

**RECONSTRUCTION OF
STREAMFLOW INTO LAKE
NAIVASHA USING CREST MODEL
AND REMOTE SENSED
RAINFALL AND
EVAPOTRANSPIRATION**

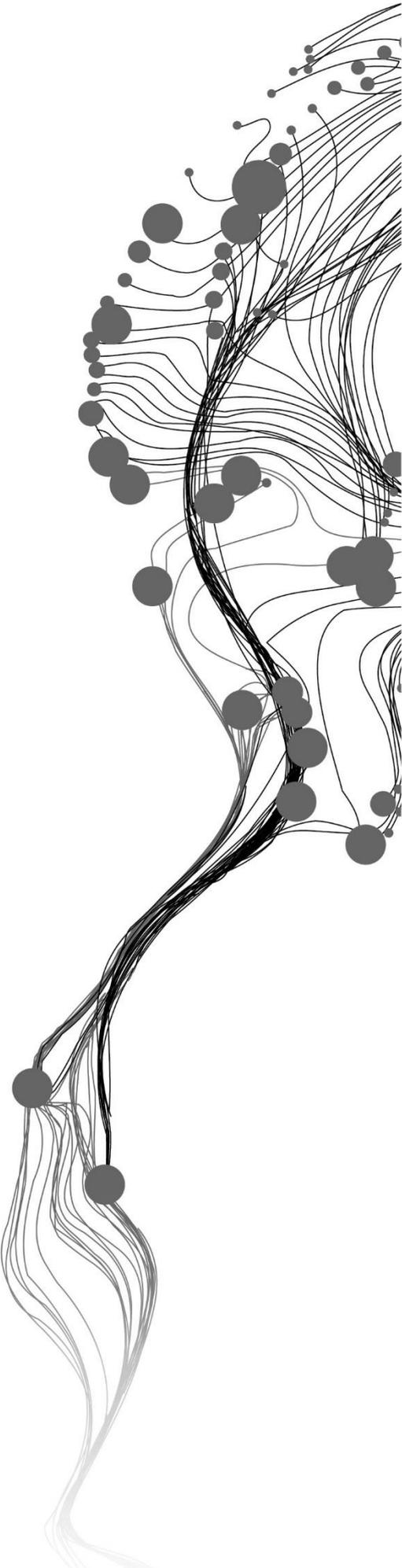
HENRY MUNYAKA GATHECHA

February, 2015

SUPERVISORS:

Dr.ir. R. van der Velde

Drs. R. Becht



RECONSTRUCTION OF STREAMFLOW INTO LAKE NAIVASHA USING CREST MODEL AND REMOTE SENSED RAINFALL AND EVAPOTRANSPIRATION

HENRY MUNYAKA GATHECHA

Enschede, The Netherlands,

February, 2015

This thesis submitted to the Faculty of Geo-Information Science and Earth Observation of the University of Twente in partial fulfillment of the requirements for the degree of Master of Science in Geo-information Science and Earth Observation.

Specialization: Water Resources and Environmental Management

THESIS ASSESSMENT BOARD:

Dr. M.C. Lubczynski (Chair)

Dr. J. van der Kwast (External Examiner, Unesco-IHE)

Dr.ir. R. van der Velde

Drs. R. Becht

DISCLAIMER

This document describes work undertaken as part of a programme of study at the Faculty of Geo-Information Science and Earth Observation of the University of Twente. All views and opinions expressed therein remain the sole responsibility of the author, and do not necessarily represent those of the Faculty.

ABSTRACT

This study aims at reconstructing the streamflow into Lake Naivasha using the CREST model (a distributed hydrologic model developed recently by the University of Oklahoma and NASA) and satellite observed rainfall inputs of TRMM3B42-v7 and CMORPH8km. In general, this study entailed the acquisition of different input data and their consequent pre-processing and quality assessment, CREST model setup, calibration and validation, and assessment of the overall streamflow input to the Lake Naivasha water balance. The streamflow was reconstructed for the period 2001 to 2010 by simulating the streamflow for the three main rivers draining into Lake Naivasha namely, Malewa, Karati and Gilgil. Although the CREST model uses the Shuffled Complex Evolution Algorithm (SCUE-A) as its inbuilt calibration procedure, in this study the PEST calibration procedure (which uses the Gauss-Marquardt-Levenberg (GML)) algorithm was used. The PEST procedure is efficient in the number of model runs and can be modified to cater for numerical problems resulting from correlation and parameter insensitivity.

The reliability assessment of both the TRMM-3B42-v7 and the CMORPH8km rainfall products was done by comparing their seasonal trends, cumulative volumes and scatterplots with the in-situ rainfall data from averages of 5 raingauges selected within the basin for a period of 10 years. The assessment was done for both the daily averages, as well as, the monthly averages to gain insights on the temporal variation of errors in the data. TRMM3B42-v7 produces the most frequent rain events of the three rainfall datasets, while CMORPH8km produces the least rainfall events but with the highest peaks. Also TRMM3B42-v7 produced the highest cumulative rainfall. The daily scatterplots were very noisy depicting the presence of temporal and spatial errors. This was further affirmed by the monthly scatterplots that were smooth and not very biased as the daily scatterplots, showing that the spatial and temporal errors were smoothed in the long run. TRMM3B42-v7 performed better for the cases of the Root Mean Square Error (daily RMSE=4.21mm, monthly= 1.77mm) and the Mean Absolute Error (Daily MAE=2.48mm, monthly= 0.883mm), than the CMORPH8km (daily RMSE=5.57mm, monthly= 1.77mm) and (Daily MAE=2.69mm, monthly= 2.1mm). On the other hand, CMORPH8km had a less bias (daily Bias= -0.334, monthly= -0.403) than the TRMM3B42-v7 (daily Bias=-0.708, monthly=-0.635).

Assessment of the in-situ discharge for the three subcatchments showed some mismatches between the seasonal dynamics in the discharge and the rainfall data, signifying presence of errors in the observation data. The study shows that the CREST model was able to reproduce the general streamflow dynamics, as well as the volumes of the streamflow from the three subcatchments in the Naivasha basin. The results from the calibration and validation of the three rivers in the Naivasha basin were better while using the TRMM3B42-v7 rainfall than when using the CMORPH8km, in terms of the seasonal flow dynamics, as well as, the volumes. Also, most of the optimum parameter values while using the PEST calibration procedure, relates closely with those obtained from literature. However, the base flow produced by the CREST model increased continuously, which is not the case with the observed discharge.

The streamflow simulated from the satellite derived rainfall were assessed for their contribution to the lake level change using the water balance for the lake against the in-situ observations. The water balance equation was used with the change in storage as the change in lake level, streamflow and rainfall on the lake surface as the input and evaporation from the lake surface as the output. Simulations from the TRMM3B42-v7 yielded better RMSE (9.57mm) and MAE (5.70mm) than those of CMORPH (RMSE=10.89 and MAE=6.15). The streamflow was seen to decrease drastically from late 2008 to early 2010, after which it started recovering.

Keywords: Lake Naivasha, CREST model, TRMM3B42-v7, CMORPH8km, streamflow, reconstruction, PEST

ACKNOWLEDGEMENTS

First and foremost, I would like to thank the Dutch government for offering me the NFP (Netherlands Fellowship Program) Scholarship, for it is because of this scholarship that I managed to pursue the MSc course in Water Resources Management.

Secondly, I want to sincerely thank my supervisors, Dr. Ir. Rogier van der Velde and Dr. Robert Becht for their immense support throughout my thesis process. I am particularly grateful to Dr. Rogier for shaping and directing this research in all its stages, providing many insights, ideas, and study materials, as well as, developing the PEST calibration procedure as used in this study. Dr. Rogier, the discussions, disagreements, and most importantly the time we spent together during the course of this study will go a long way in my future research work. I count myself very lucky for having worked with you. Also, Dr. Robert, your critical views during the study period and facilitating for the study materials and data is highly acknowledged. Thanks very much for providing information about Lake Naivasha basin and also providing the much helpful link with WARMA-Naivasha.

Thirdly, I wish to thank Mr. Vincent .O. Odongo for his contribution in this thesis. Most importantly, Mr. Vincent Odongo provided the CREST V2.0 software used in this study, and facilitated for acquirement of the rainfall and river discharges data used in the study. Apart from the data and study material for the Naivasha area, Mr. Vincent contributed immensely in the success of this study and we discussed a lot of issues that helped in the accomplishment of this study.

Fourth, I want to thank Dr. Xianwu Xue of the University of Oklahoma and the University of Oklahoma in general that developed the CREST model. I am particularly grateful to Dr. Xianwu Xue for continuously providing support with regard to the CREST.V2 model including technical support and study materials, as well as, the source code for the model.

I am also thankful to Ms Faith Mitheu of the Regional Centre for Mapping of Resources for Development (Nairobi Kenya). Thank you Faith for providing study materials for the CREST model and also for providing the python code for processing the TRMM3B42-v7 data. The discussions we had with you over the period of setting up the CREST model are highly appreciated.

