an improvement in model performance compared to the auto calibration process. The purpose of manual calibration was to improve model model performance and establish the extent to which the model could be improved.

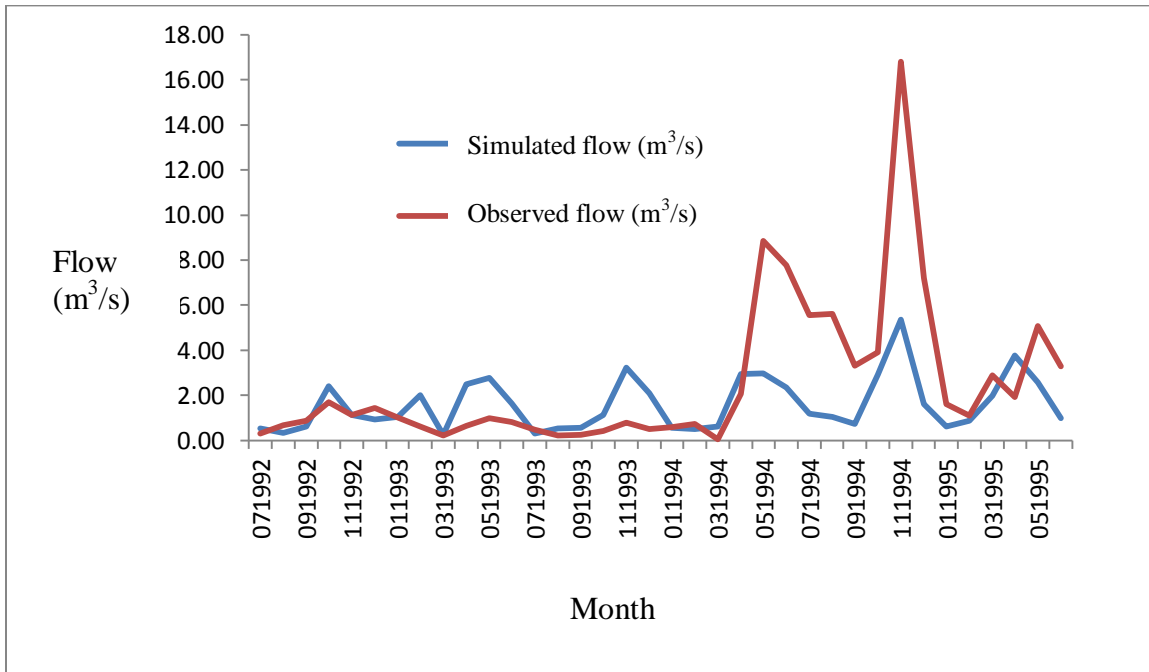


Figure 4.7. Hydrograph of observed and simulated monthly total flows after annual flow calibration

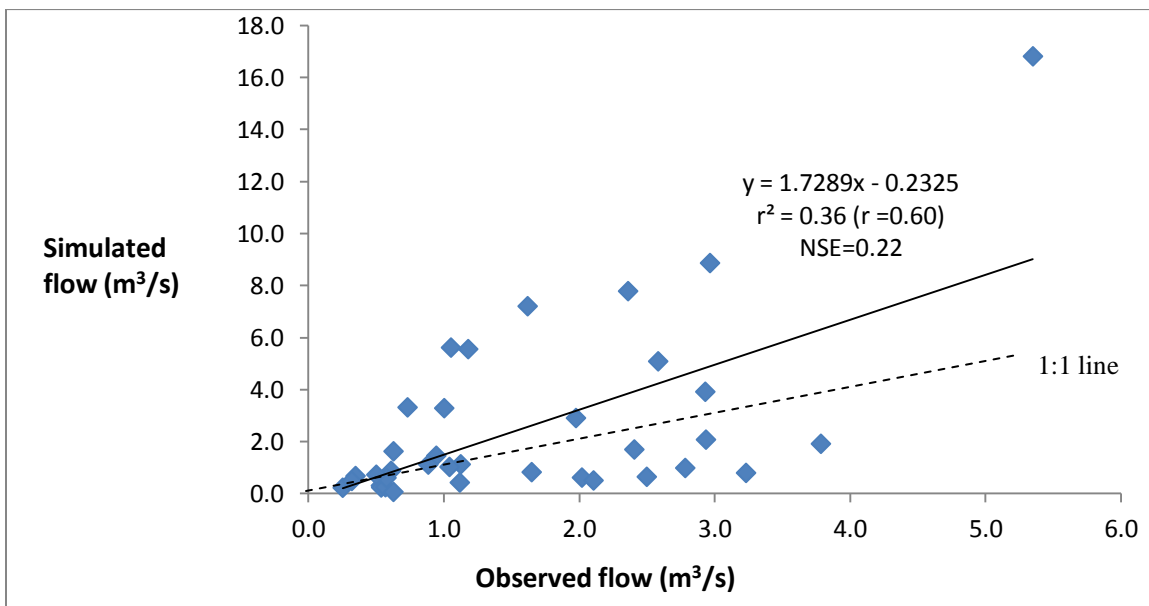


Figure 4.8 Comparison of simulated and observed average daily flow in month after calibration of average annual water yields

Table 4.10 Comparison of model performance for monthly simulation based on auto calibration and annual flow calibration processes.

Evaluation Statistic	First Simulation Run		Monthly simulation	
	Daily	monthly	Auto	Manual
Nash-Sutcliffe efficiency (NSE)	0.01	0.05	0.05	0.22
Correlation Coefficient (r)	0.30	0.69	0.72	0.60
Coefficient of Determination (r^2)	0.09	0.48	0.51	0.36
Deviation Volume (D_v)	22.57	61.7	61.30	36.86

4.5.2.2 Model Performance in surface flow and base flow simulation after annual flow calibration

An assessment was carried out to evaluate how the model performed in monthly surface and base flow simulations after the annual calibration was completed. A comparison was made between the observed and simulated base flow and surface flow as follows.

Surface flow simulation

Figure 4.9 shows a comparison of the hydrographs of observed and simulated monthly surface runoff after the annual flow calibration. The low flows are reasonably well predicted even though there are instances when there is overprediction of the same in the months of april and may 1993 and in November and December 1993. The high peak flows are underpredicted. There is however better performance in the prediction of surface flows and the total stream flow. **Figure 4.10** shows the comparison of observed and simulated surface flows. The correlation coefficient is 0.61 while the coefficient of efficiency (NSE) showed an improvement compared to comparisons based on total flow yielding a value of 0.34 compared to 0.22 observed in the case of total flow.

Baseflow simulation

The model was also assessed to examine its performance on base flow simulation after the annual flow calibration. **Figure 4.11** shows the hydrographs of simulated and observed base flows after the annual flow calibration. The low base flows are comparable while the high base flows are underpredicted. In general model performance in simulation of base flow is very poor with the coefficient of efficiency being 0.08. The correlation coefficient maintained a reasonable value of 0.52 as illustrated in **Figure 4.12**.

It can be concluded that the relatively poor simulation of the total streamflow (NSE=0.22) after the annual flow calibration may be attributed mainly to the poor simulation of the baseflow (NSE =0.08). The simulation of surface flow is however much better with an improved value of the Nash-Sutcliffe Efficiency (NSE=0.34). A graphical plot of the simulated Total flow, baseflow and surface flow is illustrated in **Figure 4.13**.

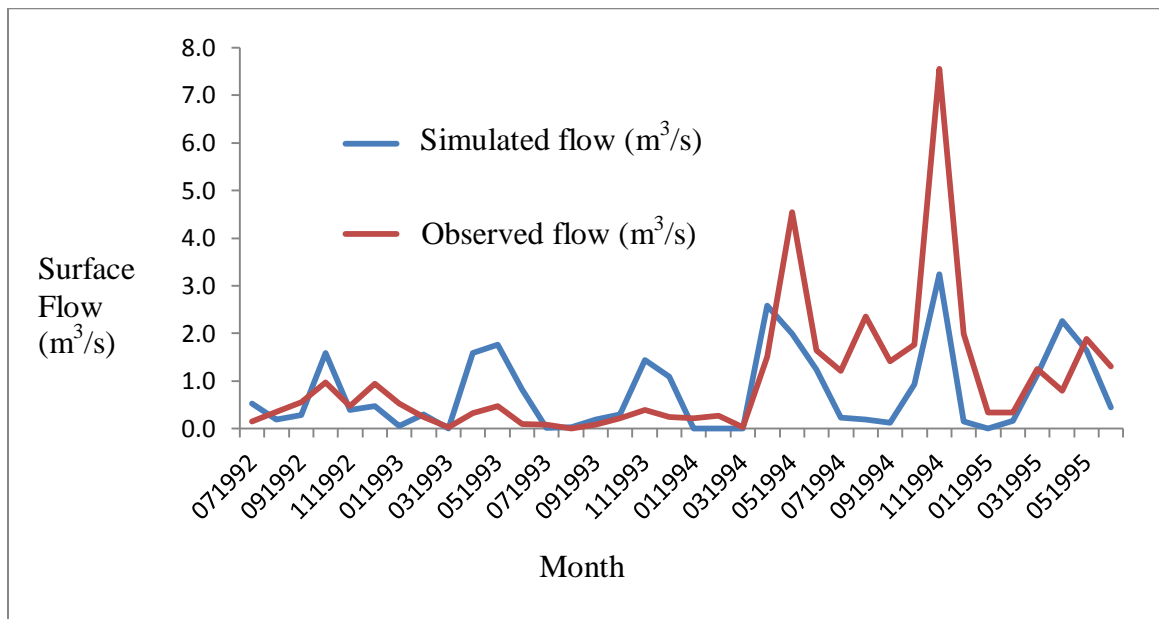


Figure 4.9. Hydrograph of observed and simulated monthly surface flows after annual flow calibration

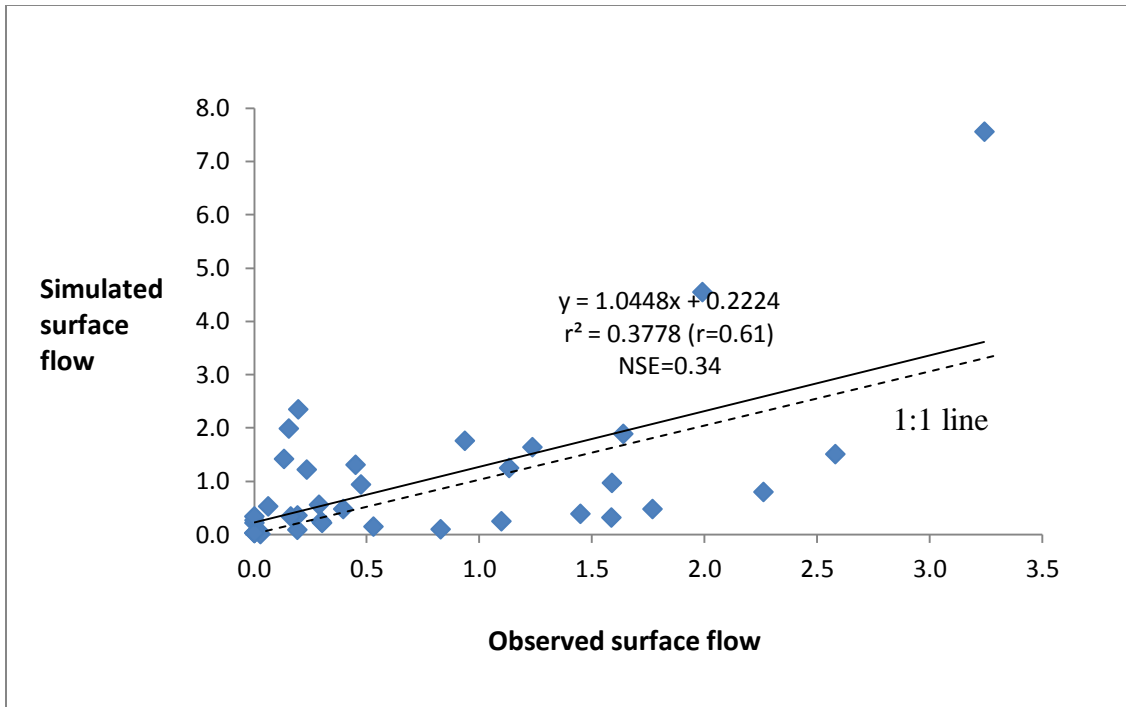


Figure 4.10. Comparison of simulated and observed average daily surface flow in month after calibration of average annual water yields

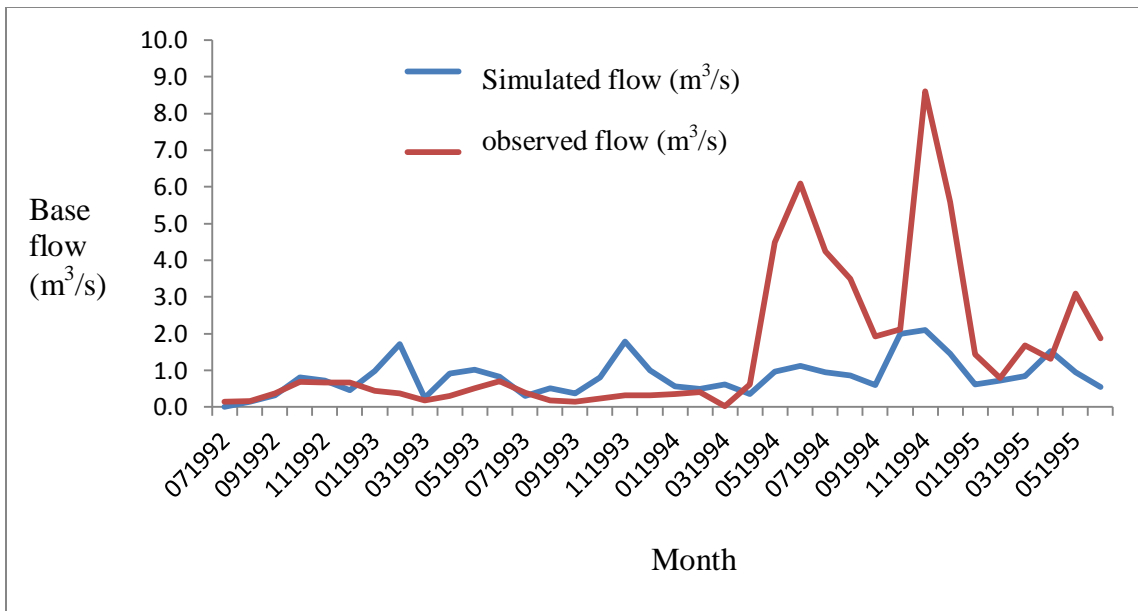


Figure 4.11 Hydrograph of observed and simulated monthly base flows after annual flow calibration

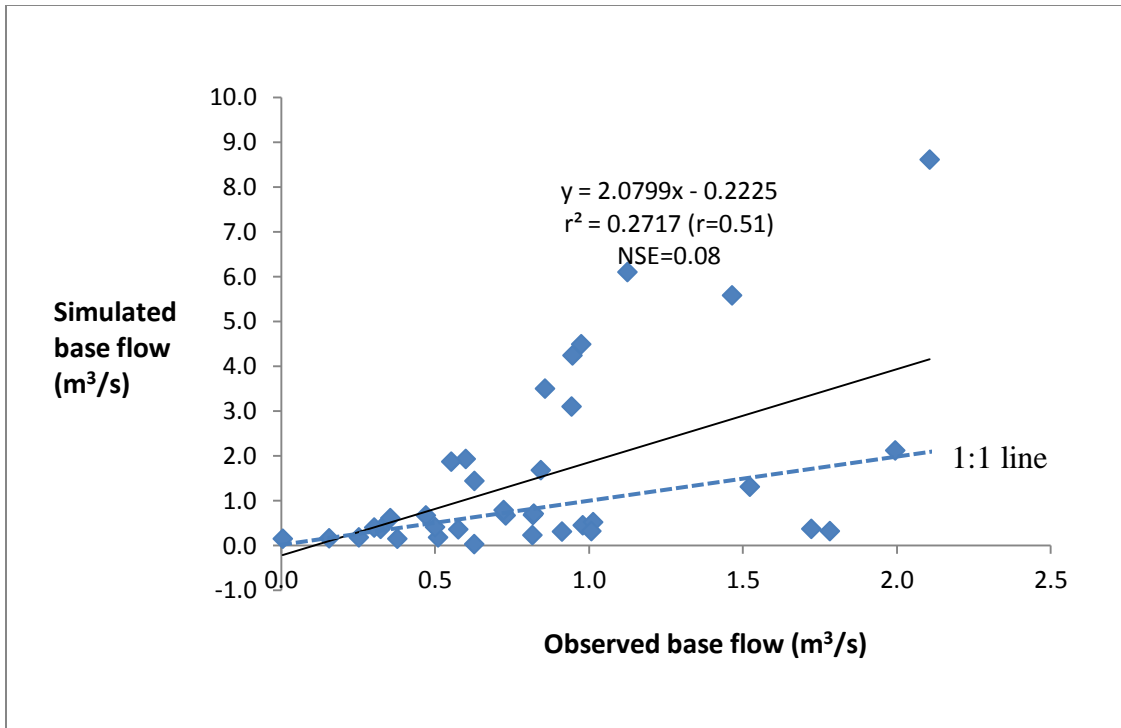


Figure 4.12 Comparison of simulated and observed daily base flow in month after calibration of average annual water yields

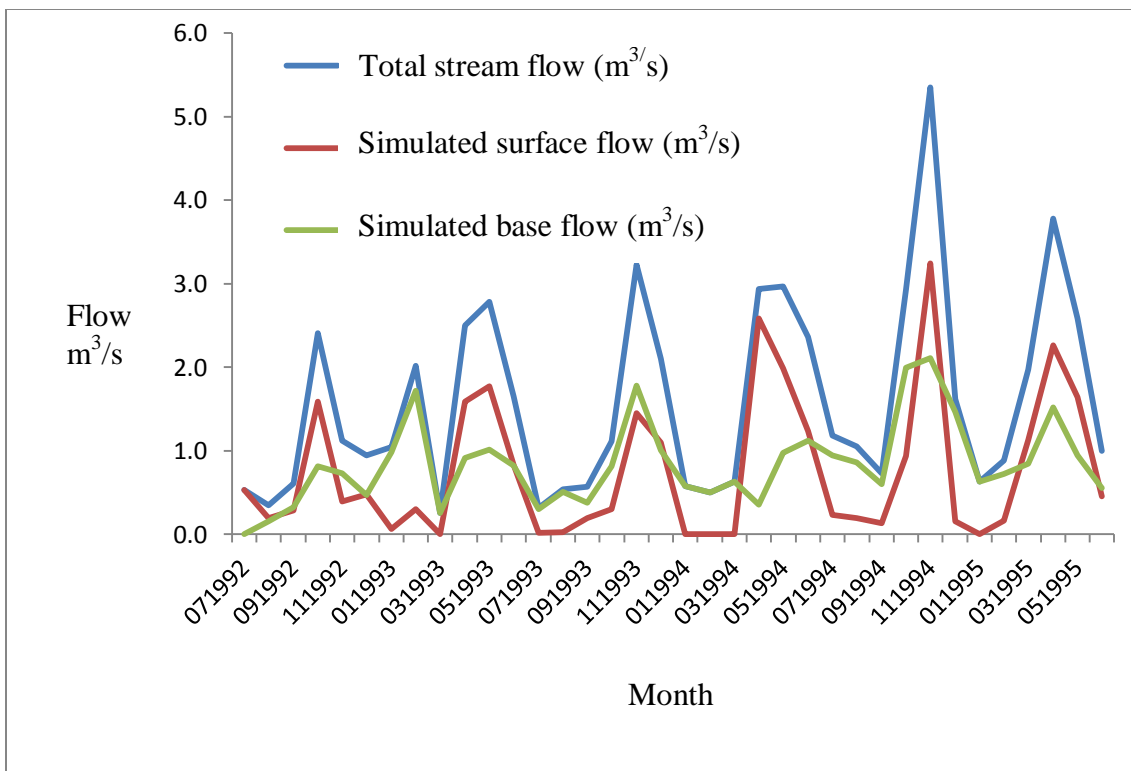


Figure 4.13 Hydrographs of simulated monthly total flow, base flow and surface flow after annual flow calibration.

4.5.2.3 Monthly flow calibration for improved simulation of surface runoff

After performing the annual calibration, it was noted that the model is a poor simulator of base flow; however the simulation of the surface flow was favourable. In order to fine tune the calibration for surface flow prediction, further adjustment of selected input parameters was carried out. In order to further increase the surface runoff, the soil available water capacity (SOL_AWC) was adjusted by reducing it while the curve number was further adjusted upwards while comparing the observed and simulated surface runoff. By so doing there is a significant increase in surface runoff thereby improving the surface runoff simulation. On adjustment of the SOL_AWC downwards by 0.03mm/mm from the default value and slightly raising the curve number, there was an improvement on the surface flow simulation with value coefficient of efficiency rising to about 0.4 while the correlation coefficient still maintaining a value of 0.62 which is acceptable. The deviation volume was significantly reduced to a value of 5.47% with percentage bias being 5.49, while the difference between the means of the surface flows was within 5% of the observed surface flow with the average observed daily surface flow in month being $1.02\text{m}^3/\text{s}$ and the simulated surface flow $0.96\text{m}^3/\text{s}$. The root mean square error (RMSE) was 6.76 while the RSR value was 0.79. These values of evaluation statistics indicate good model performance in predicting surface runoff for the catchment.