I am also grateful to Dr. B.H.P. Maathuis who assisted in developing batch processes in ILWIS software for the pre-processing of CMORPH8km rainfall, as well as other technical assistance with regard to the ILWIS software. Likewise, I am thankful to the ITC water department lecturers and staff members in general for teaching me and facilitating my studies in this noble course of Water Resources and Environmental management.

My acknowledgements cannot be complete without appreciating my friends who contributed directly to my research, including Felix Orina, Damas Mbagi, Dominique Wambua, Josyline Makena, Ir. Webster Ngumi, Ir. Lucy Chepkosgei, David Ongo, Gloria, Benson, Tabalia, Muchiri, Kibet and Grace Njenga. Also thanked are my classmates in the Water Resources Department for the wonderful time that we had together. Thank you all for the many discussions we had and for making my stay in the Netherlands enjoyable. By the same token, I am grateful to all my friends in ITC, both students and staff for your contribution of any kind to the easy and enjoyable time I had in the Institution.

Finally, I acknowledge my family in Kenya for the moral support and encouragement they provided throughout the time that I was in the Netherlands.

TABLE OF CONTENTS

List of figures.....	iv
List of tables	vi
List of acronyms	vii
1. Introduction.....	1
1.1. Background.....	1
1.2. Objectives	2
1.3. Research Questions.....	2
1.4. Research Method.....	2
1.5. Outline of the Thesis	3
2. Literature review	4
2.1. Hydrologic modelling	4
2.2. Satellite data for hydrologic modelling	4
2.3. Model selection:.....	4
2.4. Sensitivity analysis	5
2.5. Model Calibration.....	6
3. Study area and in situ data.....	8
3.1. Description of the Study Area	8
3.2. In-situ data.....	8
4. Satellite data	14
4.1. FEWSNET PET	14
4.2. HydroSHEDS	15
4.3. TRMM3B42-V7.....	15
4.4. CMORPH-8KM.....	15
5. CREST model	17
5.1. Model structure.....	17
5.2. Implementation.....	20
5.3. Coupling PEST to CREST	23
5.4. Calibration	24
6. Rainfall data assessment.....	27
6.1. Daily rainfall assessment	27
6.2. Monthly Rainfall Assessment.....	29
7. Results and discussion.....	32
7.1. Calibration and validation	32
7.2. Model calibration and validation while using TRMM3B42-V7 rainfall	32
7.3. Model calibration and validation while using CMORPH8km rainfall	38
7.4. Streamflow Flow into Lake Naivasha	41
7.5. Lake Naivasha Water balance.....	42
8. Final Remarks	46
8.1. Conclusions	46
8.2. Recommendations.....	47
9. Appendices.....	52
Appendix A: The procedure followed in obtaining streamflow data.....	52
Appendix B: The ILWIS batch process for downloading and processing yearly daily FEWSNET PET	53
Appendix C: Python Script to convert NetCDF files to ASCII format.....	54
Appendix D: ILWIS batch process for downloading and processing CMORPH8km-30min.....	55
Appendix E: The batch process used to convert the hourly CMORPH8km-1hr data to daily	56
Appendix F: HWSD soil texture lookup table	58

LIST OF FIGURES

Figure 1: The CREST modelling flowchart for reconstruction of streamflow into Lake Naivasha using the TRMM3B42-v7, CMORPH8km and FEWSNET PET.....	3
Figure 2: Map showing the hydrological and topographical features of the Lake Naivasha basin.....	8
Figure 3: Plot for daily average gauge rainfall for the Lake Naivasha basin for the period 2001 to 2010.....	9
Figure 4: Figure showing some inconsistencies between the gauge rainfall and discharge data for Gilgil for the period 2001 and 2006.....	11
Figure 5: Figure showing the impacts of the inconsistencies between the ground rainfall and discharge to the flow cumulative curve for Gilgil sub-catchment.....	11
Figure 6: Plots for in-situ rainfall and discharge data for Malewa River, with demarcations showing suspicious events.....	12
Figure 7: Plots for in-situ rainfall and discharge data for Karati River, with demarcations showing suspicious events.....	12
Figure 8: Flow chart for the CMORPH8km data processing.....	16
Figure 9: Schematic Representation of the CREST model v2.0 showing two vertical reservoirs, the variable infiltration curve and some of the processes involved in the model.....	18
Figure 10: Schematic representation of the organization of files for the CREST V2.0 model as used in this study.....	23
Figure 11: Schematic representation of the PEST calibration procedure showing how PEST encapsulates the CREST model.....	23
Figure 12: The Naivasha basin's sub-catchments as they were extracted using the CREST model based on the outlet points chosen depending on the most downstream available discharge data for each of the rivers in the basin.....	24
Figure 13: The CREST model parameter sensitivity analysis plots for Malewa sub-catchment.....	25
Figure 14: Flowchart for the steps followed in the calibration process for the 3 subcatchments in the Lake Naivasha basin.....	26
Figure 15: The time series data trends of the daily gauge, TRMM3B42-V7 and CMORPH8km for 2001 and 2002, over the Lake Naivasha Basin, showing the seasonal dynamics of rainfall.....	28
Figure 16: Cumulative rainfall plots for daily gauge, TRMM3B42V7 and CMORPH-8Km for the Lake Naivasha Basin from 2001 to 2010, showing volume relationship of the three rainfall products over the area.....	28
Figure 17: Scatter plots for daily; gauge versus TRMM3B42-v7 and gauge versus CMORPH8km over the Lake Naivasha Basin for 2001 to 2010.....	28
Figure 18: The time series data trends of the monthly gauge, TRMM3B42-V7 and CMORPH8km for 2001 and 2010.....	30
Figure 19: Cumulative rainfall plots for monthly rainfall of gauge, TRMM3B42V7 and CMORPH8Km.....	30
Figure 20: Scatter plots for monthly gauge versus TRMM3B42-v7 and gauge versus CMORPH8km.....	30
Figure 21: Streamflow hydrographs for calibration and validation of Malewa, Gilgil and Karati subcatchments using TRMM3B42-v7 rainfall.....	33
Figure 22: Streamflow mass curves for calibration and validation for Malewa, Gilgil and Karati subcatchments using TRMM3B42-v7 rainfall.....	34
Figure 23: Streamflow hydrographs for calibration and validation for Malewa, Gilgil and Karati using CMORPH8km rainfall.....	38
Figure 24: Streamflow mass curves for calibration and validation for Malewa, Gilgil and Karati subcatchments using CMORPH8km rainfall.....	39

Figure 25: Streamflow hydrographs for the flow into Lake Naivasha as modelled using the CREST v2.0 distributed model and TRMM3B42-v7 and CMORPH8km rainfall..... 41

Figure 26: Streamflow mass curves for the flow into Lake Naivasha as modelled using the CREST v2.0 distributed model and TRMM3B42-v7 and CMORPH8km rainfall..... 42

Figure 27: Plots for the daily lake level as calculated from the stream inputs using in-situ data, TRMM3B42-v7 simulations and CMORPH8km simulations. The plots shows a good coherence between the Lake levels calculated using TRMM3B42-v7 simulations and that calculated using the in-situ data..... 43

Figure 28: Plots for the Lake Naivasha daily levels from 2002 to 2010, showing the comparison between the obtained lake levels from the streamflow simulations and calculations from in-situ data against the observed lake levels 44

LIST OF TABLES

Table 1: Gap analysis of the selected raingauge stations for the period 2001 to 2010	9
Table 2: Table showing the location of the discharge stations used in this study	10
Table 3: Original streamflow gap analysis for the period 2001 to 2010 for the selected gauge stations.....	10
Table 4: Table showing the various satellite data used in this study as well as their sources.....	14
Table 5: Table showing the parameters in the CREST V2.0 model, and their respective categories.....	17
Table 6: Showing the statistical measures for the comparison of daily TRMM3B42-v7 and CMORPH8km rainfall products against the raingauge observations	29
Table 7: Table showing the relationship between monthly average rainfall values for the TRMM3B42-v7 and CMORPH 8km rainfall products with the monthly average values of the raingauge observations.	31
Table 8: Calibration and validation statistics for Malewa, Gilgil, and Karati while using TRMM3B42-V7 rainfall.....	35
Table 9: Optimum parameters for the Malewa, Gilgil, and Karati subcatchments while using TRMM3B42-v7 rainfall	36
Table 10: Calibration and validation statistics for Malewa, Gilgil, and Karati while using CMORPH8km rainfall.....	40
Table 11: Optimum parameters for Malewa, Gilgil and Karati while using CMORPH-8km rainfall	40
Table 12: The statistical measures of accuracy for the calculated Lake levels using the TRMM 3B42-v7 simulations as well as using the CMORPH8km simulations against the calculated streamflow input into the lake using the in-situ data	43
Table 13: Values for the factor representing the water abstracted form the lake as well as the ground water contribution to the lake level in (mm/day).	44
Table 14: Statistical measures of accuracy for the relationship between the calculated lake levels from the three cases of streamflow against the lake levels observed in the lake level station by WARMA-Naivasha.	45