Model validation for surface runoff prediction

After calibrating the model for monthly surface flow, monthly simulation was performed during the period 1/1/1998 to 31/12/2000 which is outside the calibration period to establish the performance of the model in predicting surface runoff after the calibration exercise. The model input parameters were the same as during the calibration period. **Figure 4.14** illustrates the monthly time series for modeled and observed surface flow during the validation period. There is an improvement in the simulation of the surface flow. The model simulates the flow patterns reasonably well. The peaks and recessions are reasonably well presented with isolated tendencies of the model to overpredict a few peak flows during this low flow period while under predicting the surface flows in some instances, however these levels of under prediction or overprediction does not appear very significant based on visual analysis. During some periods, the simulated and

observed flows fitted perfectly well. Regression analysis yielded a correlation coefficient value (r) of 0.8 i.e. ($r^2=0.64$) with the mean daily flows in month of measured and simulated surface flows of $0.958\text{m}^3/\text{s}$ and $0.905\text{m}^3/\text{s}$ (within 6% of measured surface flow). **Figure 4.15** shows the regression analysis results in which the value of Nash Sutcliffe Coefficient is also indicated being 0.62 reflecting good model performance.

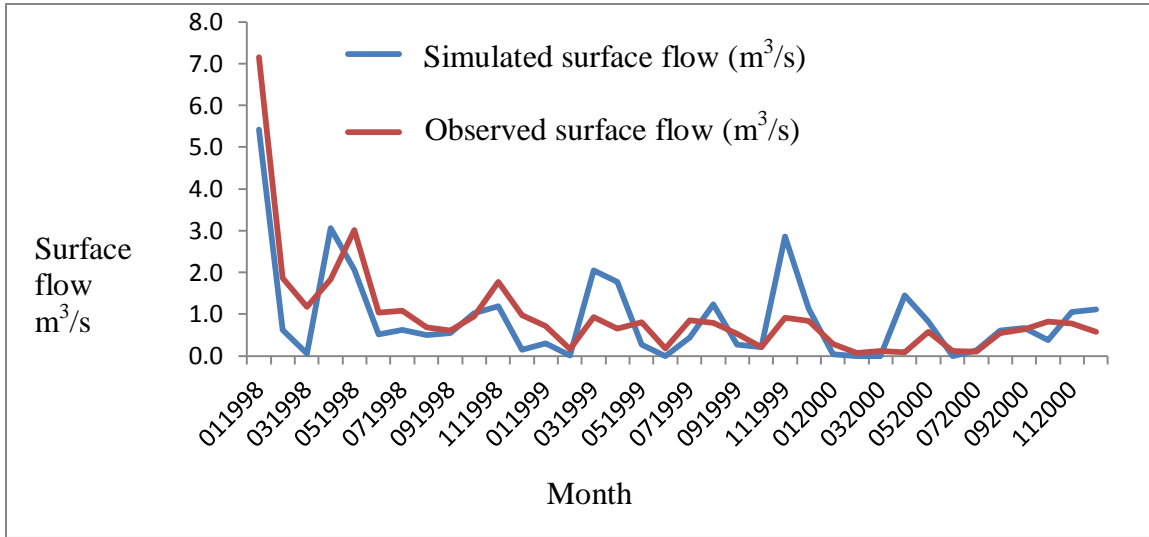


Figure 4.14. Monthly time series of simulated and observed surface flows in the Naro Moru river catchment in the period 1/1/1998 to 31/12/2000

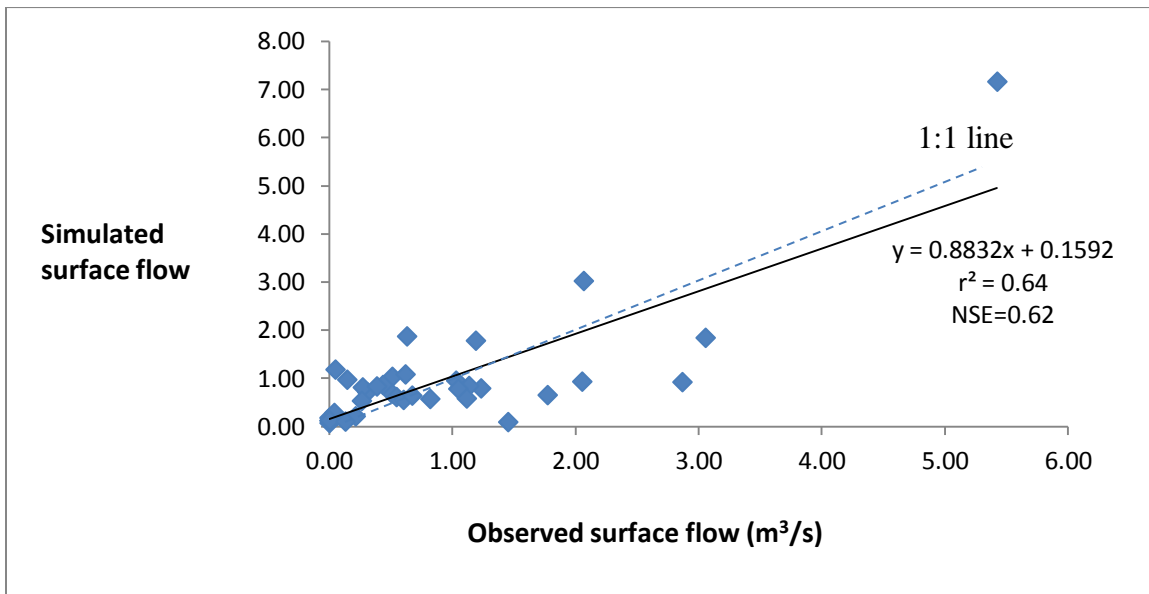


Figure 4.15. Regression of simulated average daily surface flow in month and observed surface flows during validation

The results of model performance for surface flow simulation is summarized in **Table 4.11** indicating values of various evaluation statistics during calibration and validation periods.

Table 4.11: Comparison of model performance in monthly surface flow simulation after total flow calibration, monthly surface flow calibration and validation

Monthly simulated surface flow compared to measured values			
<u>Evaluation Statistic</u>	After average annual Streamflow Calibration	After Monthly surface flow calibration	After Validation
Nash-Sutcliffe efficiency (NSE)	0.34	0.37	0.62
Correlation Coefficient (r)	0.38	0.61	0.8
Coefficient of Determination (R ²)	0.61	0.38	0.64
Deviation Volume (D _v)	25.19	5.80	5.58
Percentage bias (PBIAS)	25.30	5.85	7.04
RSR	0.81	0.79	0.61
RMSE	6.91	6.88	4.44

4.5.3 Summary of model performance assessment during manual calibration.

Model simulation during the manual calibration showed modest performance after surface flow calibration. Low values of R² were observed during the calibration period. Various reasons can be attributed to this dismal performance. It may be attributed partly to the model and partly to the input data. High flows observed during part of the calibration period indicated that the peak flows were grossly underpredicted thereby contributing to the overall low value of R² obtained (**Figures 4.5 to 4.11**). In some cases, the peak observed flow was nearly double the simulated value (May and November 1994). This may be partly attributed to the poor simulation of the base flow processes by

the model. The base flow pattern of the hydrograph does not follow the general pattern of flow in this month (**Figure 4.13**). K_s is an input parameter that influences the ground water flow processes especially flow in aquifers. This parameter has to be estimated from a programme (Soil Water Characteristics) which may underestimate or overestimate the values thereby grossly misrepresenting the actual baseflow and the subsequent flow prediction. The manner in which the model estimates peak flows also differs from the observations. It is possible that the observed data consistently captured the peak flow during the month especially where the high flows occurred close to the time when the gauge reading was taken. This would yield relative higher observed peak flow than simulated. Another possible reason is that of the unrepresentative input rainfall data. The input rainfall used in the simulations are approximations of the actual rainfall which may be overestimated or underestimated. The model uses the thiesen polygon method in rainfall estimation which is based on weighting of the stations used according to their assumed areas of influence. This method does not account for relief effects and is also approximate. If the rainfall is overpredicted, then the resulting flow would be higher than what would be observed and vice versa explaining the discrepancies between observed and predicted flows. Slope is an input parameter in the model being one among the parameters to which the model is sensitive ranking 7 in position out of the input 28 parameters. The Digital Elevation Model input in the model set up is representative of the slope. The DEM was derived from digitized contours which uses interpolation between contours to derive the DEM. This approximation may also affect the model predictions and hence low r^2 . The curve number to which the model is most sensitive relies on land use information for its determination. The tabulated land uses for determining the curve numbers may not necessarily have same descriptions as those for the catchment and so are approximations to the actual land use types. This may result into model under predictions or over predictions of surface flows compared to the simulations yielding low r^2 values. The model also requires the observed landuses to be classified according to the SWAT landuses before the model load the data. The corresponding land use in SWAT is an approximation of the actual observed land use and may yield runoff that could over predict or underpredict the flow causing some discrepancy and therefore low r^2 value when observed and predicted flow values are compared. Limitation in data availability

for model inputs may also contribute. In some instances, the observed meteorological and flow data available may be characterised with gaps which are filled by statistical approximations hence in overall, the data are approximations. This is inevitable considering data deficiency in developing countries like Kenya. In brief, the poor performance of model performances is contributed both by the model limitations and input data deficiencies.

4.6 Evaluation of selected pedotransfer functions in estimating saturated hydraulic conductivity (K_s) based on model performance in surface runoff prediction.

The Soil Characteristics Program used in the computation of saturated hydraulic conductivity has an inbuilt pedotransfer function designated as Saxton86 and Saxton 2006 which was used in the model set up, trial run, calibration and validation. The input data required in the calculation of PTFs depended on the pedotransfer function in question and ranged from simple texture data on % sand, % silt and % clay for some functions while other required additional data that included bulk density, porosity, organic matter, etc. Some of the pedotransfer functions required parameters derived from the soil moisture characteristic functions that could be estimated using available texture data. **Table 4.12** shows the values calculated for K_s using the selected pedotransfer functions.

Table 4.12 Calculated values of saturated hydraulic conductivity (K_s) based on pedotransfer functions evaluated in the study

Soil	TopDep	BotDep	SDTO	STPC	CLPC	BULK	AWC	CEC	OM	saxton06	Pucket	Dane	Saxton86	Brakensiek	Campbell	Jabro	Young
NTu	0	20	32	46	22	1.28	12.00	42.10	32.72	41.95	2.04	12.79	65.43	4.82	63.86	42.99	8.01
NTu	20	40	32	48	20	1.30	12.00	38.88	28.12	48.43	3.02	17.06	80.47	4.52	76.46	33.54	7.70
NTu	40	60	30	48	22	1.30	9.00	37.60	19.80	42.96	2.04	12.79	68.15	4.09	55.52	35.99	4.84
NTu	60	80	32	56	12	1.35	9.00	49.10	12.92	81.69	14.67	53.97	203.11	4.16	157.08	17.35	4.83
NTu	80	100	32	56	12	1.35	12.00	42.90	9.64	81.69	14.67	53.97	203.11	4.16	157.08	17.35	3.66
HSs	0	20	-1	-1	-1	0.36	35.00	15.00	137.76	0.00	0.00	0.00	0	0.00	0.00	0.00	0.00
HSs	20	40	-1	-1	-1	0.26	39.00	11.25	86.10	0.00	0.00	0.00	0	0.00	0.00	0.00	0.00
HSs	40	60	-1	-1	-1	0.18	43.00	10.00	68.88	0.00	0.00	0.00	0	0.00	0.00	0.00	0.00
PHI	0	20	28	16	56	1.10	11.00	22.00	25.83	0.76	0.00	0.10	10.78	10.88	2.26	800.31	5.08
PHI	20	40	24	14	62	1.20	10.00	23.00	21.53	9.55	0.00	0.04	11.68	2.89	1.00	362.41	1.88
PHI	40	60	18	12	70	1.30	13.00	24.25	12.66	12.83	0.00	0.01	13.58	0.34	0.32	193.03	0.35
PHI	60	70	16	12	72	1.30	13.00	24.50	8.09	14.07	0.00	0.01	14.28	0.27	0.23	219.47	0.15
PHI	0	20	34	24	42	1.10	11.00	26.60	33.75	5.40	0.04	0.72	14.15	12.78	12.14	466.36	9.97
PHI	20	40	32	26	42	1.10	10.00	26.60	27.69	6.11	0.04	0.72	14.82	11.81	10.56	466.94	7.99
PHI	40	60	32	18	50	1.20	9.00	28.20	12.57	1.64	0.01	0.23	10.99	5.32	5.14	216.08	1.10
PHI	60	80	36	17	47	1.20	12.00	25.75	6.03	1.46	0.01	0.35	7.49	6.73	8.91	199.05	0.40
PHI	80	100	44	16	40	1.20	12.00	25.60	7.23	3.54	0.06	0.96	10.89	10.11	29.28	168.00	0.78
VRe	0	20	28	22	50	1.30	12.00	35.15	23.87	2.54	0.01	0.23	12.09	1.53	3.88	72.99	1.36
VRe	20	40	22	12	66	1.40	13.00	25.80	12.56	4.41	0.00	0.02	12.28	0.08	0.60	53.29	0.21
VRe	40	60	22	12	66	1.40	13.00	25.80	12.56	4.41	0.00	0.02	12.28	0.08	0.60	53.29	0.21
VRe	60	80	22	15	63	1.40	14.00	27.67	6.36	3.33	0.00	0.03	10.26	0.11	0.79	44.48	0.07
VRe	80	100	22	16	62	1.40	15.00	24.90	3.87	1.97	0.00	0.04	5.28	0.12	0.87	42.21	0.03
PHI	0	20	20	29	51	1.20	11.00	29.00	63.71	4.62	0.01	0.20	7.68	2.92	2.03	245.10	6.69
PHI	20	40	20	20	60	1.33	11.00	20.00	17.22	1.22	0.00	0.05	12.86	0.40	0.90	82.57	0.87
PHI	40	60	20	20	60	1.29	13.00	20.00	12.92	1.22	0.00	0.05	12.86	0.73	0.90	126.60	0.69
PHI	60	75	20	20	60	1.37	13.00	20.00	8.61	1.22	0.00	0.05	12.86	0.21	0.90	53.86	0.24
LVf	0	20	26	22	52	1.48	6.00	17.50	15.93	2.25	0.01	0.17	12.1	0.11	2.82	11.56	0.56
LVf	20	40	29	15	56	1.51	8.00	10.00	11.19	0.63	0.00	0.10	10.56	0.05	2.43	10.16	0.43
LVf	40	60	25	15	60	1.51	7.00	10.00	8.61	0.49	0.00	0.05	11.39	0.02	1.28	11.95	0.25
LVf	60	70	25	15	60	1.56	10.00	10.00	8.61	0.49	0.00	0.05	11.39	0.01	1.28	7.00	0.19
CMx	0	20	57	18	25	1.28	10.00	16.25	22.82	18.99	1.13	8.30	28.61	9.34	280.55	48.99	8.73
CMx	20	40	50	10	40	1.34	12.00	20.00	13.78	1.63	0.06	0.96	9.59	5.03	44.56	47.92	1.44
CMx	40	60	50	10	40	1.36	11.00	20.00	13.78	1.63	0.06	0.96	9.59	4.29	44.56	38.70	1.31

4.6.1 Model output in Surface flow simulation

On completing the surface flow calibration exercise, the role of various pedotransfer functions in estimating hydraulic conductivity was evaluated to examine the effect on surface flow simulation and therefore establish if indeed the pedotransfer functions used in estimating hydraulic conductivity affects surface flow simulation results during calibration and validation periods. For each pedotransfer function, the calculated value of saturated hydraulic conductivity for each layer for the soil unit in each sub basin was input into the model for all the 27 sub basins into which the catchment was divided into. Other model input parameters were kept constant at the values derived after conclusion of the surface flow calibration process. Based on the saturated hydraulic conductivity values obtained from each pedotransfer function, model simulation was performed and thereafter a comparison made between the observed and simulated surface flows. **Table 4.13** shows the evaluation results during calibration and validation periods for the various pedotransfer functions when used in simulating surface runoff in which the observed and predicted surface runoff were compared for each pedotransfer function for the said periods.

4.6.2 Performance of the PTFs in surface flow prediction

From **Tables 4.13**, it is shows that the values of the performance parameters was different for each pedotransfer function used in estimating saturated hydraulic conductivity (K_s) model input when observed and simulated surface flows were compared. **Table 4.14** shows the performance ratings in surface runoff simulation based on estimation of hydraulic conductivity using various pedotransfer functions. The measures of performance ratings are appended (Appendix 5)

Table 4.13 Values of selected performance measures in model surface flow simulation based on various pedotransfer functions used in estimating K_s

	CALIBRATION							VALIDATION						
	SAX86	PUCKETT	jabro	DANE	Brakensiek	Campbell	Young	SAX86	PUCKETT	jabro	DANE	Brakensiek	Campbell	Young
NSE	0.36	0.41	0.35	0.36	0.35	0.36	0.37	0.63	0.59	0.63	0.62	0.61	0.63	0.62
r	0.61	0.67	0.60	0.62	0.60	0.60	0.61	0.8	0.80	0.80	0.80	0.79	0.80	0.79
O_{av}	1.02	1.02	1.022	1.022	1.02	1.02	1.02	0.96	0.96	0.96	0.96	0.96	0.96	0.96
P_{ave}	1.01	1.25	1.023	1.09	0.96	0.98	0.99	0.90	1.13	0.90	0.97	0.84	0.86	0.87
Dv	0.83	-22.39	-0.26	-6.87	6.19	4.07	3.48	6.20	-17.81	5.80	-1.13	12.12	9.99	9.63
PBIAS	0.84	-22.57	-0.26	-6.92	6.24	4.11	3.51	7.82	-22.48	7.32	-1.43	15.29	12.61	12.16
RMSE	6.93	6.69	6.99	6.95	6.98	6.96	6.92	4.42	4.65	4.42	4.46	4.49	4.42	4.46
RSR	0.80	0.77	0.81	0.80	0.80	0.80	0.80	0.61	0.64	0.61	0.62	0.62	0.61	0.62

Table 4.14 Rating of selected pedotransfer functions in predicting K_s for surface runoff prediction using SWAT model during calibration and validation

PTF	NSE		PBIAS		RSR	
	Calibration	Validation	Calibration	Validation	Calibration	Validation
Saxton86	0.36(Unsatisfactory)	0.63 (satisfactory)	0.84 (very good)	7.82 (Very good)	0.80 (unsatisfactory)	0.61 (satisfactory)
Puckett	0.41 Unsatisfactory)	0.59(satisfactory)	-22.57 (Satisfactory)	-22.48 (satisfactory)	0.77(unsatisfactory)	0.64 (satisfactory)
Jabro	0.35 (Unsatisfactory)	0.63(satisfactory)	-0.26 (very good)	-7.32 (very good)	0.81(unsatisfactory)	0.61(satisfactory)
Dane	0.36 (Unsatisfactory)	0.63 (satisfactory)	-6.92(very good)	-1.43 (very good)	0.80 (unsatisfactory)	0.62 (satisfactory)
Brakensiek	0.35 (Unsatisfactory)	0.61(satisfactory)	6.24 (very good)	15.29 (satisfactory)	0.80 (unsatisfactory)	0.62 (satisfactory)
Campbel	0.36 (unsatisfactory)	0.63 (satisfactory)	4.11(very good)	12.61 (good)	0.80 (unsatisfactory)	0.61(satisfactory)
Gowing	0.37 (unsatisfactory)	0.62 (satisfactory)	3.51(very good)	12.16 (good)	0.80 (unsatisfactory)	0.62(satisfactory)

Performances during calibration and validation were assessed as follows:

During calibration

The values of correlation coefficient ranged between 0.6 and 0.67 for all the pedotransfer functions which are above the acceptable value of 0.5. The simulation based on Puckett yielded the highest value of Nash-Sutcliffe efficiency (NSE=0.41) and also the highest value for the correlation coefficient. In general, the values of NSE returned for all the simulation sets indicated unsatisfactory model performance all falling in the range 0.35 to 0.41. The difference between the measured and observed surface flow was within 22.5% of the observed for Puckett. The PBIAS even though satisfactory was highest for Puckett compared to the others being -22.57, with the negative value indicating model overestimation of surface flow, however being less than 25% indicating satisfactory model performance (Dawadi and Ahmad, 2012). The RMSE and RSR also registered the lowest values compared to the other PTFs indicating better performance. The next best value for NSE was that obtained based on the use of pedotransfer function of Young with a value of 0.37. The correlation coefficient was 0.61 (acceptable) while the simulated surface flow was within 6% of the observed surface flow indicating good performance. The percentage PBIAS was 3.51 implying very good model performance based on the use of this pedotransfer function better than what was obtained from Puckett. The RSR value was 0.8 indicating unsatisfactory performance though not very far from the value of 0.7 for which the model performance would be satisfactory. The other pedotransfer functions' performance fell more or less between the two discussed, however more excellent results were obtained from Jabro based on the PBIAS value which was -0.26 i.e $-10% < \text{PBIAS} < 10%$ implying very good model performance. The difference between the means of observed and predicted surface flows was very minimal being 0.1%.

During Validation

The validation results generally indicated good performance in surface flow simulation for all sets of the pedotransfer functions. The value of NSE varied for each pedotransfer function used but the range was within 0.59 to 0.63 reflecting satisfactory model performance for all the pedotransfer functions. The highest value obtained was for

Campbell, Saxton86 and Jabro while the lowest value was obtained for Puckett. The correlation coefficient was about 0.80 for all the functions which is acceptable and reflecting good performance. The values of RSR fell in the range 0.6 and 0.7 within which the model performance is regarded satisfactory. This was the case with the other performance measures. Relative to other pedotransfer functions, Puckett registered a poorer value of PBIAS (-0.18) followed by Brakensiek, however the values still fell within the level of satisfactory performance. The difference between the observed and simulated surface flows varied depending on the pedotransfer function used in the validation but all fell within the range of 7% and 18% being lowest for saxton86 and Jabro while the highest difference occurred when Brakensiek was used in the simulation. From **Table 4.14**, it is observed that the general performance of the model in surface flow simulation ranged from satisfactory to very good during validation period for all the pedotransfer functions used to predict K_s as input to the SWAT model in predicting the surface flow.

4.7 Development of pedotransfer functions

Processes involved in developing pedotransfer functions yielded results discussed as follows;

4.7.1 Evaluation of the distribution of the moisture retention parameters.

Table 4.15 shows the values of these statistical parameters for the response variables being the Van Genuchten moisture retention parameters and their proposed transformations for the data set of 457 samples. The transformations of Van Genuchten parameters (response variables) that produce the best approximation to normal distribution are $\frac{1}{\theta_s}$, e^{θ_r} , $\sqrt{\alpha}$ and \sqrt{n} . This was based on measures of the Shapiro-Wilk (W-value), Skewness coefficient and Kurtosis. For illustration, **Figure 4.16** shows the histogram and normality plot in the normal distribution check for the saturated soil moisture content transformation (θ_s). The statistical parameters associated with the test for normality are illustrated in **Table 4.16** after analysis using the “analyse it” software embedded in microsoft excel. The transformations were then related to the available basic soil properties to develop the pedotransfer functions using multiple linear regression.

Table 4.15 Normality of distribution check for parameters in the van Genuchten water retention characteristic and their transformations:

	W-value	Skewness	Kurtosis
θ_s	0.98	0.53	0.00
θ_r	0.99	0.09	-0.21
α	0.67	4.08	29.47
n	0.78	2.65	11.13
$\ln(\theta_s)$	0.99	0.04	-0.39
$\sqrt{\theta_s}$	0.99	0.28	-0.29
$\frac{1}{\theta_s}$	0.98	0.06	0.06
e^{θ_s}	0.96	0.82	0.66
θ_s^2	0.93	1.05	1.25
$\ln(\theta_r)$	0.75	-4.16	38.54
$\sqrt{\theta_r}$	0.93	-1.09	1.77
$1/\theta_r$	0.03	21.09	444.92
θ_r^2	0.89	1.34	2.29
e^{θ_r}	0.99	0.39	0.12
$\theta_r^{1/3}$	0.83	-2.02	5.78
$\ln(\alpha)$	0.72	-2.94	11.8
$1/\alpha$	0.03	21.17	450.93
e^α	0.62	5.01	44.15
α^2	0.20	13.89	238.13
$\sqrt{\alpha}$	0.95	0.91	2.27
$\text{Log}(n)$	0.98	0.49	0.94
\sqrt{n}	0.91	1.45	3.67
n^2	0.46	6.23	56.04
$1/n$	0.93	1.32	3.93
e^n	0.24	12.69	201.48

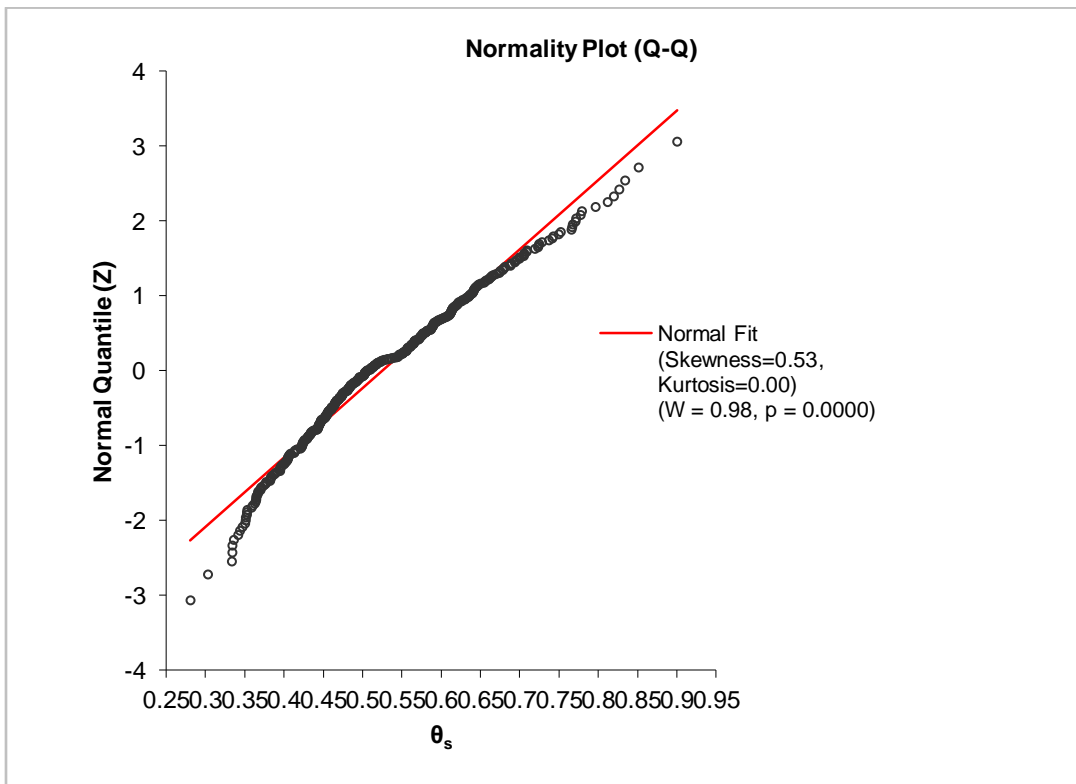
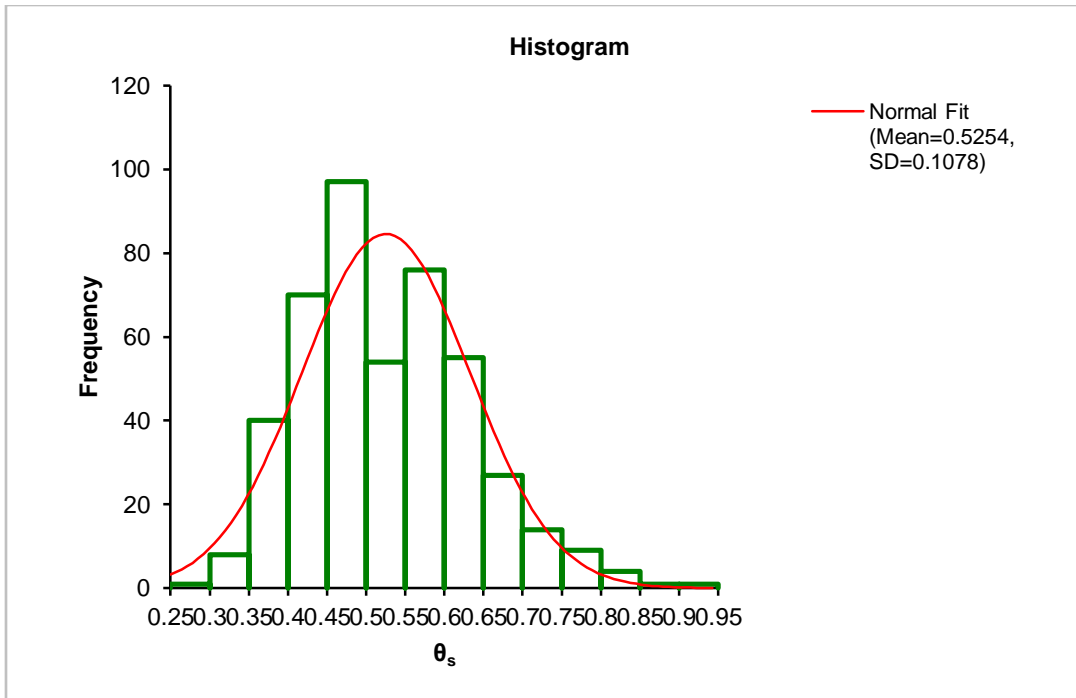


Figure 4.16. Check for normality of distribution of the saturated soil moisture content (θ_s) showing the histogram and normality plot.

Table 4.16. Statistical parameters obtained on normality check for the saturated hydraulic conductivity data

n	457		
Mean	0.5254	Median	0.5053
95% CI	0.5155 to 0.5353	95.1% CI	0.4940 to 0.5206
SE	0.00504		
Variance	0.0116	Range	0.620
SD	0.1078	IQR	0.1497
95% CI	0.1012 to 0.1153	Percentile	
CV	20.5%	0th	0.2808 (minimum)
Skewness	0.53	25th	0.4467 (1st quartile)
Kurtosis	0.00	50th	0.5053 (median)
Shapiro-Wilk W	0.98	75th	0.5964 (3rd quartile)
p	<0.0001	100th	0.9003 (maximum)

The plots of normality check for several response variables (moisture retention parameters) or their transformation are appended for illustration (Appendix).

4.7.2 Statistical linear regression analysis

Tables A6.1 through A6.5 (Appendix 6) shows the resulting values of r^2 when a linear regression analysis is performed for the corresponding response and predictor variables indicated for each of the moisture retention parameters and the transformed basic soil properties i.e percent clay, percent silt, percent sand and, percent organic carbon and bulk density.

4.7.3 Multiple regression analysis and accuracy of the pedotransfer functions developed

The quality of fit for the linear regression between each selected transformed dependent variables (moisture retention parameters) and the transformed independent variables (basic soil properties) was notably different for each transformation. For instance, the

linear regression performed between θ_s and sand yielded a quality of fit, measured by the value of r^2 , to be 0.10 while that between θ_s and e^{sand} yielded $r^2 = 0.01$. The same observation was made in the case of regression performed with silt, clay, bulk density and organic carbon. Also each selected transformed variable (i.e $\ln(\theta_s)$, e^{θ_r} , $\sqrt{\alpha}$, etc) produced different values of r^2 when a linear regression between each response variable and a transformed predictor variable for each of the variable sand, silt, clay bulk density and organic carbon. For example, a regression performed between the independent variable sand^2 and $\frac{1}{\theta_s}$ yielded $r^2 = 0.14$ while that between sand^2 and e^{θ_r} produced $r^2 = 0.20$, hence no particular transformed variable could be said to generally give the best quality of fit with all the dependent variables whether it is sand, silt, clay, bulk density or organic carbon. Also, there is no particular response variable that could be said to yield the best quality of fit compared to others for all the independent variables or transformations to which regression was performed. **Table A6.1 through A6.5** showed the values of quality of fit measured by r^2 for all the selected response variables and the corresponding predictor variable and their transformations. The purpose of this regression analysis was to help identify the pair of transformed dependent and independent variables that yield the best quality of fit between them as measured by coefficient of determination r^2 . This would assist, for each dependent variable, to identify which transformations of each of the dependent variables would be used in developing the pedotransfer functions during multiple regression to relate each response variable and selected predictor variables and appropriate transformations. This would help establish the best possible mathematical relationship between the moisture retention parameters (response variables) and selected basic soil properties (independent variables) in the pedotransfer function. From the **Tables A6.1 through Table A6.5**, a number of observations on the regression relationships can be made for each response variable and how it relates with each of the predictor variables and their transformations. The transformed response variable $\frac{1}{\theta_s}$ yielded the best quality of fit ($r^2=0.14$) with the independent variables *sand* or sand^2 compared to the other transformations when a regression analysis was performed. In the case of regression with silt, the best possible quality of fit was obtained with the transformation $1/\text{silt}$ and so was with bulk density, BD^3 ($r^2=0.60$), $\sqrt{\text{clay}}$ or clay^2

($r^2=0.12$), organic carbon, $\ln(\text{OrgC})$ ($r^2=0.27$). In general, the linear regression performed between $\frac{1}{\theta_s}$ and the bulk density transformations produced best quality of fit as compared to those of silt, sand, organic carbon and clay, showing that the best relationship was obtained with bulk density followed by organic carbon, sand, clay and silt in that order. In the case of residual moisture content, θ_r , the selected transformation e^{θ_r} related best with clay with a quality of fit $r^2=0.32$ obtained when regression was performed between e^{θ_r} and $\sqrt{\text{clay}}$. The poorest fit was obtained when the regression was performed with transformations of organic carbon e.g. $\ln(\text{OrgC})$ with $r^2=0.01$. α yielded the best possible quality of fit when a regression was performed with $\sqrt{\text{clay}}$ or $1/\text{clay}$. $\sqrt{\alpha}$ generally showed poor quality of fit with most of the independent variables or transformations. The transformation $\frac{1}{n}$ also showed poor quality fit with most of the selected response variables when regression analysis was performed. The best quality of fit was obtained with the transformation of sand when the linear regression was performed between n^{-1} and sand^3 ($r^2=0.09$). These regression results are shown in **Tables A6.1 through A6.5** (Appendix 6). **Figure 4.17** illustrates the regression between selected transformations of response variables and the predictor variables that give the best possible quality fit .

After performing linear regression between two variables involving the selected transformed response variables, associated with the moisture retention parameters, and the predictor variables associated with the basic soil properties, the following conclusions were arrived at. For the transformed response variable $\frac{1}{\theta_s}$ the predictor variables that yielded the best possible quality of fit with it when linear regression was performed were *sand*, $1/\text{silt}$, BD^3 , clay^2 , and $\ln(\text{OrgC})$ (**Tables A6.1 to A6.5**-Appendix 6). The variables would then be used in developing the multiple linear regression between the said response variable and the predictor variables.

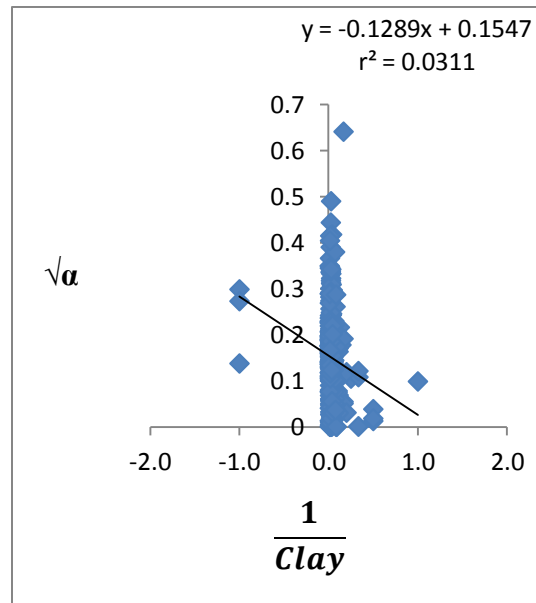
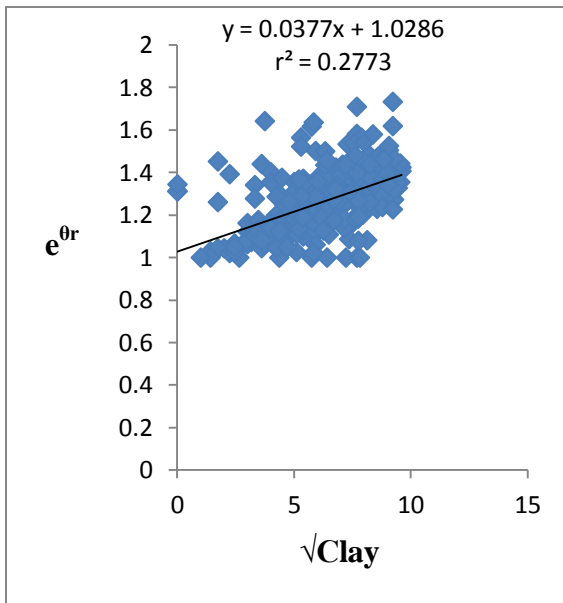
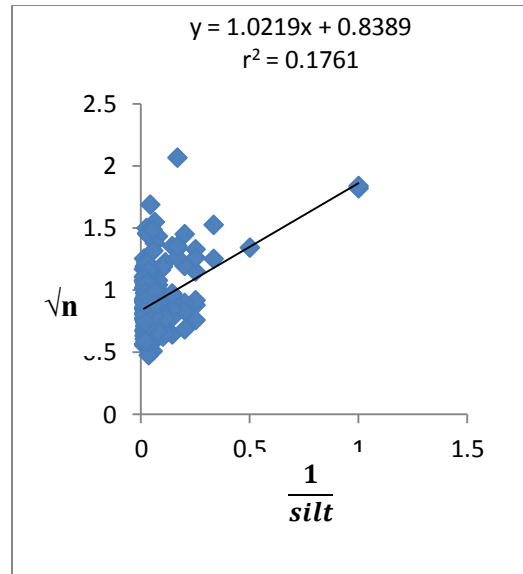
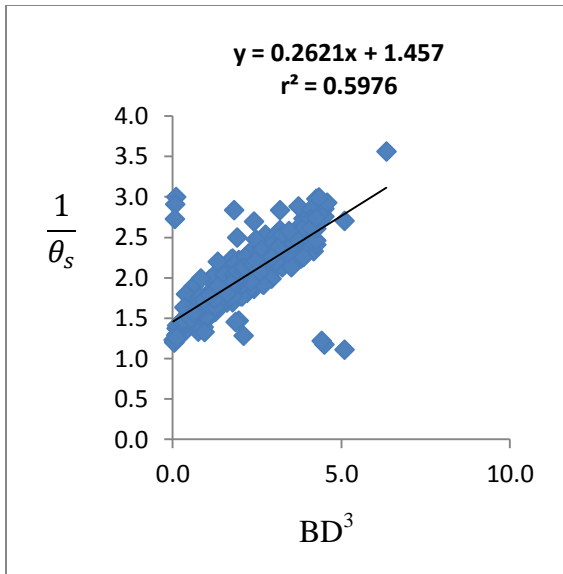


Figure 4.17: Response versus predictor variable correlations with corresponding transformations that give best possible quality fit.

In the case of e^{θ_r} , the best possible fit (measured by the value of r^2) was obtained from linear regression with *sand*, *silt*¹⁰, *BD*², $\sqrt{\text{clay}}$, and $\ln(\text{OrgC})$, while in the case of $\sqrt{\alpha}$ the best possible fits were obtained with e^{sand} , *silt*⁻², *BD*, *clay*⁻¹, and $\ln(\text{OrgC})$. In the case of \sqrt{n} , the best possible quality of fit was obtained with the predictor variables *Sand*⁵, $\frac{1}{\text{silt}}$, $\ln(\text{clay})$, *OrgC*². The predictor variables indicated above that related best in linear regression with the selected response variables were the ones considered in developing the pedotransfer functions using multiple linear regression. **Table 4.17** shows the quality of fit obtained when multiple linear regression was performed between each selected response variable and the transformed predictor variables to which it relates best e.g. a multiple linear regression performed between the response variable $\frac{1}{\theta_s}$ and *sand*, $1/\text{silt}$, *BD*³, *clay*², $\ln(\text{OrgC})$, yielded the quality of fit measured by $R^2 = 0.63$ which is considered to be of acceptable accuracy. The resulting equations are illustrated below.

$$\frac{1}{\theta_s} = 1.477 + 0.003045\text{SAND} - 0.1505 \frac{1}{\text{silt}} + 0.2126\text{BD}^3 - 9.3857\text{E-}006\text{clay}^2 - 0.04542\text{Ln}(\text{OrgC}) \quad (4.1)$$

Where:

θ_s = saturated soil moisture content (cm^3/cm^3)

silt = percent silt

sand = percent sand

BD = bulk density (g/cm^3)

Clay = percent clay

OrgC = Organic carbon content (g/kg)

The qualities of fit for the remaining transformed response variables are illustrated in **Table 4.17** (column 2) for each of e^{θ_r} , $\sqrt{\alpha}$ and \sqrt{n} when the simple linear regression was done with the selected predictor variable transformations.

Table 4.17: Quality of fit based between response variable transformations and selected predictor variable transformations in the multiple linear regression.

	$\text{sand} + \text{silt}^{-1} + \text{BD}^3 + \sqrt{\text{clay}} + \ln(\text{OrgC})$	$\text{sand} + e^{(\text{BD})}$	$\text{sand} + 1/\text{silt} + \text{BD}^3$	$\sqrt{\text{clay} + \text{BD}^3 + \text{silt}^{-1}}$		$\text{sand} + \text{BD}^3$
θ_s^{-1}	$R^2=0.63$	$R^2=0.63$	$R^2=0.63$	$R^2=0.61$		$R^2=0.63$
	$\text{sand} + \text{silt}^{10} + \text{BD}^2 + \sqrt{\text{clay}} + \ln(\text{OrgC})$	$\text{sand} + \sqrt{\text{clay}}$	$\sqrt{\text{clay}} + \text{silt}^{10} + \text{BD}^2$	$\sqrt{\text{clay} + \text{silt}^{10} + \ln(\text{orgc})}$	$\sqrt{\text{clay} + \text{BD}^2}$	$\sqrt{\text{clay} + \ln(\text{OrgC})}$
$e^{\theta r}$	$R^2=0.37$	$R^2=0.33$	$R^2=0.35$	$R^2=0.32$	$R^2=0.35$	$R^2=0.32$
	$e^{\text{sand} + \text{silt}^{-2} + \text{BD} + \text{clay}^{-1}} + \ln(\text{OrgC})$	$e^{\text{sand} + \text{BD} + \text{clay}^{-1}}$	$e^{\text{sand} + \text{clay}^{-1}} + \ln(\text{OrgC})$	$\text{silt}^{-2} + \text{BD} + 1/\text{clay} +$	$\text{BD} + 1/\text{clay}$	$\text{silt}^{-2} + \text{clay}^{-1}$
$\sqrt{\alpha}$	$R^2=0.05$	$R^2=0.05$	$R^2=0.04$	$R^2=0.05$	$R^2=0.04$	$R^2=0.04$
	$\text{sand}^5 + 1/\text{silt} + \ln(\text{clay}) + \text{OrgC}^{-2}$	$\text{sand}^5 + 1/\text{OrgC}^2$	$1/\text{silt} + \ln(\text{clay})$	$\ln(\text{clay}) + 1/\text{OrgC}^2$		
\sqrt{n}	$R^2=0.24$	$R^2=0.21$	$R^2=0.23$	$R^2=0.12$		

Based on the above analysis, the general equations that can be used to determine the Van Genuchten parameters are illustrated below.

$$\frac{1}{\theta_s} = a + b_{sand} + \frac{c}{silt} + dBD^3 + eClay^2 + f \ln(OrgC) \quad (4.2)$$

$$e^{\theta_r} = a + b_{sand} + c_{Silt}^{10} + dBD^2 + e\sqrt{clay} + f \ln(OrgC) \quad (4.3)$$

$$\sqrt{\alpha} = a + b e^{sand} + \frac{c}{silt^2} + dBD + \frac{e}{clay} + f \ln(OrgC) \quad (4.4)$$

$$\sqrt{n} = a + b_{sand}^5 + \frac{c}{silt} + d \ln clay + \frac{e}{OrgC^2} \quad (4.5)$$

Where a, b, c, d e and f are constants.