LIST OF ACRONYMS

AMSR-E	Advanced Microwave Scanning Radiometer
AMSU	Advanced Microwave Sounding Unit
BSQ	Band Sequential
CMORPH	CPC Morphing technique
CPC	Climate Prediction Centre
CREST	Coupled Routing and Excess Storage
DMSP	Defense Meteorological Satellite Program
FEWSNET	Famine Early Warning Systems Network
GDAS	Global Data Assimilation System
GES	Gorrdard Earth Sciences
GML	Gauss-Marquard-Levenberg algorithm
GPCC	Global Precipitation Climatology Centre
HydroSHEDS	Hydrological data and maps based on Shuttle Elevation Derivatives at multiple Scales
ILWIS	Integrated Land and Water Information System
IR	Infra-Red
ITCZ	Intertropical Convergence Zone
JAXA	Japanese Space Agency
LEO	Low Earth Orbit
LSA	Local Sensitivity Analysis
NASA	National Aeronautics and Space Administration
PEST	Parameter Estimation Tool
SCUE-A	Shuffled Complex Evolution Algorithm
SRTM	Shuttle Radar Topography Mission
SSM/I	Special Sensor Microwave Imager Instrument
SWAT	Soil Water Assessment Tool
TMI	TRMM Microwave Imager
TMPA	TRMM Multiple-Satellite Precipitation Analysis
TRMM	Tropical Rainfall Measuring Mission
USGS	United States Geological Survey
WARMA	Water Resources Management Authority
WGS	World Geodetic System
WWF	World Wildlife Fund

1. INTRODUCTION

1.1. Background

Lake Naivasha is a fresh water Rift Valley lake, located in Kenya (0° 45' S, 36° 20' E). The lake Naivasha basin has a wide range of terrestrial flora, and fauna as well as aquatic organisms, which support numerous anthropogenic activities in the basin thus forming a social ecological system; with strong and unique interdependence mechanism (Everard & Harper, 2002; Harper & Mavuti, 2004). Lake Naivasha is an important Ramsar site, according to the Ramsar Convention 2011 (Harper & Mavuti, 2004). As a Ramsar site, the lake's conservation and wise use of its resources falls under the Ramsar Convention.

Being located near the equator, the climate of Lake Naivasha region is influenced by the oscillations of the intertropical convergence zone (ITCZ) (Becht & Harper, 2002; Bergner, Trauth, & Bookhagen, 2003). Thus, the area has two rain seasons; the large one is from March to May and the small one is from October to December (Bergner, Trauth, & Bookhagen, 2003). The rainy seasons are characterized by rapid overland runoffs, while the long dry periods are characterized by severe droughts and famine (Everard, Vale, Harper, & Tarras-wahlberg, 2002; Golden et al., 2014; Khan, Sadiq et al., 2011). The region's local climate is, however, modified by the effects of the inland lake, land cover/land use, its proximity to the Indian Ocean, as well as topographic variations (Bergner et al., 2003). The prevailing weather cycle patterns have impacts on food security, infrastructure, human health, tourism, wildlife and other sectors (World Wide Fund for Nature & Kenya Wildlife Services, 2012).

The availability of fresh water from the lake basin has encouraged the development of a very vibrant commercial horticulture and floriculture industry over the last two decades. Other socioeconomic activities evidenced in the lake Naivasha basin include direct activities including tourism, pastoralism, and fisheries. There are also indirect socioeconomic activities within the vibrant urban center (Naivasha town) with service industries such as banks, transport and other businesses.

The water resources management of the Lake Naivasha basin is characterized by a complex socio-ecological aspect consisting of many non-linear and interrelated hydrological, economic, ecological and agronomic processes. This system poses issues including increasing competitive water supply demand, degrading water catchment areas, water quality issues, increasing population, as well as administrative issues. The Water Act (2002) places the management of the Lake Naivasha basin water resources under the Water Resources Authority (WARMA-Naivasha), which has developed a water allocation plan, to incorporate all the emerging issues with regard to this system (WARMA, 2010). According to the Integrated Water Resources Management and Water Efficiency Plan for Kenya (2009), the management of Lake Naivasha basin and indeed all the water resources in Kenya need to be equitable, sustainable and efficient (WARMA, 2009).

The recent past has seen the Lake Naivasha and its stream inputs experiencing rising demands for their depleting water resources, in an unsustainable manner. Reports show that most wetlands in the East Africa region have experienced increasing population since the 1950s (e.g. Everard & Harper, 2002; Harper & Mavuti, 2004; Obiero, 2011). The increased population in the lake Naivasha basin, coupled with unsustainable development, have led to encroachment of forest land and riparian habitat by agriculture or settlement (Bergner et al., 2003; Everard et al., 2002; Kuhn, van Oel, Pieter, & Meins, Frank, 2012; Odongo et al., 2014). Everard, Vale, Harper, & Tarras-wahlberg (2002) poses a need to explain the observed decrease in the lake Naivasha water level. The WARMA Naivasha is, therefore, faced by the need to quantify the hydrological processes in the area in order to make informed decisions and policies.

Although many studies have been performed with regard to the hydrological processes in the Lake Naivasha basin (e.g. Becht & Harper, 2002; Becht, Science, & Itc, n.d.; Everard & Harper, 2002; Muthuwatta, 2004; Odongo et al., 2014; Tiruneh, 2004; WARMA, 2010), the following gaps were noted and they motivated this study:

1. There is a clear lack of streamflow models in Naivasha catchment that caters for the impacts of the rapidly varying topographic features, land uses, soil types and weather parameters to streamflow in the catchment area
2. The raingauge data as used by previous models for the area had uncertainties due to the temporal and spatial gaps involved thereby imposing discontinuity in the streamflow produced

1.2. Objectives

The main objective of this study is to reconstruct the streamflow into Lake Naivasha using the CREST model with precipitation and potential evapotranspiration data obtained from satellite based data products. In order to meet the main objective, the following specific objectives were developed:

1. To assess the reliability of CMORPH-8km and TRMM3B42-V7 rainfall datasets as input for the CREST model in the Lake Naivasha basin
2. To assess the reliability of the available discharge data for the Lake Naivasha basin
3. To assess the performance of the CREST model in Lake Naivasha basin while using CMOPRH8km and TRMM3B42-V7 rainfall products
4. To investigate the temporal dynamics of the streamflow simulated by the CREST model for the Lake Naivasha Basin.

1.3. Research Questions

1. How does the CMORPH8km and TRMM3B42-v7 relate to the observed raingauge data in the Lake Naivasha basin?
2. How does the observed discharge data relate to the rainfall as observed from the raingauge data?
3. How is the performance of the CREST model in reproducing the in-situ measured streamflow of the three subcatchments in the Lake Naivasha basin using TRMM3B42-v7 and CMORPH8km rainfall products?
4. How does the overall simulated streamflow into Lake Naivasha as produced by the CREST model relate to the overall observed streamflow?

1.4. Research Method

The research method for this project involved data acquirement and their consequence pre-processing and preparation for their respective input into the model. Other methods included, the CREST model preparation, input data quality assessment, sensitivity assessment, model calibration and validation, data analysis and drawing of conclusion. The in-situ data of rainfall and river discharge data were obtained from the ITC archive as they were used in previous studies (Meins, 2013). Data quality assessment was done for the river discharge data based on the in-situ rainfall data. The satellite data of TRMM3B42-V7, CMORPH-8km and FEWSNET potential evapotranspiration were downloaded from their respective sources and processed as described in their respective sections in this thesis. Data quality assessment was done for the TRMM3B2-V7 and the CMORPH-8km rainfall data by comparing them with the in-situ rainfall. The HydroSHEDS topographic data were downloaded from the HydroSHEDS website, and prepared for the study area and in the appropriate format as required by the CREST model.

Setting up of the CREST model for the Naivasha basin involved separation of the three subcatchments contributing to the streamflow into the lake. These subcatchments are Malewa, Gilgil, and Karati. The discharge locations for each of the subcatchments were selected according to the location of the discharge

measurements. For each of the subcatchments, the CREST model files were prepared as well as input data of the two rainfall products and the potential evapotranspiration. The sensitivity analysis of the CREST model parameters was carried out based on the default values of the parameters so as to determine the calibration strategy. The PEST calibration files were then prepared for each of the model setup. Afterwards, calibration was done for all the three subcatchments for both satellite derived rainfall products. The validation process was then carried out and the results for the calibration and validation processes were analysed and reported. Finally the streamflow into the lake was obtained by combining all the streamflows contributed by each of the three subcatchments.

The following flowchart illustrated the research method (figure1)

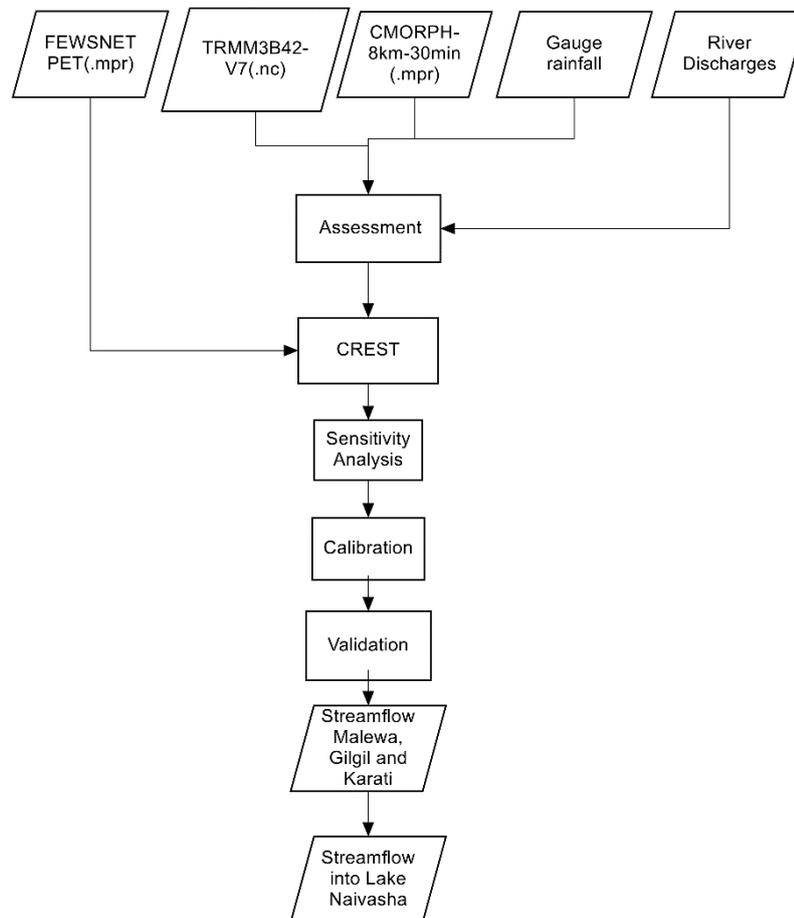


Figure 1: The CREST modelling flowchart for reconstruction of streamflow into Lake Naivasha using the TRMM3B42-v7, CMORPH8km and FEWSNET PET

1.5. Outline of the Thesis

Chapter 2 presents literature review of prior works that build the basis for the choice of the various methods and strategies used in this study. Chapter 3 contains details of the study area and the in-situ data used. Chapter 4 explains the satellite data, while chapter 5 discusses the CREST model, structure, implementation, and calibration. Chapter 6 discusses the rainfall data analysis for the various three rainfall products used in this study, i.e. gauge, CMORPH8km, and the TRMM3B42-v7 rainfall. Chapter 7 contains the results and discussions for the various model runs carried out in the study. Finally, chapter 8 discusses the final remarks of the research including conclusions and recommendations.

2. LITERATURE REVIEW

2.1. Hydrologic modelling

Hydrologic modeling involves balancing of the water budget in a given space domain over a given period of time (Gupta, S, 2011). Overtime, hydrologic models have helped in understanding the relationship between climate, hydrologic cycle, and water resources (Khan, Sadiq et al., 2010). In many cases, hydrologic models have played eminent role in understanding the hydrologic processes thus enabling skillful management as well as prediction of events and processes (Gupta, S, 2011). Also, hydrologic models can be utilized in reanalyzing past events, such as those of streamflows, to gain insights on the hydrologic trends and to bridge gaps in researches. Hydrologic models have continued to play critical roles in researches, planning, design, operation as well as management of water resources (Muleta & Nicklow, 2005).

2.2. Satellite data for hydrologic modelling

Many studies elaborate on the usefulness of satellite remote sensing data for hydrologic modeling (e.g. Artan et al., 2007; Deng, Bowman, & Jackson, 2007; Sadiq I. Khan et al., 2011; Li, Zhang, & Xu, 2012; Meng, Li, Hao, Wang, & Shao, 2014; Montzka et al., 2008). Notably, use of satellite data is attractive because of 1) minimum interruption, 2) cost effectiveness, 3) availability even in remote areas, 4) saves time and energy, and 5) good spatial and temporal resolution. Satellite data is thus crucial as many studies have cited lack of reliable geospatial data as challenges curbing distributed hydrologic modeling (e.g. Bergner, Trauth, & Bookhagen, 2003; Druyan & Fulakeza, 2013; Everard, Vale, Harper, & Tarraswahlberg, 2002; Muthuwatta, 2004; Obiero, 2011). Khan et al. (2012) emphasized the need for modeling studies involving accurate spatial and temporal information, on climatological and hydrological variables in solving the current and future trends of unsustainable water resource utilization in Kenya.

The recent past has witnessed development of satellite observed precipitation products with continuous improvement in terms of spatial coverage, temporal and spatial resolution (You, Liu, Wang, & Cao, 2011). Therefore, the various satellite-observed rainfall products have varying characteristics. For example, microwave satellite rainfall products have high spatial resolution but low temporal resolution (Bajracharya et al., 2014). On the other hand, infrared satellite rainfall products have moderately high spatial resolution. The PERSIANN-Cloud Classification System (PERSIANN-CSS), is an infrared (IR) precipitation estimate available at 0.25° spatial and 3-hourly, and 6-hourly temporal resolutions (National Center for Atmospheric Research Staff (Eds), 2013; Sorooshian, Hsu, Imam, & Hong, 2005). The dominant passive-microwave (PMW) include the National Oceanic and Atmospheric Administration (NOAA) Climate Prediction Center (CPC) morphing technique (CMORPH) which is available every 30 minutes and 8km as well as hourly 0.25° (Al, 2004). Also, NOAA incorporates IR- and PWM to produce the 3-hourly, 0.25° Tropical Rainfall Measuring Mission (TRMM) Multi-satellite Precipitation Analysis (TMPA). Notably, the TMPA products were optimized recently by gauge adjustment and are available both as real time (3B42RT) and post real time (3B42V7).

2.3. Model selection:

Some hydrologic models are lumped, while others are distributed. There are lumped empirical (black box) models as well as lumped conceptual models (Gupta, S, 2011). (Muleta & Nicklow, 2005) highlights the usefulness of distributed hydrologic models over lumped hydrologic model in accounting for

heterogeneity of environmental variables such as topographic features, land uses, soil types and weather parameters. The US Department of Agriculture's (USDA) Soil and Water Assessment Tool (SWAT) is an example of a largely used distributed hydrological model (Meins, 2013). Notably, Meins (2013) modeled the water balance for Lake Naivasha using (SWAT) and cited lack of accurate rainfall data as a challenge for distributed hydrologic modeling of the lake. Coupled Routing and Excess Storage (CREST) is another distributed hydrologic model (Xue et al., 2013). The CREST model simulates the spatial and temporal variation of land surface and subsurface water fluxes and storages by cell to cell basis (Wang et al., 2011). According to Wang et al., (2011) the primary water fluxes such as infiltration and routing are conceptualized such that they physically relate to the spatially variable land surface characteristics (i.e. topography, soil type and vegetation etc.). Also, the runoff generation and routing scheme are coupled, thereby, enabling realistic interaction between the "lower atmospheric boundary layers, terrestrial surface, and subsurface water" (Xue & Hong, 2013). The CREST model is also scalable through sub-grid soil moisture storage capacity (via a variable infiltration curve) and multi-linear reservoirs for multi scale runoff generation (Xue & Hong, 2013).

The CREST model has been used successfully to assess the performance of satellite obtained rainfall products. For example, Xue et al. (2013) used the CREST model to assess the suitability of Tropical Rainfall Measuring Mission (TRMM) 3B42V7 precipitation product for use in ungauged basins in the Wangchu region. Their general observation was that the TRMM 3B42V7 had a higher correlation with gauge rainfall simulations. Also, Khan et al. (2012) used CREST model to supplement stream gauges with remote-sensing data from microwave sensors for large sparsely gauged or ungauged basins. The CREST model is used for operational flood forecasting by SEVIR East Africa (Limaye, Gitau, & Kabuchanga, 2011).

2.4. Sensitivity analysis

According to (Zhan, Song, Xia, & Tong, 2013) Uncertainty analysis (UA) is the determination of the uncertainty that derives from uncertainty in model factors (model parameters and the model state variables). On the other hand, Sensitivity Analysis (SA) is the determination of the contributions of individual different sources of uncertain inputs to the uncertainty in the output of the model (Mishra, 2009). Sensitivity analysis enables the identification of key input parameters that contribute the most to the model's predictive uncertainty (Zhan et al., 2013).