4.7.4 Analysis of cross correlations and correlation matrix

Cross correlation was done for the independent variables (transformations) used for each of the response variables $\frac{1}{\theta_s}$, e^{θ_r} , $\sqrt{\alpha}$ and \sqrt{n} . **Tables A7.1 through A7.4 (Appendix)** Shows the relationships, based on correlation coefficient, between the various predictor variables for each of the response variables or transformations. In the case of $\frac{1}{\theta_s}$ the predictor variables *sand* and \sqrt{clay} appear to be highly negatively correlated with $r = -0.71$ while BD^3 and $\ln(OrgC)$ are also fairly correlated in a negative sense ($r = -0.59$). For the dependent variable e^{θ_r} , the response variables BD^3 and $\ln(OrgC)$ are also closely correlated in the negative sense with $r = -0.62$. Other reasonable correlations include that between e^{sand} to $\frac{1}{silt^2}$ ($r = 0.89$), BD to $\ln(OrgC)$ ($r = -0.64$) in the case of predictor variables for $\sqrt{\alpha}$. The other significant correlations observed is that between $sand^5$ to $1/silt$ ($r = 0.74$) and $sand^5$ to $\ln(Cl原因)$ ($r = -0.69$) in the predictor variables for \sqrt{n} .

4.7.5 Multiple linear regression results

For each of the selected response variables, **Table 4.17** showed the various possible combinations of the predictor variables that were used in performing the multiple linear regression and also indicates the qualities of fit obtained based on the coefficient of multiple determination R^2 . The various possible multiple linear regression equations obtained for the various independent variables are under listed. The choice of the independent variables or their transformations was based on combinations that would yield the highest possible quality of fit i.e the predictor variable combinations chosen include the variables that relate very well with the response variable, hence the equations are the best possible equations that could be obtained using the variables indicated.

Equations developed involving $\frac{1}{\theta_s}$ are as follows with the measures of accuracy based on coefficient of determination (R^2) indicated:

$$\frac{1}{\theta_s} = 0.813 + 0.003534sand + 0.3099e^{BD} (R^2=0.63) \quad (4.6)$$

$$\frac{1}{\theta_s} = 1.397 + 0.00317sand + 0.2451BD^3 (R^2=0.63) \quad (4.7)$$

$$\frac{1}{\theta_s} = 1.397 + 0.003385sand - 0.1433 \frac{1}{silt} + 0.246BD^3 (R^2=0.63) \quad (4.8)$$

$$\begin{aligned} \frac{1}{\theta_s} = & 1.477 + 0.003045sand - 0.1505 \frac{1}{silt} + 0.212BD^3 \\ & - 9.3857E - 0.006clay^2 - 0.0452\ln(orgC) (R^2=0.63) \end{aligned} \quad (4.9)$$

The set of possible multiple linear regression equations developed involving e^{θ_r} are listed as follows:

$$e^{\theta_r} = 1.083 + 0.0382\sqrt{clay} - 5.0290E - 0.21silt^{10} - 0.0392BD^2 (R^2=0.35) \quad (4.10)$$

$$e^{\theta_r} = 1.062 - 0.0007465sand + 0.03532\sqrt{clay} (R^2=0.33) \quad (4.11)$$

$$e^{\theta_r} = 1.009 + 0.04099\sqrt{clay} - 4.9035E - 0.21silt^{10} + 0.004877 \ln(orgC) (R^2=0.32) \quad (4.12)$$

$$e^{\theta_r} = 1.007 + 0.03896\sqrt{clay} - 0.03928BD^2 (R^2=0.35) \quad (4.13)$$

The set of possible multiple linear regression equations developed involving $\sqrt{\alpha}$ are listed as follows:

$$\sqrt{\alpha} = 0.18 - 1.0440E - 043e^{\text{sand} - \frac{0.071381}{\text{silt}^2}} - 0.02068\text{BD} - \frac{0.088661}{\text{clay}} + 0.0008447\ln(\text{orgC})(R^2=0.05) \quad (4.14)$$

$$\sqrt{\alpha} = 0.1847 - 2.0502E - 043\exp(\text{sand}) - 0.02556\text{BD} - 0.09316\frac{1}{\text{clay}}(R^2=0.05) \quad (4.15)$$

$$\sqrt{\alpha} = 0.1564 - 2.0817E - 043e^{\text{sand}} - 0.08877\frac{1}{\text{clay}} + 0.004372\ln(\text{orgC})(R^2=0.04) \quad (4.16)$$

The set of possible multiple linear regression equations developed involving \sqrt{n} are listed as follows:

$$\sqrt{n} = 0.9871 + 4.0051E - 011\text{sand}^5 + (\text{silt}^{-1}) - 0.03977\ln(\text{Clay}) + 0.0000639\text{orgC}^{-2}(R^2=0.24) \quad (4.17)$$

$$\sqrt{n} = 1.101 + 0.8702\frac{1}{\text{silt}} - 0.07228\ln(\text{clay})(R^2 = 0.23) \quad (4.18)$$

$$\sqrt{n} = 0.8617 + 8.4062E - 011\text{sand}^5 + 2.2259E - 005\left(\frac{1}{\text{orgC}^2}\right)(R^2=0.21) \quad (4.19)$$

$$\sqrt{n} = 1.22 - 0.09247\ln(\text{clay}) + \frac{0.00011681}{\text{orgC}^2}(R^2=0.12) \quad (4.20)$$

4.7.6 Validation of the developed pedotransfer functions to evaluate reliability

Table 4.18 shows regression equations developed for each of the selected transformations of the moisture retention parameters of the Van Genuchten moisture retention equation and measures of their performance in validation based on the statistical measures indicated. The transformed model parameter that yielded the best reliability is the saturated soil moisture content $\left(\frac{1}{\theta_s}\right)$ considering the measure of correlation coefficient of 0.86 ($r^2=0.74$). This indicates that θ_s is well predicted by the pedotransfer function developed. Prediction of transformed variable $\sqrt{\alpha}$ yielded the lowest possible value for the selected measure for the quality of fit ($r=0.25$) reflecting the lowest level of reliability compared to the other pedotransfer functions developed for moisture retention parameters under consideration. The transformed variable for residual moisture content e^{θ_r} yielded modest performance with the best possible value of correlation obtained, $r=0.64$ ($r^2=0.41$).

Table 4.18. Summary of equations for transformations of Van Genuchten parameters and their performance in reliability

<u>Response Variable</u>	<u>Equation</u>	<u>r</u>	<u>r²</u>	<u>RMSE</u>	<u>RSR</u>	<u>NSE</u>	<u>ME</u>	<u>PBIAS</u>
$\frac{1}{\theta_s}$	$\frac{1}{\theta_s} = 0.8103 + 0.003534\text{sand} + 0.3099e^{(BD)}$	0.87	0.76	2.09	0.50	0.75	0.001	0.07
$\frac{1}{\theta_s}$	$\frac{1}{\theta_s} = 1.397 + 0.003176\text{sand} + 0.2451BD^3$	0.87	0.76	2.12	0.50	0.75	0.004	0.19
$\frac{1}{\theta_s}$	$\frac{1}{\theta_s} = 1.477 + 0.003045\text{sand} - 0.1505\frac{1}{\text{silt}} + 0.2126BD^3 - 9.3857E-006\text{clay}^2 - 0.04542\ln(\text{OrgC})$	0.86	0.74	2.11	0.51	0.74	0.009	0.44
$\frac{1}{\theta_s}$	$\frac{1}{\theta_s} = 1.397 + 0.003385\text{sand} - 0.1433\frac{1}{\text{silt}} + 0.246BD^3$	0.87	0.76	2.13	0.51	0.74	0.006	0.31
e^{θ_r}	$e^{\theta_r} = 1.083 + 0.0382\sqrt{\text{clay}} - 5.0290E-021\text{silt}^{10} - 0.0392BD^2$	0.64	0.41	1.10	0.77	0.41	0.011	0.90
e^{θ_r}	$e^{\theta_r} = 1.062 - 0.0007465\text{sand} + 0.03532\sqrt{\text{clay}}$	0.60	0.36	1.14	0.80	0.36	0.008	0.63
e^{θ_r}	$e^{\theta_r} = 1.009 + 0.04099\sqrt{\text{clay}} - 4.9035E-021\text{silt}^{10} + 0.004877\ln(\text{OrgC})$	0.63	0.40	1.11	0.78	0.39	0.007	0.55
e^{θ_r}	$e^{\theta_r} = 1.077 + 0.03896\sqrt{\text{clay}} - 0.03928BD^2$	0.64	0.41	1.10	0.77	0.40	0.011	0.90
e^{θ_r}	$e^{\theta_r} = 1.003 + 0.0417\sqrt{\text{clay}} + 0.005286\ln(\text{OrgC})$	0.62	0.38	1.12	0.79	0.38	0.007	0.57
$\sqrt{\alpha}$	$\sqrt{\alpha} = 0.18 - 1.0440E-043\exp(\text{sand}) - 0.07138\frac{1}{\text{silt}^2} - 0.02068BD - 0.088661/\text{clay} + 0.0008447\ln(\text{ORGC})$	0.19	0.04	0.85	1.00	0.01	-0.012	-8.82
$\sqrt{\alpha}$	$\sqrt{\alpha} = 0.1847 - 2.0502E-043\exp(\text{sand}) - 0.02556BD - 0.093161/\text{clay}$	0.19	0.04	0.85	0.99	0.01	-0.011	-8.29
$\sqrt{\alpha}$	$\sqrt{\alpha} = 0.1564 - 2.0817E-043\exp(\text{sand}) - 0.088771/\text{clay} + 0.004372\ln(\text{OrgC}).$	0.25	0.06	0.85	0.99	0.02	-0.013	-9.39
\sqrt{n}	$\sqrt{n} = 0.9871 + 4.0051E-011\text{sand}^5 + 0.6231/\text{silt} - 0.03977\ln(\text{Clay}) - 0.00006391/\text{OrgC}^2$	0.45	0.20	1.89	0.90	0.20	-0.006	-0.62
\sqrt{n}	$\sqrt{n} = 1.101 + 0.87021/\text{silt} - 0.07228\ln(\text{Clay})$	0.42	0.18	1.91	0.91	0.17	-0.002	-0.17
\sqrt{n}	$\sqrt{n} = 0.8617 + 8.4062E-011\text{sand}^5 + 2.2259E-0051/\text{OrgC}^2$	0.45	0.20	1.88	0.89	0.20	-0.005	-0.59
\sqrt{n}	$\sqrt{n} = 1.22 - 0.09247\ln(\text{Clay}) + 0.00011681/\text{OrgC}^2$	0.29	0.08	2.02	0.96	0.08	-0.002	0.19

Wösten et al. (2001) notes that continuous pedotransfer functions can be applied in the case of more site specific applications, where measured data is available, since they do not provide site specific information. The pedotransfer functions predict soil hydraulic characteristics for broadly defined textural classes. **Figures 4.18 through 4.21** illustrates the performance of the developed pedotrasfer functions for the transformed response variables based on analysis by linear regression.

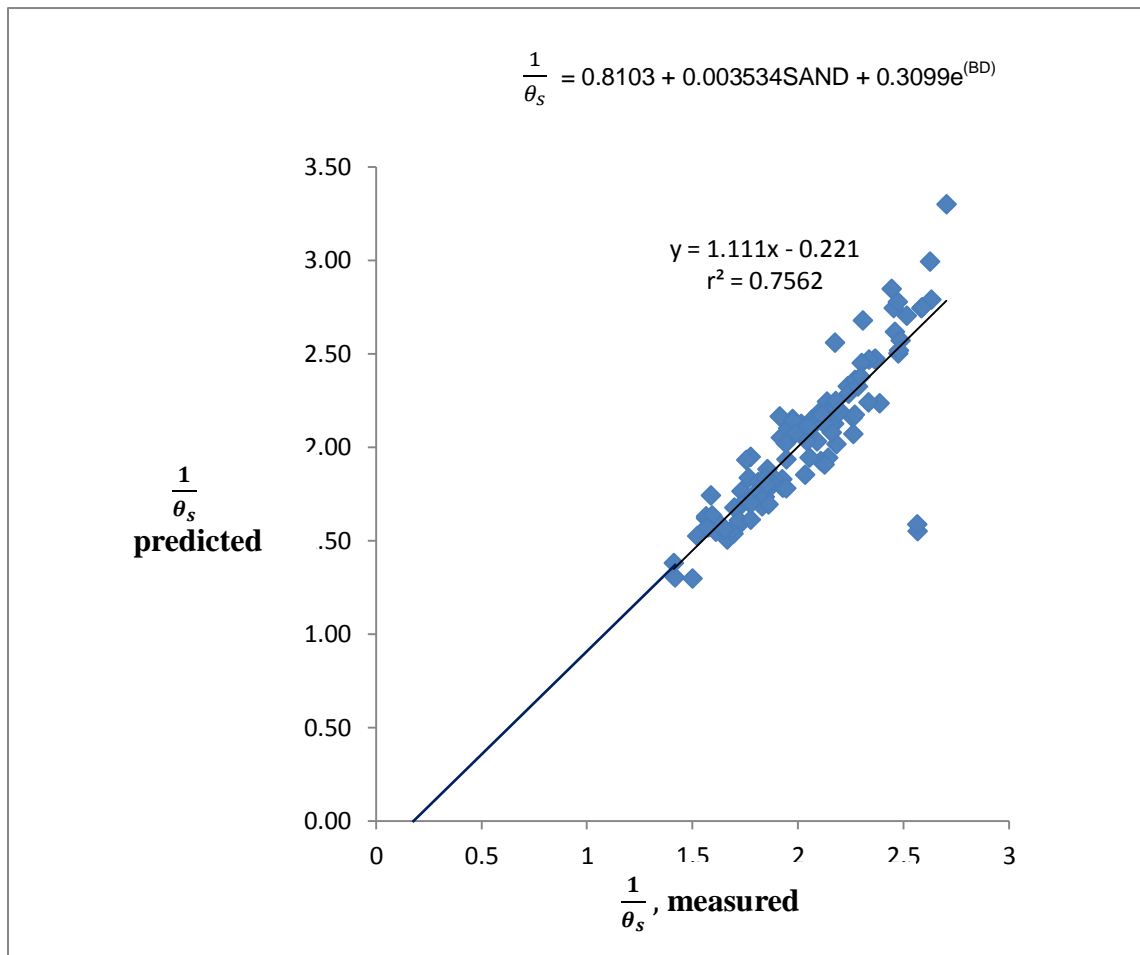


Figure 4.18 : Predicted value of transformed variable $\frac{1}{\theta_s}$ versus measured by correlation for a developed equation that yielded the best possible quality of fit.

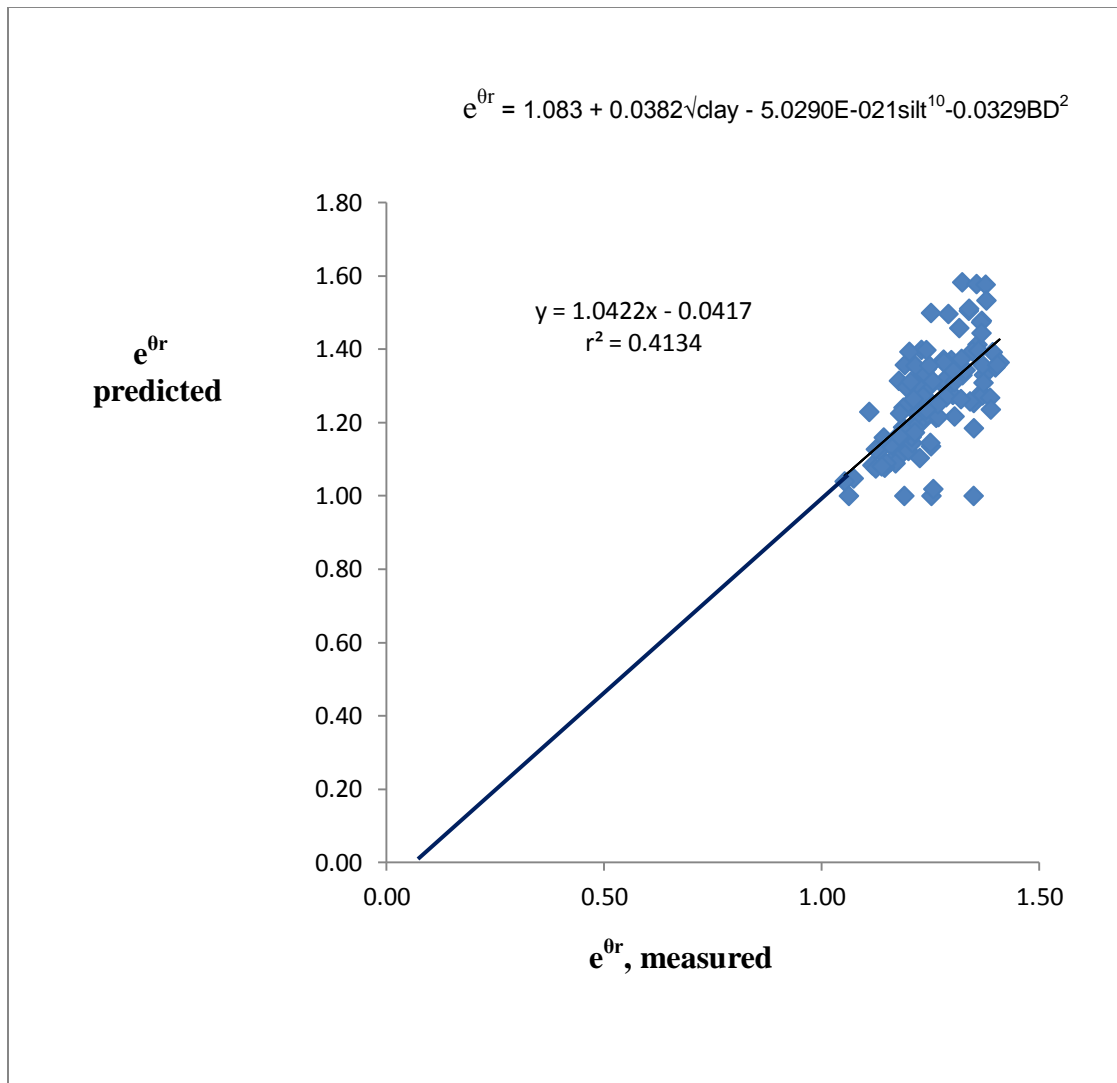


Figure 4.19 : Predicted value of transformed variable e^{θ_r} versus measured by correlation for a developed equation that yielded the best possible quality of fit.

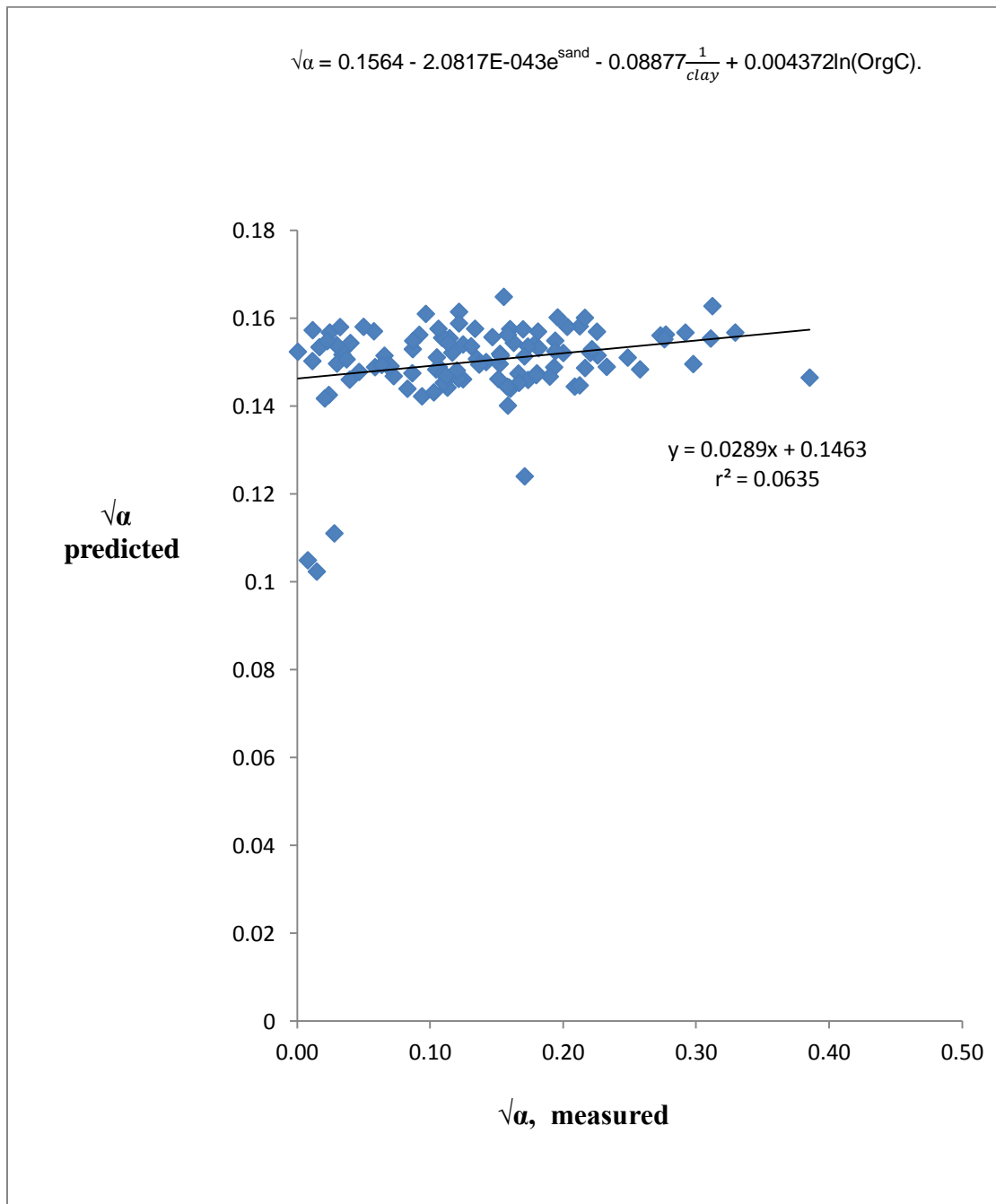


Figure 4.20 : Predicted value of transformed variable $\sqrt{\alpha}$ versus measured by correlation for a developed equation that yielded the best possible quality of fit.

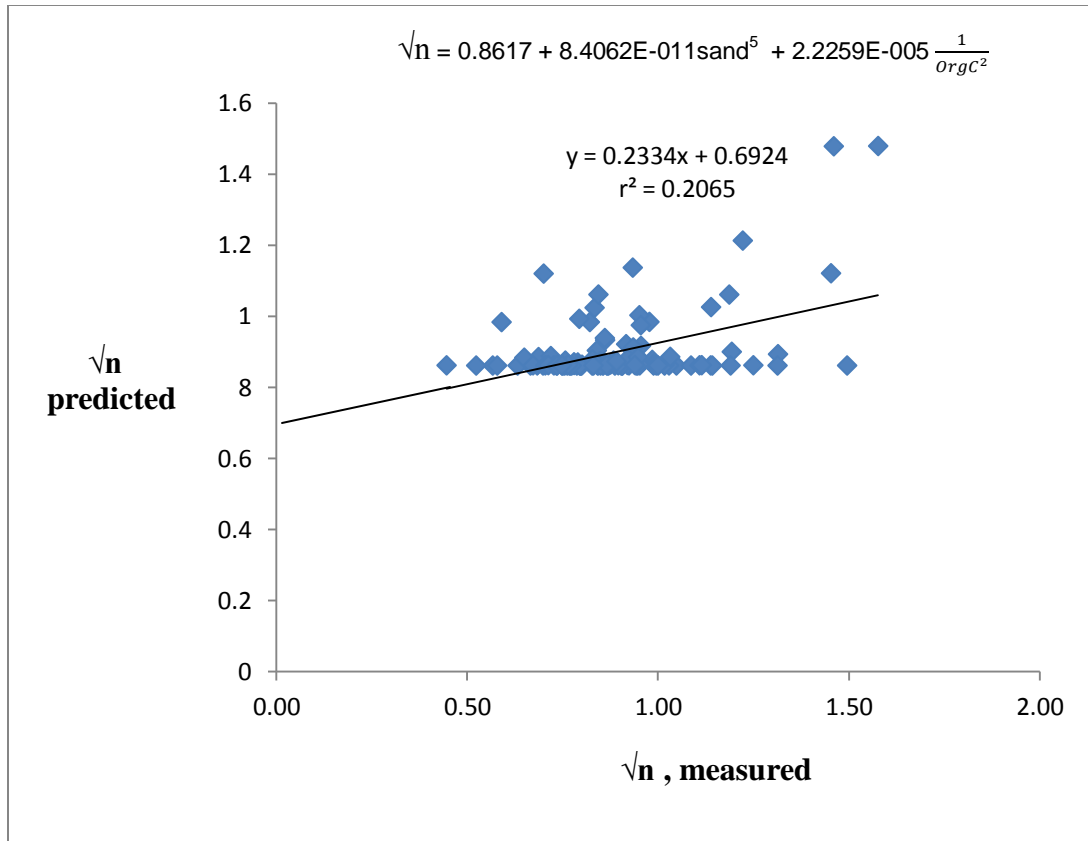


Figure 4.21: Predicted value of transformed variable \sqrt{n} versus measured by correlation for a developed equation that yielded the best possible quality of fit.

Based on the above indicated equations that produced the best possible fit, equations relating the moisture retention parameters that are not transformed to the transformed basic properties were also compared to the measured (fitted) parameters and yielded the values of R^2 indicated below alongside the equations.

$$\theta_s = \frac{1}{0.8103 + 0.003534SAND + 0.3099e^{BD}} \quad (R^2 = 0.77) \quad (4.21)$$

$$\theta_r = \ln(1.083 + 0.0382\sqrt{\text{clay}} - 5.0290E-021\text{silt}^{10} - 0.0329BD^2) \quad (R^2 = 0.43) \quad (4.22)$$

$$n = (0.8617 + 8.4062E-011\text{sand}^5 + 2.2259E-005 \frac{1}{\text{ORG}^2})^2 \quad (R^2 = 0.28) \quad (4.23)$$

$$\alpha = (0.1564 - 2.0817E-043e^{\text{sand}} - 0.08877 \frac{1}{\text{clay}} + 0.004372\ln(\text{OrgC}))^2 \quad (R^2 = 0.04) \quad (4.24)$$

4.8. Surface runoff simulation using saturated hydraulic conductivity values predicted using developed pedotransfer functions.

The simulation was performed during the period 1/1/1998 to 31/12/2012 being also the period for model validation from previous analyses of performance of pedotransfer functions earlier assessed. **Figure 4.22** shows the hydrograph of surface runoff predicted using K_s values obtained from developed pedotransfer functions and the observed surface runoff for the said period. **Figure 4.23** shows the comparison of predicted and observed surface flow based on regression plot.

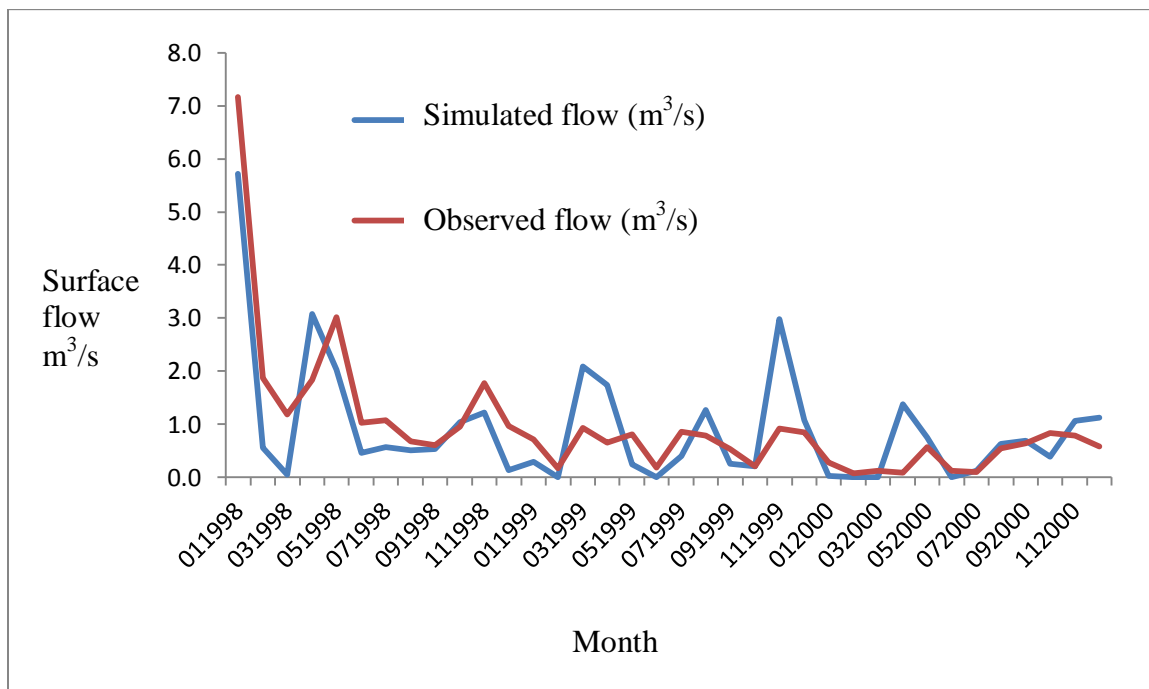


Figure 4.22: Hydrographs of simulated and observed surface runoff based on K_s values determined using developed pedotransfer functions

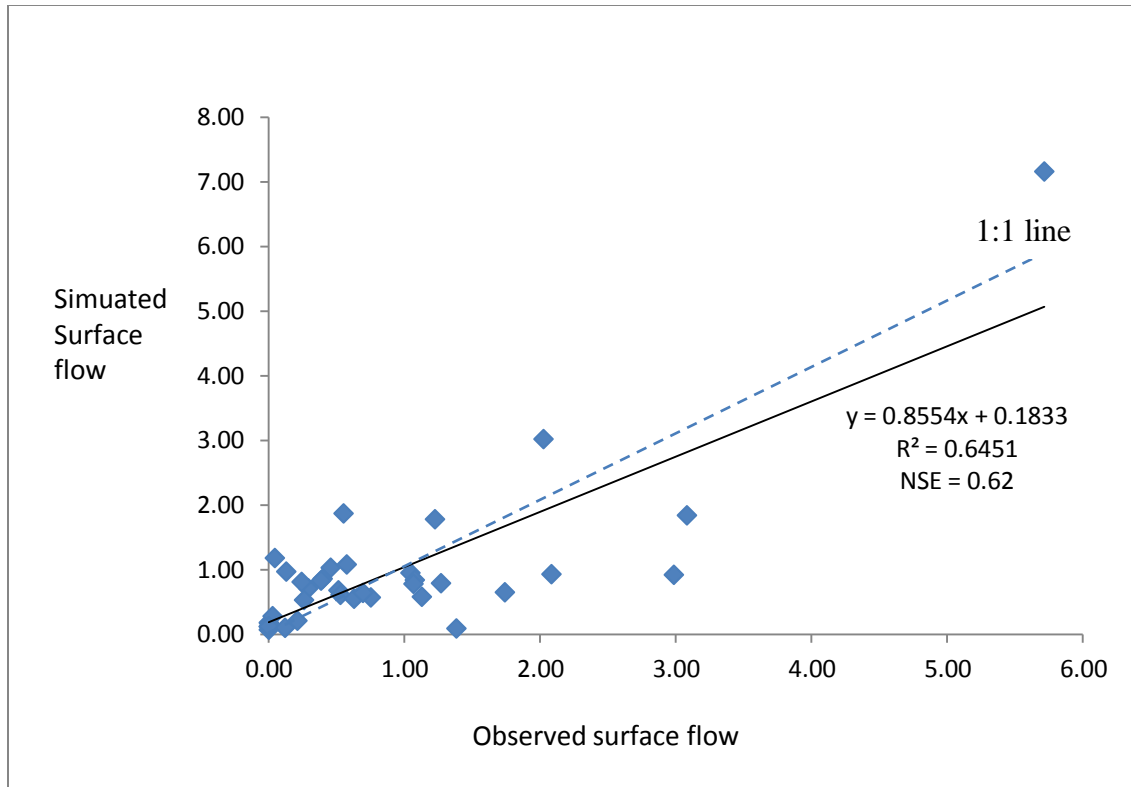


Figure 4.23: Comparison of simulated and observed surface runoff using linear regression based on the value of K_s calculated using developed pedotransfer functions.

From **Figure 4.22**, it is evident that the predicted flow tends to follow the pattern of the observed flow. There are few instances where the simulated flow slightly under predicts the observed flow during the period January 1998 to March 1999. During the period April 1999 to December 2000, the model either overpredicts the flow or follows well the flow pattern. The coefficient of determination (r^2) yielded a value of about 0.65 (>0.5) reflecting acceptable performance of the model. The Nash-Sutcliff efficiency value of 0.62 reflected satisfactory model performance of the model in predicting the low flows. The performance of the model in surface flow simulation using the K_s values based on developed pedotransfer functions was compared to the performance of other pedotransfer functions earlier discussed (section 4.6) using known performance measures. **Table 4.19** shows the comparison based on the performance measures. The NSE value is comparable to the other pedotransfer function. The value is higher than obtained for Jabro, Puckett and Brakensiek and lower than for Saxton and Campbell, but equal to that of Young. The K_s value from developed PTFs yielded the highest value for the coefficient of

determination ($r^2=0.65$), however this value was not so much higher than for the other PTFs studied. **Figure 4.24** illustrates the comparison between the developed pedotransfer functions and some selected existing PTFs indicating their performances in surface runoff simulation based on the coefficient of determination (r^2). The difference between means of the observed and predicted mean flow based on the K_s value from developed PTFs was 5% indicating nearly close agreement between observed and simulated average flows and hence good performance. The percentage bias (PBIAS) value of 6.88 for the developed model was better than for all the other models except for Dane and indicated very good performance. The RSR value was also among the lowest possible value compared to the other PTFs and showed satisfactory performance. In general the performance of the developed PTFs in predicting surface runoff is considered acceptable. The PTFs performed better than the others considered while it was comparable to the others. The PTFs generally performed better than most of the others considered in surface flow prediction. For each of the performance measures, the developed PTFs ranked best or second best in surface flow prediction when used in predicting K_s .

Table 4.19: Comparison of performance of developed pedotransfer function for K_s (Obiero) and existing PTFs in surface flow simulation.

	SAX86	PUCKET T	Jabro	DANE	Brakensiek	Campbell	Young	Obiero
NSE	0.63	0.59	0.63	0.62	0.61	0.63	0.62	0.62
r	0.80	0.80	0.80	0.80	0.79	0.80	0.79	0.81
r^2	0.64	0.64	0.64	0.64	0.62	0.64	0.62	0.65
O_{av}	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96
P_{ave}	0.90	1.13	0.90	0.97	0.84	0.86	0.87	0.91
Dv	6.20	-17.81	5.80	-1.13	12.12	9.99	9.63	5.46
PBIAS	7.82	-22.48	7.32	-1.43	15.29	12.61	12.16	6.88
RMSE	4.42	4.65	4.42	4.46	4.49	4.42	4.46	4.43
RSR	0.61	0.64	0.61	0.62	0.62	0.61	0.62	0.61

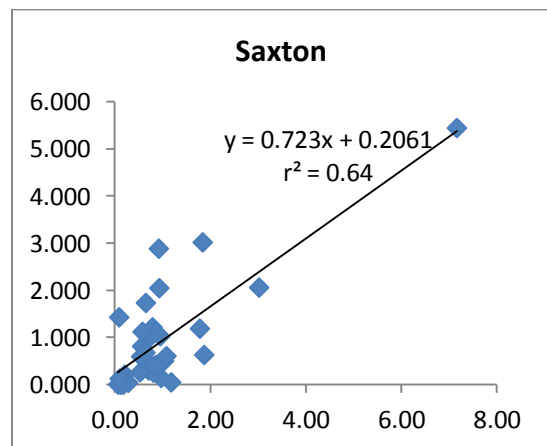
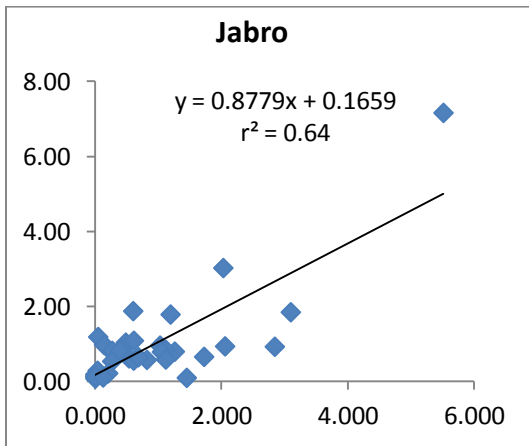
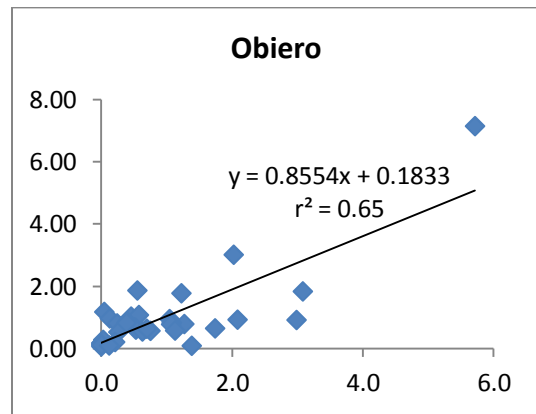
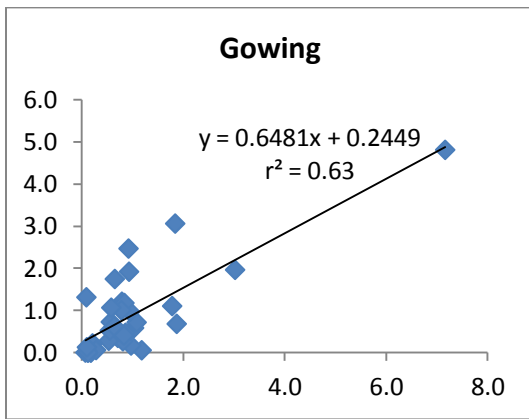
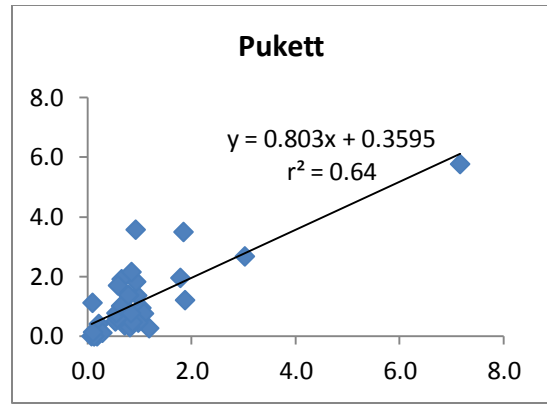
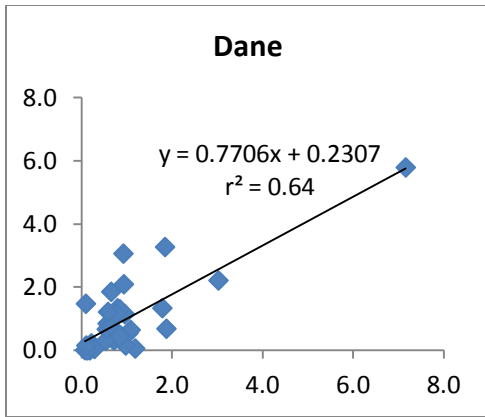


Figure 4.24: Comparison of simulated and observed surface flows for the developed and some existing PTFs.

5 CONCLUSIONS AND RECOMMENDATIONS

5.1 Conclusions

Within the range of this study, the following conclusions can be drawn:

1. Parameters used in Pedotransfer Functions (PTFs) for predicting saturated hydraulic conductivity (K_s) are based on basic soil properties that can be measured and which are can easily be obtained in most soil data bases.
2. Preliminary model assessment, calibration and validation efforts showed that Soil Water Assessment Tool (SWAT) model is a good simulator of surface runoff and could be used in verification of developed Pedotransfer Functions in surface runoff prediction.
3. Performance of the developed Pedotransfer Functions for estimating Van Genuchten moisture retention parameters in saturated hydraulic conductivity prediction varied for the individual moisture retention paramaters, but in general showed acceptable performance.
4. Developed Pedotransfer Functions for saturated hydraulic conductivity showed satisfactory performance in flow prediction during verification yielding a coefficient of determination value higher than those obtained with existing Pedotransfer Functions.

5.2 Recommendations for further study

The following are recommendations for further study:

1. In developing PTFs, it is preferable to use soil properties that are available in soil data bases. Emphasis should be put in developing PTFs based on the use of moisture retention parameters in predicting K_s where little effort have been made;

2. Efforts should be made in developing data bases that would provide more elaborate and good quality data for use in hydrological modeling; and
3. Continued effort should be made in developing more pedotransfer functions to provide users with more alternative equations and especially those PTFs that use the very basic soil properties available in many local soils data bases.

6. REFERENCES

Andualem, G and M. Yonas. 2008. Prediction of sediment flow to Legegadi Reservoir Using SWAT Watershed and CCHELD Sediment Transport Models. Nile Basin Water Engineering Scientific magazine 1: 66-74.

Arnold, J. G , P. M. Allen, R. Muttiah and G. Bernhardt. 1995. Automated baseflow separation and recession analysis techniques. Ground Water Journal 33(6): 1010-1018.

Bartes, N. H. and P. Gicheru. 2004. Soil data derived from SOTER for studies of carbon stocks and change in Kenya (version 1.0), project report 2004/01. International Soil Reference and Information Centre (ISRIC)-World Soil Information.

Bekele E.G., and H. N Knapp. 2008. Hydrologic. Modeling of Fox Watershed: Model Development, Calibration, and Validation. Proceedings of the World Environmental and Water Resources Congress. Honolulu, Hawaii, USA, 12-16 May 2008. ASCE/EWRI: 1-10.

Chinnarasi, C., T. Tinsanchai, and S. Banchuen, 2008. Field validation of two river morphological models on the Pasak River, Thailand. Hydrological Sciences Journal 53(4): 818-833.

Dawadi, D. and S. Ahmad. 2012. Changing climatic conditions in the Colorado River Basin: Implications for water resources management. Journal of Hydrology 430-431: 127-141.

Fooladmad, H. R. 2011. Pedotransfer functions for point estimation of soil moisture characteristic curve in Some Iranian soils. African Journal of Agricultural Research 6(6), March 2011: 1586 -1591.

Gassman, P. W, M. R. Reyes, C. H. Green and J. G. Arnold. 2007. The Soil water

Assessment Tool: Historical Development, Applications and Future Research Directions. Transactions of the ASABE 5(4): 1211-1250.