Analysis of uncertainty in hydrologic model is important in determining the reliability of the model. In many cases, use of the best-guess or worst-case assumptions about model inputs are utilized in the traditional deterministic analysis of uncertainty in hydrologic models, to quantify their impacts on model predictions (Trucano, Swiler, Igusa, Oberkampf, & Pilch, 2006). A more objective based approach involves utilizing a set of optimistic and pessimistic values to provide upside and downside forecasts around a reference scenario (Wainwright, Finsterle, Jung, Zhou, & Birkholzer, 2014). Evidently, such approach is simplistic and is not capable of dealing with complex problems where the model parameters have significant correlations or the system response is nonlinear. Moreover, the systematic combinations of optimistic and pessimistic values can also lead to confidence intervals that are too wide and whose reliability is difficult to assess (Chen, Jin, & Sudjianto, 2005).

Sensitivity analysis can either be local or global. Local sensitivity analysis (LSA) compute or approximate the local response of the model outputs by varying input factors or parameters at some nominal settings, known as the "baseline" or "nominal value" point, in the hyperspace of the input factors (Wainwright et al., 2014). On the other hand, global sensitivity analysis (GSA) evaluates the effects of input variations on the outputs in the entire allowable ranges of the input space (Zhan et al., 2013). For this study, local

sensitivity analysis was done to gain insight on the contribution of each parameter so as to reduce the number of parameters requiring calibration.

2.5. Model Calibration

Models are simplification of reality and no matter how sophisticated they are, they entail conceptualization and empiricism. This renders the reliability of models to depend on the model assumptions and algorithms, inputs details and quality, as well as parameter estimates (Muleta & Nicklow, 2005). Thus, model calibration plays a vital role in determining the model reliability. Model calibration can be manual or automatic. Two classes of automatic calibration procedures are local and global search procedures based on the evolving pattern of the solutions (Blasone, Madsen, & Rosbjerg, 2007). The local search calibration techniques have a definite direction (deterministic), and may involve the steepest descent or the simplex downhill method (gradient-based) (Blasone et al., 2007). On the other hand, the global search seeks a population of solutions using both stochastic and deterministic rules. The global methods thus take advantage of the robustness of the stochastic random search and definite direction to converge into the space containing the objective function's optima (Zabinsky, 2003). An example of a local search calibration method is the gradient-based Gauss-Marquard-Levenberg (GML) algorithm as implemented in the PEST software by (Doherty, 1994). On the other hand, an example of global search is the Shuffled Complex Evolution (SCE) algorithm developed by Duan, Sorooshian, & Gupta, (1994)

Notably, increasing model complexity results in increased number of parameters to be calibrated and the need to account for the spatial distribution of the catchment's hydrological properties (Blasone et al., 2007). This increases the parameter dimensionality thus increasing the number of model runs needed for the calibration of the model. According to Gupta, S,(2011), attainment of a sound model for a complex hydrologic model requires multi-objective model calibration. Reducing the dimensionality of calibration reduces the number of parameters to be calibrated, according to "the principle of parsimony" (Blasone et al., 2007). In such a case, the estimated parameters are sufficient to ensure a satisfactory model fit. The selection of the few parameters should however follow an extensive sensitivity analysis to determine those parameters with the most pronounced effect on the model response.

Blasone, Madsen, and Rosbjerg, (2007) investigated the performance of a global and a local optimization technique, i.e., the Shuffled Complex Evolution algorithm and the gradient based Gauss-Marquard-Levenberg algorithm respectively, in calibration of physically based distributed models of different complexity. Their motivation was the emerging problems with automatic calibration procedures for distributed hydrologic models including computational time, parameter identifiability, large number of parameters, model response surface complexity, and handling of parameter compensation (equifinality) and multiple objectives(Blasone et al., 2007). The Gauss-Marquard-Levenberg (GML) algorithm involves optimization based on gradient descent. The algorithm seeks to minimize the sum of the squared deviations between the outputs and the corresponding observations (Blasone et al., 2007).

The GML method is implemented in the PEST software and is used in this study. Notably, the GML optimization method directs the basic gradient estimation process. The basic gradient method searches new parameter estimates by shifting the current set along the direction of the maximum improvement of the objective function. Thus the Levenberg-Marquardt variation corrects the direction and the length of the parameter upgrade vector, such that the search is not trapped within the proximity of the optimum without actually achieving it (Blasone et al., 2007). The PEST software is thus widely applicable for nonlinear parameter estimation and it has found particular application in distributed hydrological models such as MIKE SHE (Blasone et al., 2007).

Skahill and Doherty, (2006) compared the advantages and disadvantages of PEST method compared to global optimization procedure like SCE. For the sake of this study, the particular advantages of PEST over the inbuilt SCE procedure in the CREST model is the possibility of modifying the procedure in cases of potential numerical problems resulting from correlation and parameter insensitivity and its efficiency in the number of the model runs. On the other hand, the notable disadvantages of the PEST procedure in this study lies on the dependency of the optimization result on the point of initialization of the search, and the procedure being trapped in local objective function minima (Blasone et al., 2007). To overcome this disadvantage, a calibration strategy was designed by grouping the parameters in 2 classes, calibrating for one class and using the obtained optimum point as the starting point for the next class and looping on until a satisfactory result is obtained or the calibration yields no more significant change.

3. STUDY AREA AND IN SITU DATA

3.1. Description of the Study Area

Lake Naivasha is a fresh water body located between latitudes 0° 42' S, 0° 48' S and longitudes 36° 20' E, 36° 26' E (WGS84) (see figure 2 below). The lake Naivasha basin is located at latitude, 0° 10' S, 0° 55' S and longitude, 36° 10' E, 36° 40' E. Two main rivers; Malewa and Gilgil, originating from the eastern escarpment flow into the lake. The basin has three subcatchments that contribute to the flow into the lake, which are Malewa, Gilgil and Karati. The lake area is 140km², the Malewa catchment area is 1849km², the Gilgil catchment is 370km² and that of the Karati sub-catchment is 146km². Thus, the total basin area is about 2500km². The Malewa River is the main contributor to the streamflow into the lake, contributing for about 80% of the streamflow. River Gilgil contributes for about 10% and the remainder is contributed by the seasonal streams including the Karati. The maximum altitude of the Naivasha basin is 3990 m above mean seal level (a.m.s.l) on Eastern side of the Aberdares ranges. The minimum altitude is 1980m (a.m.s.l), located at the Rift Valley floor. The major soils in the escarpment are of volcanic origin,

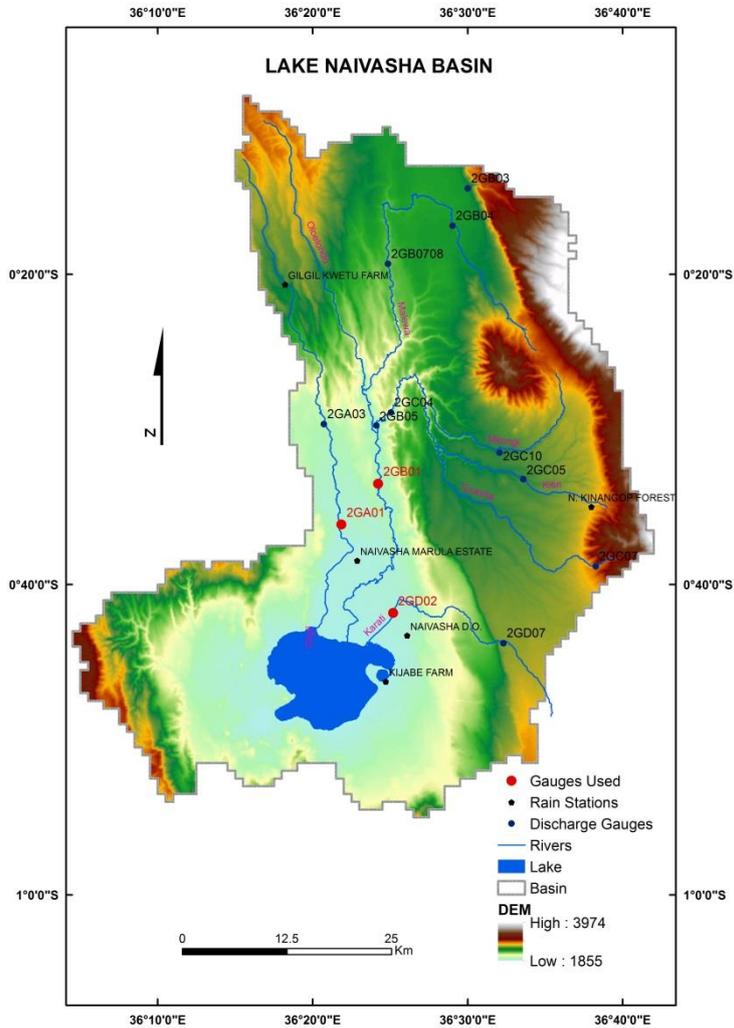


Figure 2: Map showing the hydrological and topographical features of the Lake Naivasha basin

Ashes of major old volcanoes (Odongo et al., 2014). Notably, the soils are deep (1.2-1.8m) and well drained. The rainfall distribution changes rapidly due to topographic effects. At the Rift valley floor, it is 600 mm per annum, and goes up to 1700 mm per annum at the Aberdares ranges. The mean annual temperatures varies from 8°C to 30°C (Odongo, Onyando, Mutua, van Oel, & Becht, 2013).