Gathenya, J. M, H. P. Liniger, and F. N. Gichuki. 1993. Problems of River-Water Management for a Basin West of Mount Kenya: Challenge to Water Resource Planners, Proceedings of the Fourth National Workshop on Land and Water Management in Kenya: Towards Sustainable Land Use, Kikuyu, Kenya, 15-19 February: 175-180.

Githui, F; Gitau, W; Mutua, F. and W. Bauwens. 2008. Climate Change Impact on SWAT simulated streamflow in Western Kenya. International Journal of Climatology 29 (12): 1823-1834

Githui, F, F. Mutua and W. Bauwens. 2009. Estimating Impacts of Land Cover Change on Runoff using the Soil Water Assessment Tool (SWAT): Case study of Nzoia Catchment, Kenya. Hydrological Sciences Journal 54(5) 5: 899-908.

Govender, M. and C.S. Everson. 2005. Modeling Streamflow from two small African experimental Catchments using SWAT model. Hydrological Processes Journal 19(3): 683-692.

Gowing, J.W and M.D.B. Young. 1996. Evaluation and Promotion of Rainwater Harvesting in Semi-Arid areas. The PARCHED-THIRST model, User Guide Version 1.0. University of Newcastle: 69-70.

Heuvelmans, G, J.F.G. Qujano, B. Muys,; J. Feyen, P. Coppin. 2005. Hydrological Processes Journal 19 (3): 729-748.

Jayakrishnan, R, R. Srinivasan, C. Santhi and J. G. Arnold. 2005. Advances in the application of the SWAT model for water resources management. Hydrological Processes Journal 19(3): 749-762.

Jain, S. K. and K. P. Sudheer. 2008. Fitting of Hydrologic Models: A close look at the Nash-Sutcliffe Index. *Journal of Hydrologic Engineering* 13(10): 981-986

Lamont , S.J., R. N. Eli, and J. J. Fletcher. 2008. *Journal of Hydrologic Engineering* 13 (7): 621-635.

Levesque, E, F. Anctil and A. Van Griensven. 2008. Evaluation of the Streamflow Simulation by SWAT model for two Small Watersheds under Snowmelt and Rainfall. *Hydrological Sciences Journal* 53(5): 961–976.

Lorentz, S., P. Goba, and J. Pretorius. 2001. Hydrological processes research: Experiments and measurements of soil hydraulic characteristics. Project Report to the Water Research Commission on the project “Experimentation and Laboratory Measurement for Hydrological Process Research” (Report No.:K5/744), School of Bioresources and Environmental Hydrology, University of Natal, Pietermaritzburg, South Africa.

Maidment, D.R. 1993. *Handbook of Hydrology*. Mc Grawhill, Inc. USA:17.1-17.55.

Marshal, T. J. 1958. A relation between permeability and size distribution of pores. *Journal of Soil Science* 9(1): 1-8.

Marshal, T. J. and J. W. Holmes. 1988. *Soil Physics*. Cambridge University Press, United Kingdom.

Moriasi, D. N., J. G. Arnold, M. W. Van Liew, R. L. Binger, R. D Harmel, and T. L. Veith. 2007. Model Evaluation Guidelines for Systematic Quantification of Accuracy in Watershed Simulations. *Transactions of the American Society of Agricultural and Biological Engineers (ASABE)* 50(3): 885-900.

Nandagari, L. and S. Prasad. 1996. Field Evaluation of Unsaturated Hydraulic Conductivity Models and Parameter Estimation from Retention Data. *Journal of Hydrology* 179 (1-4): 197-205.

Nandagari, L. and R. Prasad .1997. Relative Performances of Textural Models in Estimating Soil Moisture Characteristic. *Journal of Irrigation and Drainage Engineering*. 23 (3): 211-214.

Ndomba, P. M. 2010. Modeling of Sediment Upstream of Nyumba Ya Mungu Reservoir in Pangani River Basin. *Nile Basin Water Science & Engineering Journal* 3(2): 25-38.

Ndomba, P. M. and B. Z. Birhanu. 2008. Problems and Prospects of SWAT Model Applications in NILOTIC Catchments: A Review. *Nile Water Engineering Scientific Magazine* 1: 41-52.

Neitsch, S. L., J. G. Arnold, J. R. Kiniry, and J. R. Williams. 2005. Soil Water Assessment Tool Theoretical Documentation (version 2005). Grassland Soil and Water Research Laboratory, Temple, Texas: 1-27.

(a) Neitsch , S. L. J. G. Arnold, J. R. Kiniry, J. R. Williams and K. W. King. 2002. Soil Water Assessment Tool Theoretical Documentation (version 2000). Grassland Soil and Water Research Laboratory, Temple, Texas: 215-230.

(b) Neitsch , S. L. J. G. Arnold, J. R. Kiniry, J. R. Williams and K. W. King. 2002. Soil Water Assessment Tool Theoretical Documentation (version 2000). Grassland Soil and Water Research Laboratory, Temple, Texas: 93-115.

(c) Neitsch, S. L. J. G. Arnold, K. R. Kiniry, R. Srinivasan and J. R. Williams. 2002. Soil Water and Assessment Tool User's Manual, Grassland Soil and water research

Laboratory, Temple, Texas 76502: 341-378.

Ngigi, S. N. 2006. Hydrological Impacts of Land Use Changes on Water Resources Management and Socio-economic Development of Upper Ewaso Ng'iro River Basin in Kenya. PhD Thesis. Delft University of Technology, Netherlands.

Obiero, J. P. O. 1996. Evaluation of Infiltration Using Green Ampt Model and Rainfall Runoff Data for Lagan and Sambret catchments, Kericho, Kenya. M.Sc. Thesis. Dept. of Agricultural Engineering, University of Nairobi.

Obiero, J. P. O., C. T. Omuto and D. O. Muge. 2003. Rainwater Harvesting for Crop Production in Semi-Arid areas: A hydrology based teaching manual. Department of Agricultural Engineering, University of Nairobi.

Obiero, J. P. O, M. A. Hassan, and L. O. M. Gumbe. 2011. Modelling of Streamflow of a catchment in Kenya. *Journal of Water Resource and Protection* 3 (9): 667-677

Obuobie, E. and D. Bernd. 2008. Using SWAT to Evaluate Climate Change Impact on Water Resources in the White Volta River Basin, West Africa. Conference on International Research on Food Security, Natural Resource Management and Rural Development. University of Hohenheim, October 7-9.

Odira, P. M. A, M. O. Nyadawa, B. Okello, N. A. Juma and J. P. O. Obiero. 2010. Impact of Land Use/Cover dynamics on Stream flow: A case Study of Nzoia river Catchment, Kenya. *Nile Water Science and Engineering Journal* 3(2): 64-78.

Pathak, P.; K. B. Laryea; and R. Sudi. 1989. A runoff model for Small Watershed in Semi-Arid Tropics. *ASAE*. 32(5): 1619-1624.

Rasoulzadeh, A. 2011. Estimating Hydraulic Conductivity Using Pedotransfer Functions,

Hydraulic Conductivity - Issues, Determination and Applications. Lakshmanan Elango (Ed.), ISBN: 978-953-307-288-3, InTech, DOI: 10.5772/22753: 145-164: <http://www.intechopen.com/books/hydraulic-conductivity-issues-determination-and-applications/estimating-hydraulic-conductivity-using-pedotransfer-functions>.

Rostamian, R., A. Jaleh, M. Afyuni, S. F. Mousavi, M. Heidarpour, A. Jalalian and K. C. Abbaspour. 2008. Application of a SWAT Model for Estimating Runoff and Sediment in two Mountainous basins in Central Iran. *Hydrological Sciences Journal* 53 (5): 977-988.

Sarangi, A., D. K. Singh, and A. K. 2008. *Current science* 94 (12): 1620-1626.

Saxton, K. E and W. J. Rawls. 2006. Soil Water Characteristic Estimates by Texture and Organic Matter for Hydrologic Solutions. *Soil Science Soc. Am. J.* 70: 1569-1578.

Schmalz, B; F. Tavares, and N. Fohrer. 2008. Modeling hydrological processes in mesoscale lowland river basins with SWAT-capabilities and challenges. *Hydrological Sciences Journal* 53(5): 989-1000.

Setegen, S. G., R. Srinivasan and B. Dargahi. 2008. Hydrological Modelling in Lake Tana Basin, Ethiopia Using SWAT Model. *The Open Hydrology Journal* 2: 49-62.

Sharda, V.N., R. M. Patel, S. O Prasher, P. R. Ojasui, and C. Prakash. 2006. *Agricultural Water Management*, 83. Elsevier Science. Amsterdam: 233-242.

Singh, P., U.K. Haritashya, and N. Kumar. 2008. Modeling and estimation of different components of stream flow for Gangotri Glacier basin, Himalayas. *Hydrological Sciences Journal* 53(2): 309-322.

Siriba, D. 2004. Documentation on procedure used in digitization of key features in the Naro Moru Catchment. Personal communication. Department of Geospatial and space Technology, University of Nairobi.

Sobieraj, J. A., H. Elsenbeer and R. A. Vertessy. 2001. Pedotransfer functions for estimating saturated hydraulic conductivity: implications for modeling storm flow generation. *Journal of Hydrology* 251: 202-220.

Schaap, M. G. and Leij, F.J. 1998. Using neural networks to predict soil water retention and soil hydraulic conductivity. *Soil and Tillage Research* 47: 37-42.

Stehr, A., A P. Debeles, F. Romero and H. Alcayaga. 2008. Hydrological modeling with SWAT under conditions of limited data availability: evaluation of results from Chilean case study. *Hydrological Sciences Journal* 53(5): 588-601.

Tadele, K. and G. Förch. 2007. Impact of Land Use/Cover Change on Stream flow: The case of Hare River watershed, Ethiopia. *Catchment and Lake Research. Proceedings 2nd Lake Abaya Research Symposium (LARS). Arba Minch, Ethiopia 7th -11th May, 2007.*

Touma, J. And J. Albergel 1992. Determining Soil Hydrological Properties from Rain Simulator and Double Ring Infiltrometer experiments: A comparison. *Journal of Hydrology* 135: 73-86.

Technical Brief 2. 2007. Hydrological Modelling of the Upper Malaprabha Catchment using ArcView SWAT. Development of Tools and Methodologies to Implement the Payment for Environmental services Concept in Watersheds in India (PES India Project). Norwegian Institute for Water Research, supported by Royal Norwegian Embassy, India.

Thomas, M.K., H. P. Liniger, and F. N. Gichuki. 1993. Development of a Stream flow Model for Rural Catchments in Kenya. *Proceedings of the Fourth National Workshop on Land and Water Management in Kenya: Towards Sustainable Land Use. Kikuyu, Kenya,*

15-19 February, 1993: 191-202.

Tripathi, M.P., R. K. Panda and N. S. Raghuwanshi. 2005. Development of Effective Management Plan for Critical Sub Watersheds using SWAT Model. *Hydrological Processes Journal* 19(3): 809-826.

Tomasella, J. and M. Hodnett. 2004. *Developments in Soil science* 30. ISSN 0166-2481/DOI 10.1016/S0166-2481(04)30021-8, Elsevier, B. V: 415-429.

Van Mullem, J. A. 1991. Runoff and Peak Discharges using Green-Ampt Infiltration Model. *Journal of Hydraulic Engineering* 117 (3). ASCE: 354-370.

Vereecken, H. and M. Herbst. 2004. *Developments in Soil Science* 30, ISSN 0166-2481/DOI 10.1016/S0166-2481(04)30021-8. Elsevier, B. V: 3-19.

Walczack, R. T., F. Moreno, C. Slawinski, E. Fernandez and J. L. Arrue. 2006. Modelling of soil water retention curve using soil solid phase parameters. *Journal of Hydrology* 329: 527-533.

Wells, T.; Fityus, S; Smith, D. W. & H. Moe .2006. The direct estimation of saturated hydraulic conductivity of soils using measurements of gas permeability, Laboratory testing with dry granular soils. *Australian Journal of Soil research*, Nov. 2006.

Wösten, J. H. M, Y. A. Pachepsky and W. J. Rawls. 2001. Pedotransfer functions: bridging the gap between available basic soil data and missing soil hydraulic characteristics. *Journal of Hydrology* 251: 123-150.

Young, M.D.B, and J.W. Gowing. 1996. Evaluation and Promotion of Rainwater Harvesting in Semi-arid Areas: The Patched-Thirst model, User Guide Version 1.0.

Department of Agricultural and Environmental Science, University of Newcastle Upon Tyne, U.K.: 61-64.

Zayani, K. G. Vachaud and N. Ennabli. 1992. Estimation of Unsaturated Hydraulic Conductivity from Inflow Data. *Journal of Hydrology* 138: 503-514.

APPENDIX 1: PHOTOGRAPHIC MAPPING PROCEDURE FOR NARO MORU CATCHMENT

The procedure used in the photogrammetric mapping of the Naro Moru catchment based on the use of aerial photographs and the relevant topographical map proceeded as follows:

- suitable control points were identified and marked in the topo sheets. These included river junctions, road junctions, railway/road junctions, hilltops etc. These points were likewise located in the aerial photographs. As many of these points as possible are identified and marked.
- The coordinates as well as the spot heights of the above mentioned control points were established. **Table A1** shows recorded coordinates, heights and descriptions of the selected points and serial numbers as marked on the topo sheets and aerial photographs.
- The identified control points were then plotted onto a gridded transparent plotting sheet prepared to plot the contours to a reduced scale of 1: 20,000.
- The next step involved setting up of the plotting sheet for mapping of the described features (e.g. contours). This involved mounting the pairs of overlapping aerial photographs (stereo pairs) on plotting equipment known as the stereo top appropriately and systematically removing parallax through an iterative process to ensure the plotted coordinates and spot heights for the selected control points are correct.
- The parallax removal is accomplished through a standardized computational procedure with a special sheet for tabulation. Once the photo pairs have been mounted on the stereotype machine, adjustments are made using appropriate adjustment screws until a floating mark is observed.

Table A1.1 selected control points for mapping using aerial photos and toposheets

Control point	Easting	Northing	Height(m)	Description
1	267,300	9994342	1795.00	River junction
2	268360	9994620	1800.00	Road junction
3	270140	9995660	1821.00	Road junction
4	267960	9989060	1818.00	Road/river junction
5	267340	9986460	1850.00	Road/cutline junction
6	266880	9984780	1882.00	Road/footpath junction
7	269200	9985140	1830.00	Road/footpath junction
8	270560	9985620	1885.00	Cutlines junction
9	272460	9984740	1885.00	Roads junction
10	272680	9989740	1845.00	Cutlines junction
11	272060	9995370	1823.00	Roads junction
12	272560	9993360	1846.00	Road junction
12A	276240	9995440	1838.00	Road and river junction
13	275900	9985640	1838.00	River and rail junction
14	275800	9985560	1895.00	River and rail junction
15	274780	9984400	1911.00	Cutline and road junction
16	275800	9985560	1895.00	River and road junction
17	277430	9984850	1915.00	River junction
18	279040	9984565	1945.00	River junction
19	281120	9986140	1981.00	Roads junction
20	279040	9982050	1975.00	Roads junction
21	277640	9979980	1970.00	River/road junction
23A	278040	9977000	1952.00	Footpath junctions
24	-	-	1981.00	River/footpath junction
25	279940	9979210	1978.00	River junction
26	283180	9983620	2040.50	River/track junction
27	-	-	1984.30	
28	283100	9979500	2068.00	Road junction

29	284700	9978680	2075.00	Road junction
30	-	-	2070.00	Road junction
31	-	-	2073.30	Road junction
32	285240	9975480	2101.60	Road/track junction
33	287400	9976440	2118.40	River/track junction
34	286960	9979960	2103.50	River and footpath junction
35	285280	9981980	2097.90	River/track junction
36	287800	9983180	2163.50	River junction
37	290180	9985440	2254.90	River/road junction
38	289760	9980460	2193.00	River/road junction
38A	288760	9980100	2190.00	Roads junction
39A	292400	9976200	2194.00	Roads junction
40	292120	9978480	2318.00	Roads junction
41	291620	9980940	2318	Roads junction
42	-	-	2423	Open ground
43A	295100	9977330	7895.00	
43	294780	9975140	2377.50	River/road junction
44A	299400	9984650	9395.00	River junction
44	-	-	2775.20	Top of hill
45A	300400	9981550	9600.00	Road corner
45	299400	9981180	2836.50	Road corner
46	298060	9979900	2704.50	River junction
48	298240	9976620	2592.00	River junction
49	301000	9979680	2835.00	River junction
50	301900	9982270	3002.30	River junction
51	300980	9984880	3112.00	Isolated hill
52	300430	9975920	2804.16	River junction
53	301120	99874200	2719.0	River junction
54	303350	9976240	3104.39	River junction
55	302450	9977640	3075.43	River junction

55A	306370	9979520	3834.38	River junction
56	304640	9984820	3750.56	Peak of mountain
57	303780	9983530	3689.60	Peak of mountain
58	308730	9979810	4038.60	River junction
58A	308700	9978470	4044.70	River junction
59	309510	9981630	4146.80	River junction
60	308810	9982740	4212.34	River junction
61	307150	9983220	3989.83	River junction
60A	-	-	4024.88	River junction
61	307150	9983220	3989.83	River junction
62	311700	9983450	5199.28	Peak of mountain
63	310570	9982850	4504.90	Road junction
63A	309250	9984600	4116.32	River junction
64	310930	9982740	4352.54	River junction
65	311870	9982160	4669.54	Road junction
66	310310	9978370	4020.31	River junction
67	311880	9977930	3931.92	River junction
67A	312830	9976.05	3659.12	River junction
68	312110	9979440	4090.42	River junction
69	312920	9981840	4325.11	River junction
70	312200	9984740	4206.24	River junction
70B	313200	9983850	4593.34	Centre line of pond

- The marked control points are then rechecked again to confirm if they are correct and accurately georeferenced. This involves again some series of adjustments on the stereo top machine and cross checking of the points until a good number of points are observed to be properly positioned.

- The process of mapping then follows once the setting is completed. The contours are then drawn and also the river drainage networks as well as other important features like roads, railway lines etc.

The above described technique of mapping is manual/analogue.

APPENDIX 2: PROCEDURE USED IN THE DIGITIZATION OF NARO MORU CATCHMENT

The digitization of Naro Moru River catchment area was done from a topographical map of the area at a scale of 1:20,000 mapped by photogrammetric means. The photogrammetric mapping was based on aerial photography at a scale of 1:70,000, with a camera having a focal length of 152.822cm and for the year 1988. Photogrammetric mapping was manual/analogue and for the same map information to be useful for GIS analysis, it had to be converted to digital form. The conversion can result in raster format by scanning or in vector format by either vectorizing/digitizing using a digitizing tablet or by first scanning and then performing on-screen vectorization. This latter alternative was the one used for this exercise. The resulting digital map was obtained by digitizing three analogue maps. Figure A2.1 shows the outline of the three map sheets. The maps sheets had the areal extents of the neat line after georeferencing as indicated in **Table A2.1**.

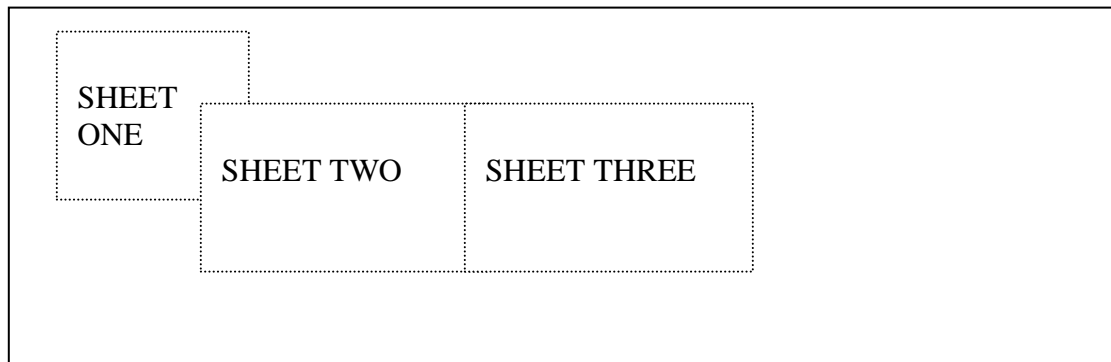


Figure A2.1 Outline of map sheets covering Naro-Moru catchment

Table A2.1 Areal extents of map sheets covering the test catchment

Corner	Sheet one		Sheet two		Sheet three	
	<i>Easting</i>	<i>Northing</i>	<i>Easting</i>	<i>Northing</i>	<i>Easting</i>	<i>Northing</i>
Top left	265637.7	9995694.4	274488.7	9988983.7	294747.9	9988949.7
Top right	279468.7	9995772.8	297586.7	9988948.9	314670.6	9988734.2
Bottom left	265425.4	9982344.5	274415.1	9975198.1	294640.2	9975035.7
Bottom right	280158.6	9983182.1	297491.1	9975200.0	314473.1	9974774.6

Scanning

Scanning is the process of converting an analogue document to digital form in raster format, i.e., in pixels. For the purpose of this project, scanning was done at a resolution of 300 dpi (*dots per inch*) at Gath Management (A consulting firm). The raster files were supplied in TIF format

Georeferencing

Since scanning and on-screen vectorization was the alternative adopted in this project, among the available GIS software packages, **ILWIS 3.0** (i.e Integrated Land and Water Information System) was chosen, because it has capabilities for georeferencing, on-screen digitization, and importing and exporting results in different file formats, especially the TIF and DXF, the industry raster and vector standards, besides having various image processing functionalities.

First and foremost, a folder that is to contain the project files was created, followed by importing the scanned topographical maps in TIF format. ILWIS version 3.0 requires uncompressed raster files. Before commencing digitization, attention is pay attention to the *coordinate system* in which the map will be digitized. One may for instance use user-defined co-ordinates, coordinates defined by a national standard or coordinates of a certain UTM zone. The information on the *map coordinates* one uses in a project is stored in a *coordinate system*. A coordinate system defines the possible minimum and maximum X's and Y's that can be used in one or more maps, along with other information on map datum, projection, etc. A coordinate system may include a *projection*, which defines the relationship between map coordinates and geographic coordinates in latitude and longitude.