3.2. In-situ data

The in-situ data used in this study are:

1. Measured rainfall data
2. Measured river discharge data
3. Monthly pan evaporation data for the lake
4. Daily lake levels

3.2.1. Measured Rainfall data

The gauge rainfall data was used to validate the TRMM3B42-v7 and the CMORPH8km rainfall data. Raingauge data from 1960 to 2010 is available for 65 rain stations in the area from ITC archive as used by Meins, (2013) in the study for the area. The original dataset was incomplete as it had gaps and only two stations had continuous dataset for the study period, i.e. Gilgil Kwetu and Kijabe farm. The quality assessment of the observed rainfall data entailed gap analysis from which, out of the 65 rain stations available only 5 stations had significantly reliable data (significant here means at most 50% gaps). The chosen stations were also representative of the various regions in the catchment, namely the upper catchment and the lower catchment. The upper catchment was not well represented though as no station was selected from the Northeastern part of the basin. The gap analysis of the 5 stations chosen was as shown in table1 below. The daily averages of the data from these stations was obtained and used to validate the satellite rainfall products.

Table 1: Gap analysis of the selected raingauge stations for the period 2001 to 2010

Station name	Location in decimal degrees(WGS84)	Station ID	N ^o of days	N ^o gaps	% gaps
Naivasha D.O.	36.436275 ^o , -0.721085 ^o	9036002	3588	64	0.018
Gilgil Kwetu farm	36.302655 ^o , -0.344527 ^o	9036999	3588	0	0
Kijabe farm	36.411507 ^o , -0.771107 ^o	9036666	3588	0	0
Naivasha Marula estate	36.381397 ^o , -0.640467 ^o	9036109	3588	1823	0.508
N. Kinangop forest station	36.633448 ^o , -0.584132 ^o	9036025	3588	364	0.101

The daily average rainfall for the selected five stations was as shown in the plot in figure 3. From the figure, the 2 rainy seasonal trends per year are evident. Also, some rainfall outliers can be seen, which could be indicators of unique events or errors in the data. Nevertheless, the in-situ rainfall data was assumed to be the ground truth, the uncertainties due to the interpolation notwithstanding.

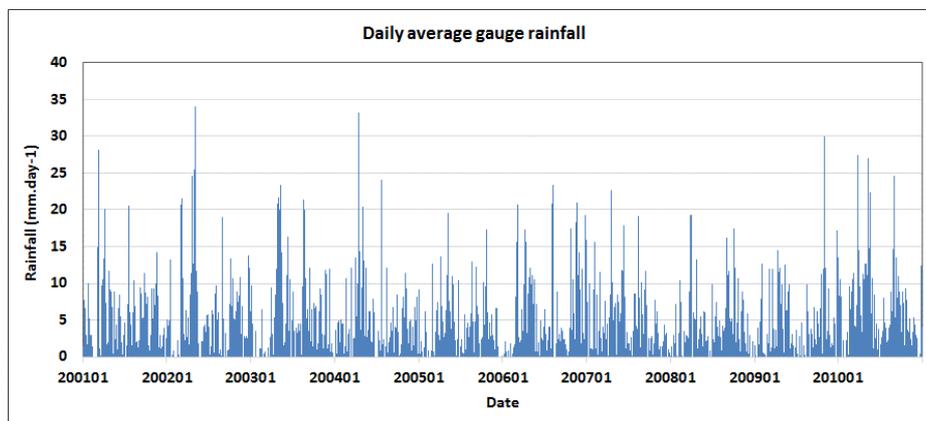


Figure 3: Plot for daily average gauge rainfall for the Lake Naivasha basin for the period 2001 to 2010

3.2.2. River discharge data

River discharge is used in calibrating and validating the CREST model. Discharge has been observed at several gauging stations within the streams in the Lake Naivasha basin and is available from 1960 to December 2014. In this study, the discharge data adopted is as used in a study by Meins, (2013). The process used in obtaining the discharge data by Meins, (2013) is shown in appendix (A). For this study, the most downstream gauging stations with appreciable continuous discharge data were chosen for each stream. Thus, three stations were used for the calibration in this study with the values for 2GB01 obtained by adding the data from 2GC04 and 2GB05. The information of the discharge stations used in this study is presented in table 2.

Table 2: Table showing the location of the discharge stations used in this study

Station	River	Location in decimal degrees (WGS84)	Elevation (m)
2GC04	Turasha	36.41701 ^o , -0.480988 ^o	2000
2GB05	Malewa	36.401537 ^o , -0.495355 ^o	1987
2GB01	Malewa	36.403195 ^o , -0.558904 ^o	1951
2GD02	Karati	36.419773 ^o , -0.697605 ^o	1896
2GA01	Gilgil	36.362856 ^o , -0.602006 ^o	1920

During the fieldwork, the reconnaissance was done for the area and all the river gauge stations were visited. It was noted that the reliability of the discharge data was compromised by siltation, burglary of the gauging station materials, and unreliable readers. Over time, siltation in the stream courses has affected the calibrations of the gauge staffs. Thus the gauge staffs indicate higher readings than the actual streamflow. It was also noted that the persons responsible for reading the river gauge staffs lived at a distance from the discharge stations, which had an implication of unreliable readings, as most of the time they would avoid the actual reading. This also contributed to the missing discharge data in all of the stations. Some farms also discharged their effluents to the stream, which contributed to the obtained streamflow yet it was not as a result of rainfall.

The discharge dataset was not continuous with the original data having more than 75% gaps and had been interpolated by Meins, (2013). However, data interpolation always introduces uncertainties, and in this case, the uncertainties introduced by more than 75% of the data were expected to contribute significant errors in to the data. Table 4 shows the gap analysis of the original streamflow over the period 2001 to 2010.

Table 3: Original streamflow gap analysis for the period 2001 to 2010 for the selected gauge stations

Sub-catchment	Total days	Total gaps	% gaps
Malewa	3651	2935	0.80
Gilgil	3651	3012	0.82
Karati	3651	2768	0.76

To gain more insight on the uncertainties in the observed discharge data, the interpolated streamflow was compared with the in-situ rainfall data. Since the in-situ rainfall data had less gaps, it was expected that the streamflow dynamics of wet and dry seasons would match those of the rainfall. The outstanding mismatches between the rainfall and the streamflow were treated as suspicious events as shown in the following figures:

Figure 4 below shows the plots of rainfall and discharge for the Gilgil River, for visual inspection of suspicious events.

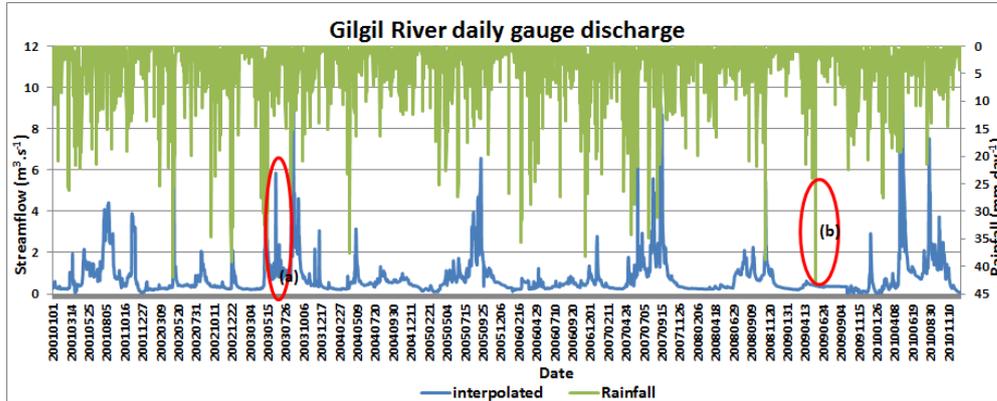


Figure 4: Figure showing some inconsistencies between the gauge rainfall and discharge data for Gilgil for the period 2001 and 2006

From figure 4 above, typical examples of suspicious events are demarcated. The demarcated suspicious events are expected to introduce discontinuities in the general trend of the streamflow mass curve. See figure 5 below.

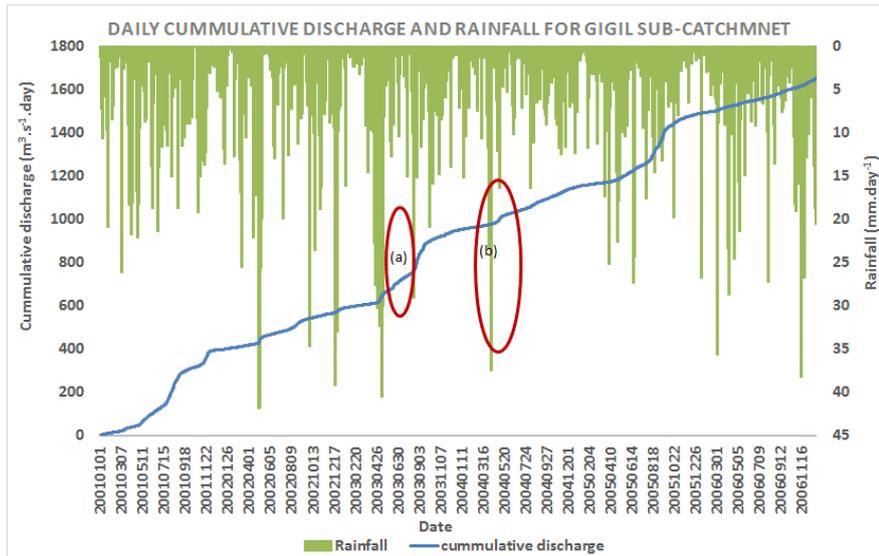


Figure 5: Figure showing the impacts of the inconsistencies between the ground rainfall and discharge to the flow cumulative curve for Gilgil sub-catchment

From the figures 4 and 5 above, the demarcated events are as follows:

- (a) High flow peaks without observable contributing rainfall
- (b) Large rainfall events without significant response to the observed flow,

Similar suspicious events were observed in the Malewa and Karati discharge and rainfall data as shown in figures 6 and 7 below:

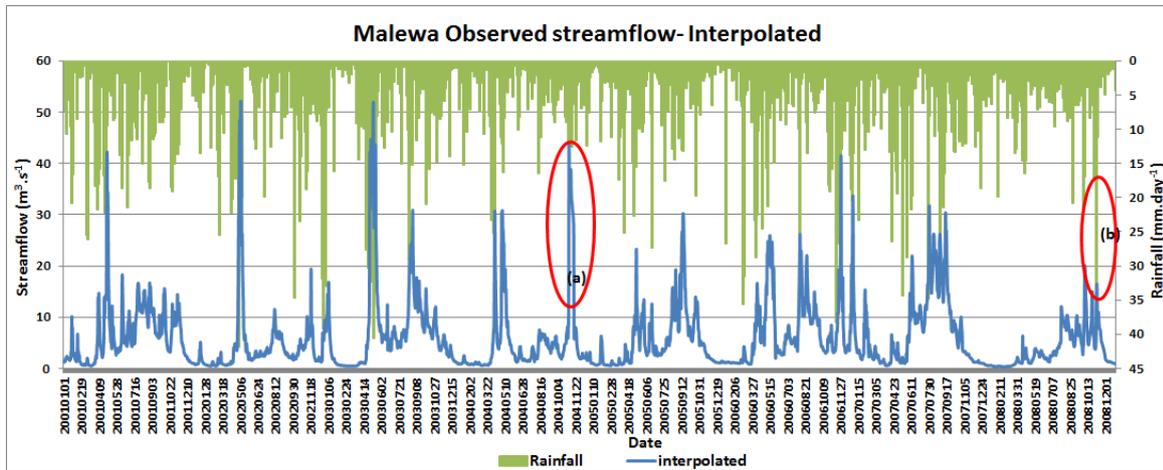


Figure 6: Plots for in-situ rainfall and discharge data for Malewa River, with demarcations showing suspicious events

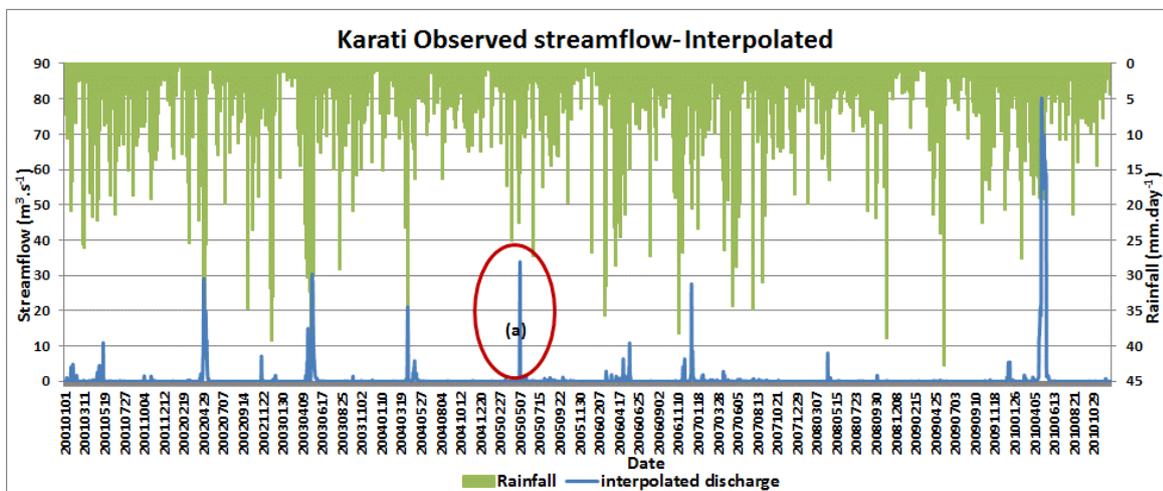


Figure 7: Plots for in-situ rainfall and discharge data for Karati River, with demarcations showing suspicious events

Data quality assessment for the Malewa and Gilgil streamflow as compared to the rainfall showed significant inconsistencies indicating that there were significant errors with the streamflow after it was interpolated to fill the missing gaps. There were notable events of high peaks in the streamflow without coinciding causative rainfall events, with the implications that such high peaks are difficult to reproduce with a model that depends on the rainfall input as the driver. Also, there were notable events of high rainfall without coinciding high peak discharges in the observed discharge, implying a model can produce some peaks in its simulation that does not coincide with peaks in the observed input discharge, thereby compromising its performance.

The Karati River, the streamflow lacked base flow, which can be explained from the fact that Karati is a seasonal river. Also, the abrupt peaks can be associated with discharge from some farms that were noted to be discharging their effluent upstream of the discharge station 2GD02. Also, the Gilgil River and the Malewa River are noted to have delayed quick flows, showing that the soil moisture contributes

significantly to the quick flow. This is also affirmed by their relatively low base flow, showing that most of the water forms quick flow. Such flow characteristics can also be caused by the terrain and the surface characteristics.

3.2.3. Monthly evaporation data

The monthly evaporation data for the lake was obtained from the WARMA-Naivasha offices. The monthly evaporation is derived from a pan evaporation station located near the lake. This data was used in calculating the loss of water by evaporation from the lake surface as a component of the lake level water balance.

3.2.4. Daily Lake levels

The daily lake levels data was obtained from the WARMA-Naivasha offices. This dataset includes the daily lake level measured using a tide gauge within the lake. This data was used in assessing the consistency between the actual observed lake levels trends and those calculated from the simulations and the in-situ data used in this study.

4. SATELLITE DATA

The satellite data used in the research include:

1. Rainfall data: TRMM3B42-v7 and CMORPH8km
2. Potential Evapotranspiration from Famine Early Warning Systems Network , here called (FEWSNET-PET)
3. HydroSHEDS topographical data including Shuttle Radar Topography Mission (SRTM) at 30 arc-second resolution Digital Elevation Model, Flow accumulation map, and Flow direction map

These satellite datasets were downloaded from their respective sources as shown in Table 4.

Table 4: Table showing the various satellite data used in this study as well as their sources

Data	Source
TRMM rainfall	http://mirador.gsfc.nasa.gov/
FEWSNET PET	http://earlywarning.usgs.gov/ftp2/bulkdailydata/global/pet/years
SRTM DEM	http://hydrosheds.cr.usgs.gov/dataavail.php
Flow Accumulation Map	http://hydrosheds.cr.usgs.gov/dataavail.php
Flow Direction Map	http://hydrosheds.cr.usgs.gov/dataavail.php
CMORPH 8km-1hr	ftp://ftp.cpc.ncep.noaa.gov/prcip/CMORPH_V1.0/CRT/8km-30min

4.1. FEWSNET PET

Famine Early Warnings Systems Network (FEWSNET) PET is a daily global potential evapotranspiration product that is calculated using climate parameter data obtained from Global Data Assimilation System (GDAS) analysis fields (USGS, 2012). Notably, the inputs to the PET obtained from GDAS include air temperature, wind speed, atmospheric pressure, solar radiation (long wave, short wave, outgoing and incoming), and relative humidity. The daily PET is calculated using Penman-Monteith equation on a spatial basis as given below:

$$E_p = \frac{\Delta}{\Delta + \gamma} \frac{R_n}{\lambda} + \frac{\gamma}{\Delta + \gamma} E_a \quad (\text{Eq 1})$$

Where: E_p is the daily potential evapotranspiration (mm day^{-1}), R_n is the net radiation to the evaporating surface ($\text{MJ m}^{-2} \text{day}^{-1}$), E_a is a function of the daily wind speed (ms^{-1}), average vapour pressure (kPa), and saturation vapour pressure (kPa), Δ is the slope of the saturation vapour pressure temperature relationship ($\text{kPa } ^\circ\text{C}^{-1}$), γ is the psychrometric constant ($\text{kPa } ^\circ\text{C}^{-1}$), λ is the latent heat of vaporization (MJ kg^{-1}) and E_a is the aerodynamic component.