Use of **Tie-Points** as a coordinate system in ILWIS entails determining relation between pixel coordinates and map co-ordinates using a conformal, affine, second order, projective, etc. transformation. The tie points/ control points used include grid intersections, which have precisely known map co-ordinates. The following **Tables (A2.2, A2.3, & A2.4)** indicate results of georeferencing each sheet.

A2.2 Georeference of map sheet 1 using second order transformation

Tie Point	X	Y	Row	Col	Δ Row	Δ Col
1	266000	9994000	1647	843	1.47	-2.13
2	268000	9984000	7534	2075	1.57	-2.26
3	272000	9984000	7542	4437	0.98	-1.59
4	276000	9986000	6375	6802	-1.98	3.05
5	278000	9992000	2850	7970	1.43	-2.24
6	270000	9994000	1649	3224	-2.22	3.10
7	274000	9994000	1658	5595	-0.39	0.79
8	266000	9986000	6352	896	-3.68	5.44
9	268000	9988000	5187	2063	2.82	-4.17

A2.3 Georeference of sheet two using Affine transformation

Tie Point	X	Y	Row	Col	Δ Row	Δ Col
1	278000	9988000	884	2741	2.54	1.50
2	286000	9988000	908	7467	-3.16	-0.51
3	296000	9986000	2126	13376	0.59	-0.90
4	278000	9976000	7943	2735	-1.19	-0.82
5	288000	9978000	6805	8645	0.81	-1.44
6	294000	9976000	8004	12194	0.41	2.17

A2.4 Georeference of map sheet three using full second order transformation

Tie Point	X	Y	Row	Col	Δ Row	Δ Col
1	296000	9986000	2170	1223	0.00	2.11
2	302000	9986000	2117	4749	-0.01	-3.69
3	310000	9986000	2048	9479	0.00	1.58
4	296000	9978000	6862	1298	0.00	-1.11
5	304000	9980000	5615	6004	0.00	0.00
6	306000	9978000	6778	7201	0.00	2.95
7	312000	9978000	6729	10749	0.00	-1.84

On-screen digitization/ on-screen vectorization

Digitization or vectorization in this sense entails following each feature with a mouse.

The following features were vectorized:

- Contours
- Spot heights
- Rivers
- Dams
- Railway
- Tracks and footpaths
- Trees/shrubs

During on-screen digitization of contours, the following omissions and commissions were encountered, but were rectified:

- Contours labeled 2300, 2340, 2360 on the lower end of sheet two were wrongly labeled, instead they should respectively have been labeled 2340, 2360 and 2370.
- Contour labeled 2540 was discontinuous about the middle.
- There was a mix up in contours labeled 2600, 2620, 2640 and 2660.
- In sheet three, the following contours were incomplete: 2820, 2860, 2900, 2940 and 3020

Throughout the three sheets, the contour interval was not constant. The contour intervals in the topographic sheets were spread (for clarity) in the following ranges

<u>Contour interval ranges</u>	<u>Contour interval</u>
1790m - 2120m	5m
2120m – 2400m	10m
2400m – 2800m	20m
2800m - 4000m	40m
4000m – 4560m	80m

spot heights were only shown in sheet one and three.

APPENDIX 3: RUNOFF CURVE NUMBERS FOR VARIOUS LAND COVER TYPES

Table A3.1 Runoff curve numbers for other agricultural watersheds

Cover type	Hydrologic Condition	<u>Hydrologic Soil Group</u>			
		A	B	C	D
Pasture, grassland, or range— continuous forage for grazing ¹	Poor	68	79	86	89
	Fair	49	69	79	84
	Good	39	61	74	80
<hr/>					
Meadow—continuous grass, protected from grazing and generally mowed for hay. - - - -		30	58	71	78
<hr/>					
Brush—brush-weed-grass mixture with brush the major element ²	Poor	48	67	77	83
	Fair	35	56	70	77
	Good	30	48	65	73
<hr/>					
Woods—grass combination (orchard or tree farm)	Poor	57	73	82	86
	Fair	43	65	76	82
	Good	32	58	72	79
<hr/>					
Woods ³	Poor	45	66	77	83
	Fair	36	60	73	79
	Good	30	55	70	77
<hr/>					
Farmsteads—buildings, lanes, driveways, and surrounding lots. - - - -		59	74	82	86

¹ *Poor*: < 50% ground cover or heavily grazed with no mulch
Fair: 50 to 75% ground cover and not heavily grazed
Good: > 75% ground cover and lightly or only occasionally grazed

² *Poor*: < 50% ground cover
Fair: 50 to 75% ground cover
Good: > 75% ground cover

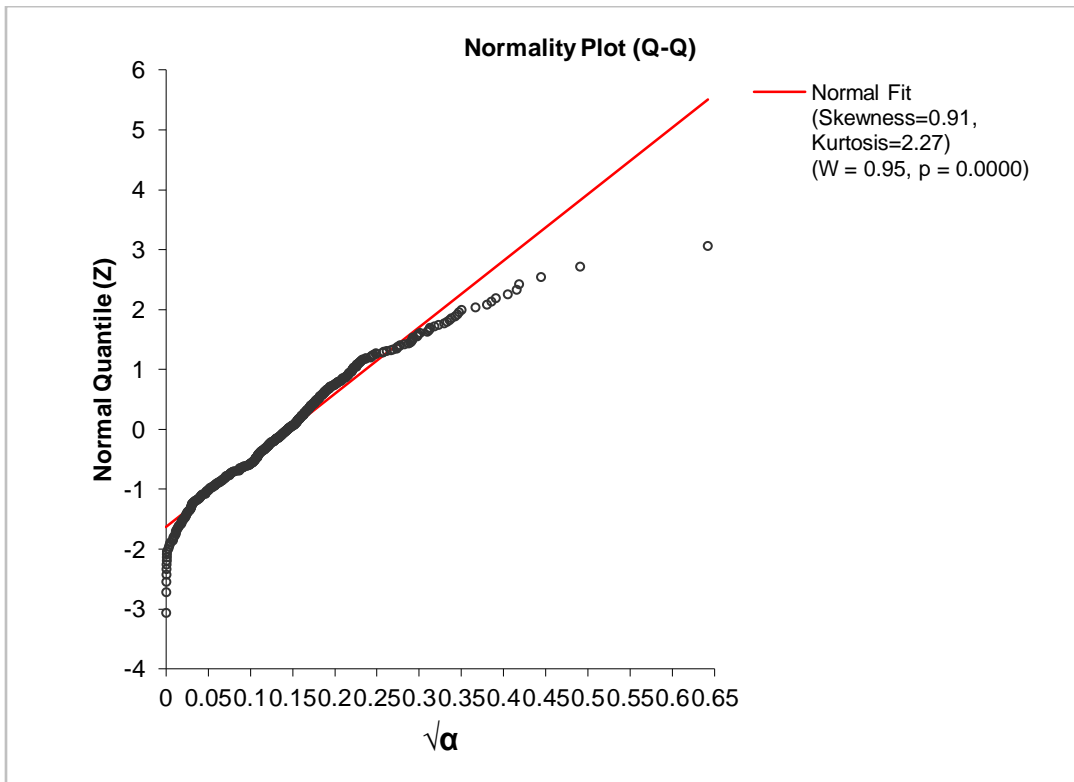
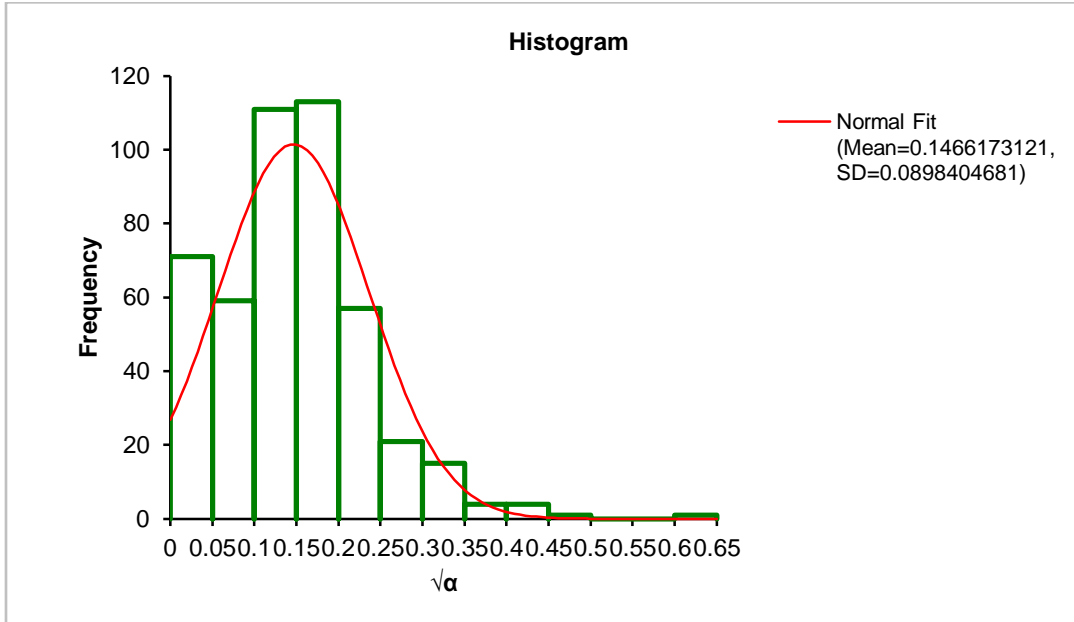
³ *Poor*: Forest litter, small trees, and brush are destroyed by heavy grazing or regular burning
Fair: Woods are grazed but not burned, and some forest litter covers the soil.
Good: Woods are protected from grazing, and litter and brush adequately cover the soil.

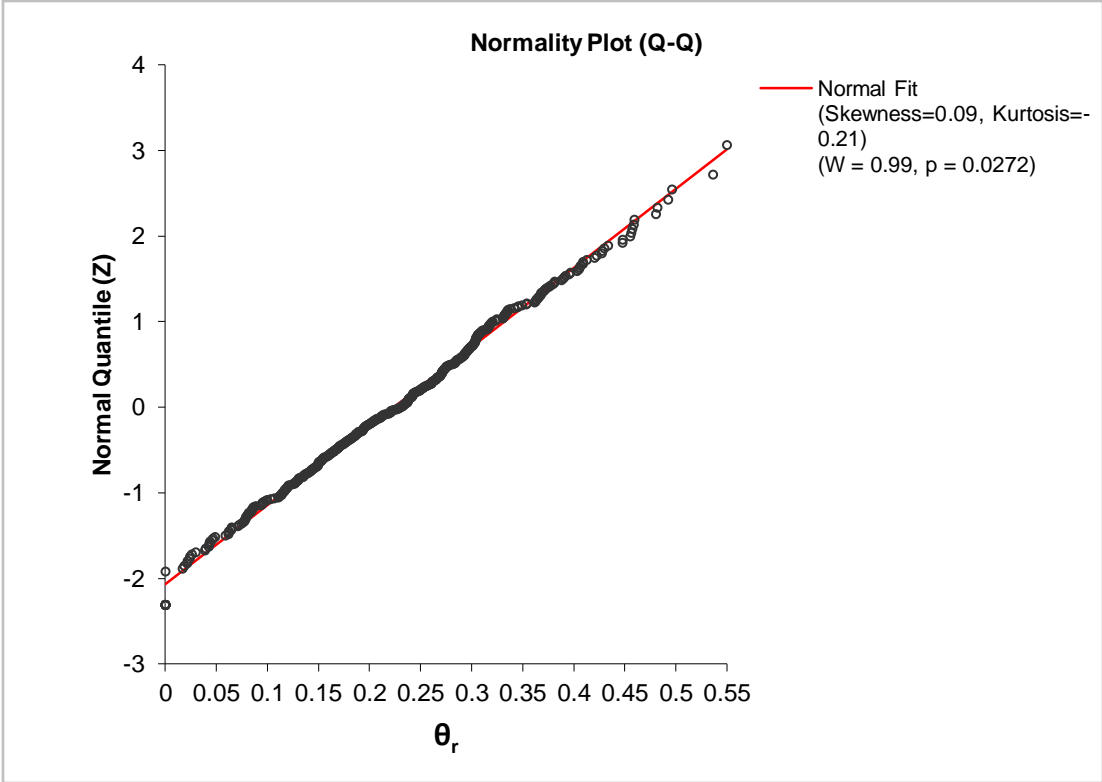
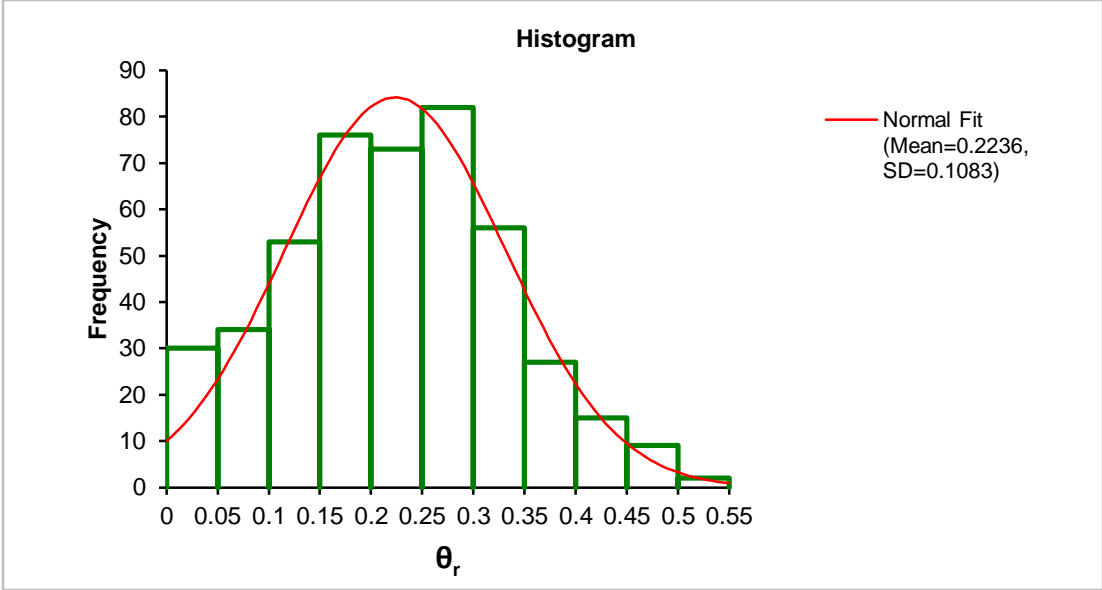
(Adapted from Neitsch et al., 2002(b))

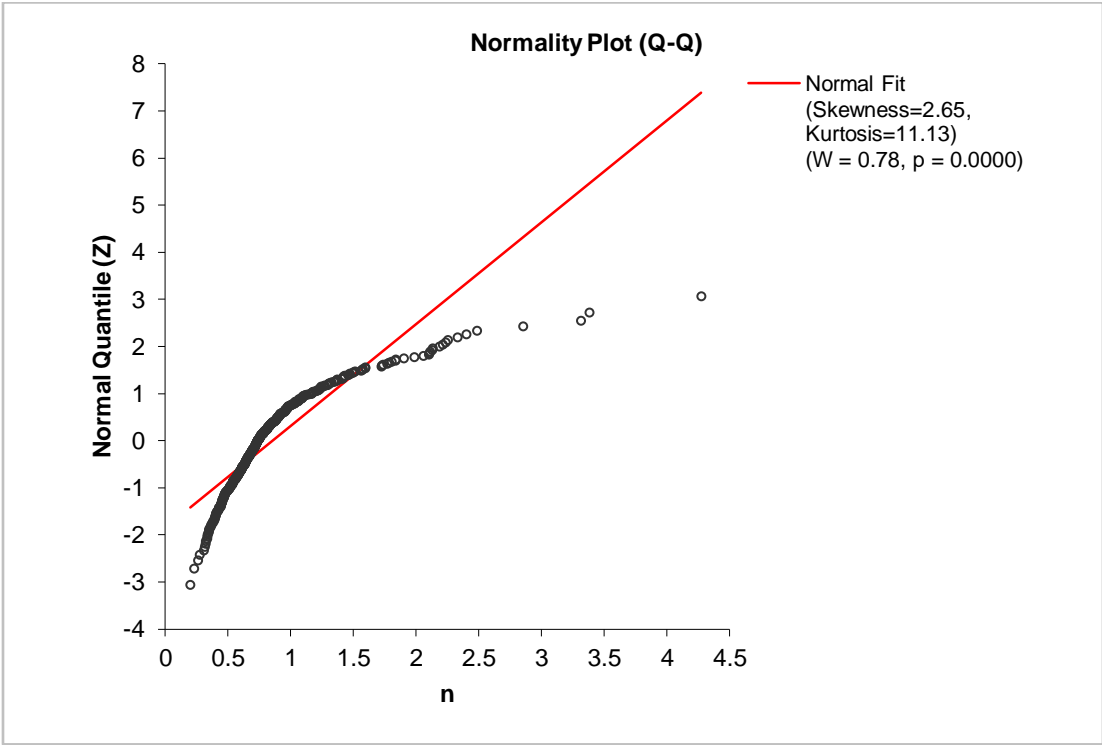
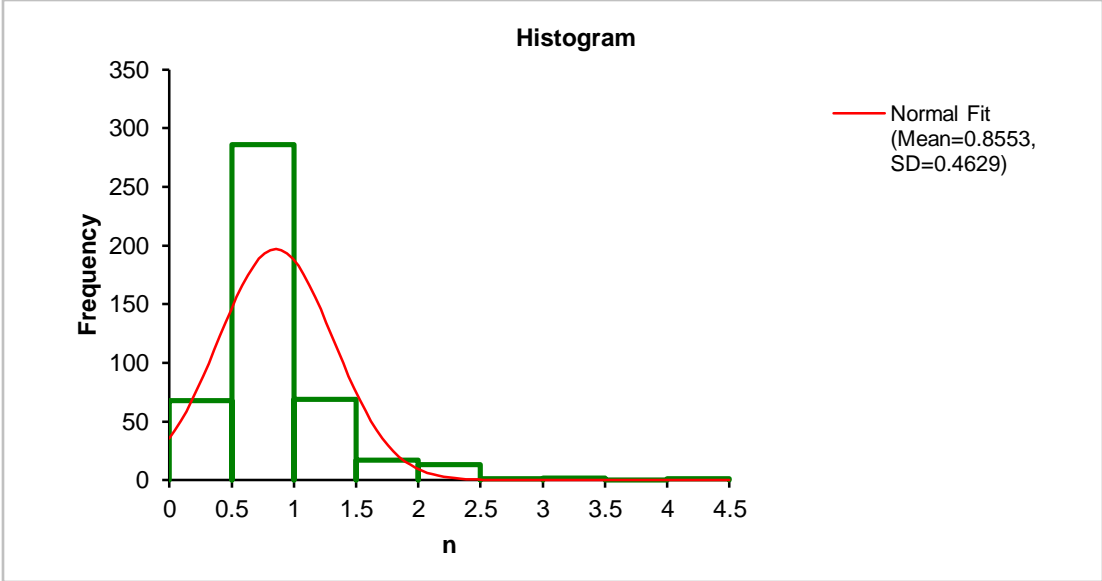
Table A3.2 Runoff curve numbers for urban areas

Cover type	Hydrologic	Hydrologic Soil Condition			
	Condition	A	B	C	D
Fully developed Urban Areas					
Open spaces (lawns, parks, golf courses, cemeteries, etc.)	Poor	68	79	86	89
	Fair	49	69	79	84
	Good	39	61	74	80
Impervious areas					
Paved parking lots, roofs, driveways, etc. (excluding right-of-way)	----	98	98	98	98
Paved streets and roads; curbs and storm sewers (excluding right-of-way)	----	98	98	98	98
Paved streets and roads; open ditches (including right-of-way)	----	83	89	92	93
Gravel streets and roads (including right-of way)	----	76	85	89	91
Dirt streets and roads (including right-of way)	----	72	82	87	89
Urban district:					
Commercial and business		89	92	94	95
Commercial and business		89	92	94	95
Industrial		81	88	91	93
Residential Districts by average lot size:					
1/8 acre (0.05 ha) or less (town houses)		77	85	90	92
1/4 acre (0.10 ha)		61	75	83	87
1/3 acre (0.13 ha)		57	72	81	86
1/2 acre (0.20 ha)		54	70	80	85
1 acre (0.40 ha)		51	68	79	84
2 acres (0.81 ha)		46	65	77	82
Developing urban areas:					
Newly graded areas (pervious areas only, no vegetation)		77	86	91	94

APPENDIX 4: PLOTS OF NORMALITY CHECK FOR RESPONSE VARIABLES







APPENDIX 5: PERFORMANCE RATINGS FOR STATISTICAL MEASURES

Table A5.1: Performance ratings for recommended statistics in flow prediction for monthly time step

Performance Rating	RSR	<u>Evaluation Statistic</u>	
		NSE	PBIAS (%)
Very Good	$0.0 \leq \text{RSR} \leq 0.50$	$0.75 < \text{NSE} < 1.00$	$\text{PBIAS} < \pm 10$
Good	$0.50 < \text{RSR} \leq 0.60$	$0.65 < \text{NSE} \leq 0.75$	$\pm 10 \leq \text{PBIAS} < \pm 15$
Satisfactory	$0.6 < \text{RSR} \leq 0.70$	$\text{NSE} \leq 0.65$	$\pm 15 \leq \text{PBIAS} \leq \pm 25$
Unsatisfactory	$\text{RSR} > 0.70$	$\text{NSE} \leq 0.50$	$\text{PBIAS} \geq \pm 25$

(Adapted from Moriasi et al., 2007).

APPENDIX 6: QUALITIES OF FIT BETWEEN VAN GENUCHTEN MOISTURE RETENTION PARAMETERS AND BASIC SOIL PROPERTIES IN LINEAR REGRESSION

Table A6.1. Quality of fit between transformations of moisture retention parameters and sand transformations in the linear regression.

	sand	sand ²	sand ³	sand ⁵	sand ¹⁰	sand ¹⁵	√sand	ln (sand)	1/sand	e ^{sand}
θ_s	r ² =0.10	r ² =0.10	r ² =0.08	r ² =0.06	r ² =0.03	r ² =0.03	r ² =0.09	r ² =0.06	r ² =0.01	r ² =0.01
θ_r	r ² =0.21	r ² =0.21	r ² =0.19	r ² =0.15	r ² =0.08	r ² =0.06	r ² =0.18	r ² =0.18	r ² =0.02	r ² =0.03
Θ_s^{-1}	r ² =0.14	r ² =0.14	r ² =0.12	r ² =0.08	r ² =0.05	r ² =0.04	r ² =0.12	r ² =0.08	r ² =0.01	r ² =0.02
ln (θ_s)	r ² =0.12	r ² =0.12	r ² =0.10	r ² =0.07	r ² =0.04	r ² =0.03	r ² =0.11	r ² =0.07	r ² =0.01	r ² =0.02
e ^{θ_r}	r ² =0.20	r ² =0.20	r ² =0.18	r ² =0.13	r ² =0.07	r ² =0.05	r ² =0.18	r ² =0.13	r ² =0.02	r ² =0.03
√ α	r ² =0.01	r ² =0.01	r ² =0.02	r ² =0.02	r ² =0.01	r ² =0.01	r ² =0.00	r ² =0.00	r ² =0.00	r ² =0.02
ln (α)	r ² =0.01	r ² =0.03	r ² =0.05	r ² =0.07	r ² =0.09	r ² =0.09	r ² =0.01	r ² =0.00	r ² =0.00	r ² =0.08
Log(n)	r ² =0.10	r ² =0.13	r ² =0.15	r ² =0.16	r ² =0.15	r ² =0.15	r ² =0.08	r ² =0.05	r ² =0.02	r ² =0.08
√n	r ² =0.11	r ² =0.15	r ² =0.18	r ² =0.21	r ² =0.20	r ² =0.12	r ² =0.08	r ² =0.05	r ² =0.01	r ² =0.12
1/n	r ² =0.07	r ² =0.08	r ² =0.09	r ² =0.08	r ² =0.09	r ² =0.06	r ² =0.06	r ² =0.04	r ² =0.01	r ² =0.03

Table A6.2. Quality of fit between transformations of the moisture retention parameters and silt transformations in the linear regression

	silt	silt ²	silt ³	silt ⁵	silt ¹⁰	silt ¹⁵	√silt	(silt) ^{1/3}	silt ^{1/5}	silt ^{1/10}	silt ^{1/20}	ln(silt)	1/silt	1/silt ²	1/silt ³
θ_s	r ² =0.01	r ² =0.00	r ² =0.00	r ² =0.00	r ² =0.00	r ² =0.00	r ² =0.02	r ² =0.02	r ² =0.02	r ² =0.03	r ² =0.03	r ² =0.03	r ² =0.03	r ² =0.02	r ² =0.01
θ_r	r ² =0.00	r ² =0.01	r ² =0.02	r ² =0.03	r ² =0.03	r ² =0.03	r ² =0.00	r ² =0.00	r ² =0.00	r ² =0.00	r ² =0.00	r ² =0.00	r ² =0.02	r ² =0.02	r ² =0.02
θ_s^{-1}	r ² =0.01	r ² =0.00	r ² =0.02	r ² =0.00	r ² =0.00	r ² =0.00	r ² =0.02	r ² =0.02	r ² =0.02	r ² =0.03	r ² =0.03	r ² =0.03	r ² =0.04	r ² =0.02	r ² =0.01
ln(θ_s)	r ² =0.01	r ² =0.00	r ² =0.00	r ² =0.00	r ² =0.00	r ² =0.00	r ² =0.02	r ² =0.02	r ² =0.02	r ² =0.03	r ² =0.03	r ² =0.03	r ² =0.04	r ² =0.02	r ² =0.01
e ^{θ_r}	r ² =0.00	r ² =0.01	r ² =0.02	r ² =0.03	r ² =0.04	r ² =0.03	r ² =0.00	r ² =0.00	r ² =0.00	r ² =0.00	r ² =0.00	r ² =0.00	r ² =0.02	r ² =0.02	r ² =0.02
√ α	r ² =0.01	r ² =0.01	r ² =0.00	r ² =0.02	r ² =0.02	r ² =0.02	r ² =0.00	r ² =0.00	r ² =0.00	r ² =0.00	r ² =0.00	r ² =0.00	r ² =0.02	r ² =0.02	r ² =0.02
Ln(α)	r ² =0.00	r ² =0.00	r ² =0.00	r ² =0.00	r ² =0.01	r ² =0.01	r ² =0.00	r ² =0.00	r ² =0.00	r ² =0.01	r ² =0.01	r ² =0.01	r ² =0.07	r ² =0.09	r ² =0.09
Log(n)	r ² =0.02	r ² =0.01	r ² =0.00	r ² =0.00	r ² =0.00	r ² =0.01	r ² =0.04	r ² =0.05	r ² =0.06	r ² =0.08	r ² =0.08	r ² =0.07	r ² =0.13	r ² =0.09	r ² =0.07
√n	r ² =0.03	r ² =0.01	r ² =0.00	r ² =0.00	r ² =0.00	r ² =0.00	r ² =0.05	r ² =0.06	r ² =0.07	r ² =0.08	r ² =0.08	r ² =0.09	r ² =0.18	r ² =0.14	r ² =0.12
1/n	r ² =0.02	r ² =0.01	r ² =0.00	r ² =0.00	r ² =0.00	r ² =0.00	r ² =0.03	r ² =0.03	r ² =0.04	r ² =0.04	r ² =0.04	r ² =0.04	r ² =0.06	r ² =0.03	r ² =0.02

Table A6.3. Quality of fit between transformations of the moisture retention parameters and silt transformations in the linear regression.

	BD	BD ²	BD ³	BD ⁵	√BD	Ln(BD)	1/BD	e ^(BD)
θ _s	r ² =0.56	r ² =0.56	r ² =0.53	r ² =0.43	r ² =0.54	r ² =0.51	r ² =0.38	r ² =0.55
θ _r	r ² =0.09	r ² =0.09	r ² =0.09	r ² =0.10	r ² =0.09	r ² =0.08	r ² =0.05	r ² =0.09
θ _s ⁻¹	r ² =0.54	r ² =0.59	r ² =0.60	r ² =0.56	r ² =0.49	r ² =0.44	r ² =0.28	r ² =0.59
ln (θ _s)	r ² =0.56	r ² =0.59	r ² =0.58	r ² =0.50	r ² =0.53	r ² =0.48	r ² =0.34	r ² =0.58
e ^{θ_r}	r ² =0.09	r ² =0.10	r ² =0.09	r ² =0.08	r ² =0.09	r ² =0.08	r ² =0.06	r ² =0.10
√α	r ² =0.02	r ² =0.02	r ² =0.02	r ² =0.02	r ² =0.02	r ² =0.01	r ² =0.01	r ² =0.02
Ln(α)	r ² =0.01	r ² =0.01	r ² =0.01	r ² =0.01	r ² =0.00	r ² =0.00	r ² =0.00	r ² =0.01
Log(n)	r ² =0.00	r ² =0.00	r ² =0.00	r ² =0.00	r ² =0.00	r ² =0.00	r ² =0.00	r ² =0.00
√n	r ² =0.00	r ² =0.00	r ² =0.01	r ² =0.01	r ² =0.00	r ² =0.00	r ² =0.00	r ² =0.01
1/n	r ² =0.00	r ² =0.00	r ² =0.00	r ² =0.00	r ² =0.00	r ² =0.00	r ² =0.00	r ² =0.00

Table A6.4. Quality of fit between transformations of the moisture retention parameters and clay transformation in the linear regression analysis

	clay	clay ²	clay ³	clay ⁵	√clay	(clay) ^{1/3}	ln(clay)	1/clay	1/clay ²	e ^(clay)
θ_s	r ² =0.07	r ² =0.08	r ² =0.09	r ² =0.07	r ² =0.08	r ² =0.02	r ² =0.06	r ² =0.05	r ² =0.01	r ² =0.00
θ_r	r ² =0.30	r ² =0.26	r ² =0.20	r ² =0.16	r ² =0.34	r ² =0.23	r ² =0.31	r ² =0.07	r ² =0.26	r ² =0.01
√ α	r ² =0.02	r ² =0.01	r ² =0.01	r ² =0.01	r ² =0.04	r ² =0.02	r ² =0.04	r ² =0.04	r ² =0.00	r ² =0.01
Ln(α)	r ² =0.02	r ² =0.01	r ² =0.01	r ² =0.01	r ² =0.03	r ² =0.02	r ² =0.04	r ² =0.03	r ² =0.03	r ² =0.00
θ_s^{-1}	r ² =0.11	r ² =0.12	r ² =0.11	r ² =0.09	r ² =0.12	r ² =0.05	r ² =0.10	r ² =0.05	r ² =0.00	r ² =0.00
ln (θ_s)	r ² =0.09	r ² =0.11	r ² =0.10	r ² =0.08	r ² =0.12	r ² =0.04	r ² =0.08	r ² =0.08	r ² =0.00	r ² =0.00
e ^{θ_r}	r ² =0.29	r ² =0.26	r ² =0.22	r ² =0.16	r ² =0.32	r ² =0.22	r ² =0.29	r ² =0.06	r ² =0.00	r ² =0.01
Log(n)	r ² =0.05	r ² =0.02	r ² =0.01	r ² =0.01	r ² =0.08	r ² =0.05	r ² =0.10	r ² =0.05	r ² =0.00	r ² =0.00
√n	r ² =0.05	r ² =0.02	r ² =0.01	r ² =0.01	r ² =0.02	r ² =0.06	r ² =0.11	r ² =0.05	r ² =0.00	r ² =0.00
1/n	r ² =0.03	r ² =0.02	r ² =0.01	r ² =0.00	r ² =0.06	r ² =0.04	r ² =0.07	r ² =0.03	r ² =0.00	r ² =0.00

Table A6.5. Quality of fit between the moisture retention parameters and organic carbon transformations in the linear regression analysis

	OrgC	OrgC ²	√OrgC	OrgC ^{1/3}	ln(OrgC)	e ^(OrgC)	1/OrgC	1/OrgC ²	1/OrgC ³
θ _s	r ² =0.20	r ² =0.07	r ² =0.30	r ² =0.22	r ² =0.30	r ² =0.02	r ² =0.06	r ² =0.01	r ² =0.00
θ _r	r ² =0.00	r ² =0.00	r ² =0.01	r ² =0.00	r ² =0.01	r ² =0.00	r ² =0.01	r ² =0.01	r ² =0.00
√α	r ² =0.01	r ² =0.00	r ² =0.01	r ² =0.01	r ² =0.01	r ² =0.01	r ² =0.01	r ² =0.00	r ² =0.00
Ln(α)	r ² =0.01	r ² =0.00	r ² =0.01	r ² =0.01	r ² =0.01	r ² =0.01	r ² =0.03	r ² =0.03	r ² =0.00
θ _s ⁻¹	r ² =0.14	r ² =0.04	r ² =0.24	r ² =0.19	r ² =0.27	r ² =0.01	r ² =0.06	r ² =0.01	r ² =0.00
ln(θ _s)	r ² =0.17	r ² =0.06	r ² =0.27	r ² =0.21	r ² =0.29	r ² =0.01	r ² =0.07	r ² =0.01	r ² =0.00
e ^{θ_r}	r ² =0.00	r ² =0.00	r ² =0.01	r ² =0.00	r ² =0.01	r ² =0.01	r ² =0.01	r ² =0.01	r ² =0.00
Log(n)	r ² =0.01	r ² =0.00	r ² =0.01	r ² =0.00	r ² =0.01	r ² =0.00	r ² =0.01	r ² =0.02	r ² =0.02
√n	r ² =0.01	r ² =0.00	r ² =0.01	r ² =0.00	r ² =0.01	r ² =0.00	r ² =0.03	r ² =0.03	r ² =0.03
1/n	r ² =0.00	r ² =0.00	r ² =0.00	r ² =0.00	r ² =0.00	r ² =0.00	r ² =0.00	r ² =0.00	r ² =0.00

APPENDIX 7: ANALYSIS OF CROSS CORRELATION RESULTS FOR PREDICTOR VARIABLES IN PEDOTRANSFER FUNCTION DEVELOPMENT

Table A7.1 Cross correlation tests for transformed predictor variables for $\frac{1}{\theta_s}$

	sand	1/silt	BD ³	$\sqrt{\text{clay}}$	LN(OrgC)
sand	r=1	r=0.42 95%CI (0.32-0.50) t statistic (8.43) 2-tailed p <0.0001	r=0.26 95%CI (0.16-0.36) t statistic (5.0) 2-tailed p <0.0001	r=-0.71 95%CI (-0.76 to -0.65) t statistic (-18.53) 2-tailed p<0.0001-	r=-0.10 95%CI (-0.20-0.01) t statistic (-1.81) 2-tailed p <0.0001
1/silt	r=0.42 95%CI (0.32-0.50) t statistic (8.43) 2-tailed p <0.0001	r=1	r=0.19 95%CI (0.08 to 0.29) t statistic (3.55) 2-tailed p = 0.0004	r=-0.13 95%CI (-0.24 to -0.03) t statistic (-2.49) 2-tailed p=0.0132	r=-0.25 95%CI (-0.35 to -0.15) t statistic (-4.76) 2-tailed p <0.0001
BD ³	r=0.26 95%CI (0.16-0.36) t statistic (5.0) 2-tailed p <0.0001	r=0.19 95%CI (0.08 to 0.29) t statistic (3.55) 2-tailed p = 0.0004	r=1	r=-0.28 95%CI (-0.37 to -0.17) t statistic (-5.26) 2-tailed p<0.0001	r=-0.59 95%CI (-0.65 to -0.51) t statistic (-13.28) 2-tailed p <0.0001
$\sqrt{\text{clay}}$	r=-0.71 95%CI (-0.76 to -0.65) t statistic (-18.53) 2-tailed p<0.0001	r=-0.13 95%CI (-0.24 to -0.03) t statistic (-2.49) 2-tailed p= 0.0132	r=-0.28 95%CI (-0.37 to -0.17) t statistic (-5.26) 2-tailed p<0.0001	r=1	r=0.08 95%CI (-0.02 to 0.19) t statistic (1.53) 2-tailed p =0.1271
ln(OrgC)	r=-0.10 95%CI (-0.20-0.01) t statistic (-1.81) 2-tailed p <0.0001	r=-0.25 95%CI (-0.35 to -0.15) t statistic (-4.76) 2-tailed p <0.0001	r=-0.59 95%CI (-0.65 to -0.51) t statistic (-13.28) 2-tailed p <0.0001	r=0.08 95%CI (-0.02 to 0.19) t statistic (1.53) 2-tailed p =0.1271	r=1

Table A7.2 Cross correlation tests for transformed predictor variables in e^{0r}

	sand	Silt ¹⁰	BD ²	$\sqrt{\text{clay}}$	ln(OrgC)
sand	r=1	r=-0.25 95%CI (-0.35 to -0.16) t statistic (-4.81) 2-tailed p <0.0001	r=0.23 95%CI (0.13 to 0.33) t statistic (4.41) 2-tailed p< 0.0001	r=-0.71 95%CI (-0.76 to -0.65) t statistic (-18.53) 2-tailed p<0.0001	r=-0.10 95%CI (-0.20-0.01) t statistic (-1.81) 2-tailed p <0.0001
Silt ¹⁰	r=-0.25 95%CI (-0.35 to -0.16) t statistic (-4.81) 2-tailed p <0.0001	r=1	r=0.06 95%CI (-0.04 to 0.17) t statistic (1.15) 2-tailed p= 0.2528	r=-0.21 95%CI (-0.31 to 0.11) t statistic (-4.01) 2-tailed p< 0.0001	r=-0.09 95%CI (-0.20 tp 0.01) t statistic (-1.71) 2-tailed = 0.0882
BD ²	r=0.23 95%CI (0.13 to 0.33) t statistic (4.41) 2-tailed p< 0.0001	r=0.06 95%CI (-0.04 to 0.17) t statistic (1.15) 2-tailed p= 0.2528	r=1	r=-0.24 95%CI (-0.34 to -0.13) t statistic (-4.50) 2-tailed p<0.0001	r=-0.62 95%CI (-0.68 to -0.55) t statistic (-14.31) 2-tailed p<0.0001
$\sqrt{\text{clay}}$	r=-0.71 95%CI (-0.76 to -0.65) t statistic (-18.53) 2-tailed p<0.0001	r=-0.21 95%CI (-0.31 to 0.11) t statistic (-4.01) 2-tailed p< 0.0001	r=-0.24 95%CI (-0.34 to -0.13) t statistic (-4.50) 2-tailed p<0.0001	r=1	r=0.08 95%CI (-0.02 to 0.19) t statistic (1.53) 2-tailed p =0.1271
ln(OrgC)	r=-0.10 95%CI (-0.20-0.01) t statistic (-1.81) 2-tailed p <0.0001	r=-0.09 95%CI (-0.20 tp 0.01) t statistic (-1.71) 2-tailed = 0.0882	r=-0.62 95%CI (-0.68 to -0.55) t statistic (-14.31) 2-tailed p<0.0001	r=0.08 95%CI (-0.02 to 0.19) t statistic (1.53) 2-tailed p =0.1271	r=1

Table A7.3 Cross correlation tests for transformed predictor variables in $\sqrt{\alpha}$.

	e^{sand}	$silt^{-2}$	BD	$clay^{-1}$	$\ln(\text{OrgC})$
e^{sand}	r=1	r=0.89 95%CI (0.86 to 0.91) t statistic (35.22) 2-tailed p<0.0001	r=0.13 95%CI (0.03 to 0.24) t statistic () 2-tailed p	r=0.30 95%CI (0.20 to 0.39) t statistic (5.74) 2-tailed p<0.0001	r=-0.21 95%CI (-0.31 to -0.11) t statistic (-3.98) 2-tailed p<0.0001
$silt^{-2}$	r=0.89 95%CI (0.86 to 0.91) t statistic (35.22) 2-tailed p<0.0001	r=1	r=0.14 95%CI (0.03 to 0.24) t statistic (2.58) 2-tailed p=0.0103	r=0.23 95%CI (0.12 to 0.33) t statistic (4.32) 2-tailed p<0.0001	r=-0.24 95%CI (-0.33 to -0.13) t statistic (-4.44) 2-tailed p<0.0001
BD	r=0.13 95%CI (0.03 to 0.24) t statistic () 2-tailed p	r=0.14 95%CI (0.03 to 0.24) t statistic (2.58) 2-tailed p=0.0103	r=1	r=0.22 95%CI (0.12 to 0.32) t statistic (4.14) 2-tailed p<0.0001	r=-0.64 95%CI (-0.70 tp -0.57) t statistic (-15.13) 2-tailed p<0.0001
$clay^{-1}$	r=0.30 95%CI (0.20 to 0.39) t statistic (5.74) 2-tailed p<0.0001	r=0.23 95%CI (0.12 to 0.33) t statistic (4.32) 2-tailed p<0.0001	r=0.22 95%CI (0.12 to 0.32) t statistic (4.14) 2-tailed p<0.0001	r=1	r=0.08 95%CI (-0.02 to 0.19) t statistic (1.53) 2-tailed p=0.5271
$\ln(\text{OrgC})$	r=-0.21 95%CI (-0.31 to -0.11) t statistic (-3.98) 2-tailed p<0.0001	r=-0.24 95%CI (-0.33 to -0.13) t statistic (-4.44) 2-tailed p<0.0001	r=-0.64 95%CI (-0.70 tp -0.57) t statistic (-15.13) 2-tailed p<0.0001	r=0.08 95%CI (-0.02 to 0.19) t statistic (1.53) 2-tailed p=0.5271	r=1

Table A7.4. Cross correlation tests for transformed predictor variables in \sqrt{n}

	$Sand^5$	$silt^{-1}$	$\ln(\text{clay})$	OrgC^{-2}
$Sand^5$	r=1	r=0.74 95%CI (0.68 to 0.78) t statistic (20.09) 2-tailed p<0.0001	r=-0.69 95%CI (-0.74 to -0.63) t statistic (-17.60) 2-tailed p<0.0001	r=0.34 95%CI (0.24 to 0.43) t statistic (6.64) 2-tailed p<0.0001
$Silt^{-1}$	r=0.74 95%CI (0.68 to 0.78) t statistic (20.09) 2-tailed p<0.0001	r=1	r=-0.25 95%CI (-0.35 to -0.15) t statistic (-4.81) 2-tailed p<0.0001	r=0.52 95%CI (0.43 to 0.59) t statistic (11.08) 2-tailed p<0.0001
$\ln(\text{clay})$	r=-0.69 95%CI (-0.74 to -0.63) t statistic (-17.60) 2-tailed p<0.0001	r=-0.25 95%CI (-0.35 to -0.15) t statistic (-4.81) 2-tailed p<0.0001	r=1	r=-0.21 95%CI (-0.31 to -0.10) t statistic (-3.92) 2-tailed p=0.0001
OrgC^{-2}	r=0.34 95%CI (0.24 to 0.43) t statistic (6.64) 2-tailed p<0.0001	r=0.52 95%CI (0.43 to 0.59) t statistic (11.08) 2-tailed p<0.0001	r=-0.21 95%CI (-0.31 to -0.10) t statistic (-3.92) 2-tailed p=0.0001	r=1