Since the GDAS data are generated at 6 hours intervals by the National Oceanic and Atmospheric Administration (NOAA), the PET so computed is aggregated to daily totals. The FEWSNET PET is at 100km spatial resolution.

The daily PET data from FEWSNET was downloaded using a batch process integrated in the ILWIS software (see appendix B for the batch process used in processing the FEWSNET PET). The dataset was then converted to ASCII format and put in the PET folder in the CREST model.

4.2. HydroSHEDS

HydroSHEDS is coined from Hydrological data and maps based on Shuttle Elevation Derivatives at multiple Scales. “HydroSHEDS is a mapping product providing hydrographic information for applications at regional and global-scale in a consistent format” (“USGS HydroSHEDS,” 2013). HydroSHEDS have been developed by the Conservation Science Program of World Wildlife Fund (WWF) and are available for free. For this study, the 90m HydroSHEDS products of conditioned digital elevation model (DEM) flow accumulation map (FAC), and drainage directions map (FDR) were used.

4.3. TRMM3B42-V7

Tropical Rainfall Measuring Mission (TRMM) is a joint project between NASA and Japanese Space Agency (JAXA), launched in November 1997. TRMM generates various products that are released freely through various websites. The TRMM3B42-v7 is one of the various products of TRMM Multiple-Satellite Precipitation Analysis (TMPA). The TMPA combines various precipitation datasets from different satellite sensors as well as raingauge data from Global Precipitation Climatology Centre (GPCC) (Prakash, Mahesh, & Gairola, 2013). TRMM3B42-v7 is the latest version and is available in daily basis. The spatial resolution for TRMM3B42-v7 is 0.25° (approximately 25km). This study used the daily TRMM3B42-v7.

The daily TRMM3B42-v7 data for the period 2001 to 2010 was downloaded from Goddard Earth Sciences (GES) data through the Mirador search tool. Mirador is an earth science data search tool developed at the GES DISC for the users of the data from their portal. This data was in netcdf format and a code was developed in python to convert it to ASCII format, as required by the CREST model. The code is as shown in appendix C.

4.4. CMORPH-8KM

NOAA CPC Morphing Technique (“CMORPH”) is a process of producing global precipitation analysis at high spatial and temporal resolution (Climate Prediction Center Internet Team, 2013). Precipitation estimates exclusively derived from low earth orbit (LEO) satellite microwave observations have their features transported by spatial propagation information obtained from IR data of geostationary satellite (Joyce, Janowiak, Arkin, Xie, & Cmorph, 2004). The current algorithm for the CMORPH 8km-30min product incorporates precipitation estimates obtained from the passive microwaves aboard the DMSP 13, 14 & 15 (SSM/I), the NOAA-15, 16, 17 & 18 (AMSU-B), and AMSR-E and TMI aboard NASA’s Aqua and TRMM spacecraft respectively (NOAA, 2014). CMORPH 8km-30min is produced with a grid resolution of 0.07277 degrees (0.078km at the equator). The temporal resolution is 30 minutes, in a global domain and is freely available from various websites.

A procedure is developed and incorporated in the ILWIS software to download the CMORPH8km-30min product from ftp site, unzip it and process it to hourly values. The ILWIS procedure was used to download the data, but was customized to process only one half of the globe so as to save time. (See appendix D for the customized ILWIS function used to download CMORPH8km-30min from the ftp site in BSQ format and convert it to hourly in ILWIS format).

The hourly dataset was then aggregated to daily values by a batch process using the ILWIS software as well as other preprocessing procedures as follows:

Aggregation of hourly CMORPH 8km to daily

The hourly CMORPH 8km was aggregated to daily by a batch process in ILWIS. The map list utility in ILWIS was used in this process. This entailed developing a template to group the days 24 hourly images to

a map-list, and using the **map list function of sum**, the daily values per pixel were obtained in daily images in mm/day. (See appendix E for the batch process used to process the hourly CMORPH8km to daily)

Sub mapping for Naivasha basin

The daily CMORPH 8km data was sub mapped for the area of interest in ILWIS through the sub map spatial processing utility. Finally the daily sub map images were converted to ASCII format using the export function in ILWIS.

Figure 8 shows the processing of CMORPH8km rainfall.

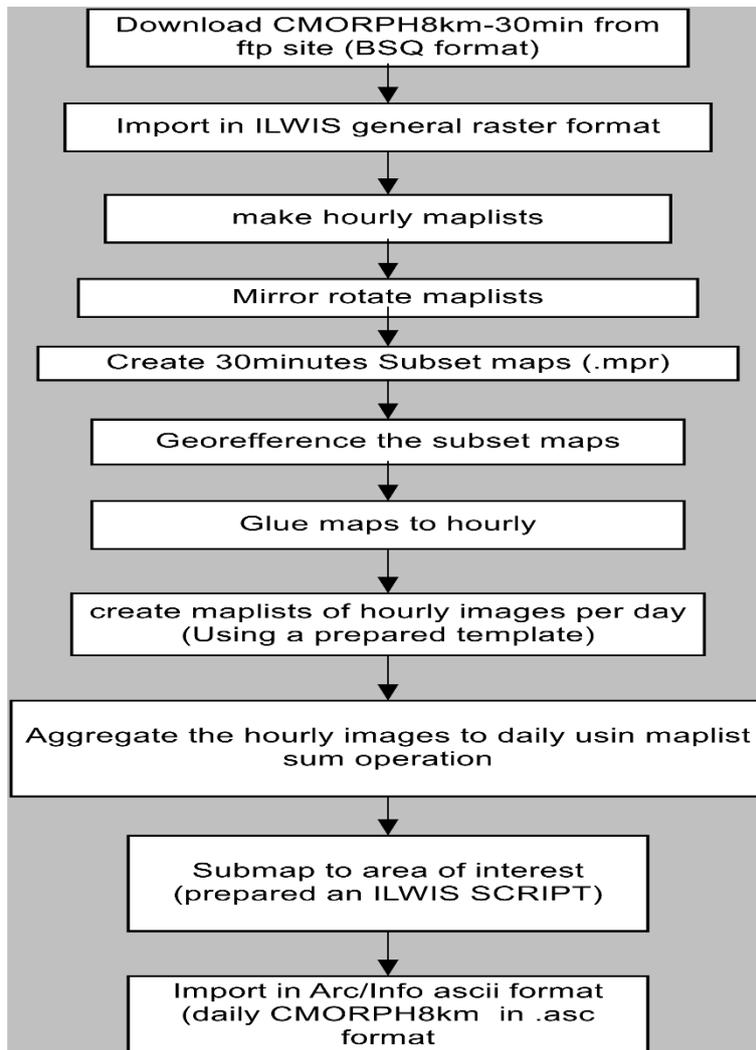


Figure 8: Flow chart for the CMORPH8km data processing

5. CREST MODEL

5.1. Model structure

The hydrological modelling software used in this study is the CREST distributed hydrologic model version 2.0. CREST model was developed recently by the University of Oklahoma and NASA SERVIR Project Team (Wang et al., 2011). CREST model is coded in FORTRAN and is open source software. The special feature of the CREST model is that it uses a variable infiltration curve (VIC) to compute the components of runoff including surface runoff, infiltration and interflow runoff (Wang et al., 2011b; Xue et al., 2013). Furthermore, the surface and the subsurface flow process are linked in the CREST model by coupling the runoff generation processes and cell-to-cell routing scheme (Xue et al., 2013). CREST is applicable to small and medium size basins at high spatial resolutions and as such, it has found wide application (e.g. Khan, Sadiq et al., 2011; Wang et al., 2011b; Xue et al., 2013). The distribution aspect of CREST V2.0 is achieved through grid cells whose resolution is user-defined. In this way, we can represent the spatial variability of the hydrologic processes of soil moisture, and runoff generation for each of the sub-catchments in the Naivasha basin. The CREST V2.0 model used has 15 parameters as shown in table 5 below. The parameters were divided into initial conditions, physical, and conceptual parameters.

Table 5: Table showing the parameters in the CREST V2.0 model, and their respective categories

CREST v2.0 PARAMETERS		
MODULE	SYMBOL	DESCRIPTION
Initial condition	W0	Initial value of soil moisture
	SS0	Initial value of Overland Reservoir
	SI0	Initial value of Interflow Reservoir
Physical Parameters	Ksat	The soil saturated hydraulic conductivity
	RainFact	The multiplier on the precipitation field
	WM	The Mean Water Capacity
	B	the exponential of the variable infiltration curve
	IM	Impervious area Ratio
	KE	The factor to convert the PET to local actual
	coeM	Overland runoff velocity coefficient
Conceptual Parameters	expM	Overland flow speed exponent
	coeR	Multiplier used to convert overland flow speed to channel flow speed
	coeS	Multiplier used to convert overland flow speed to interflow flow speed
	KS	Overland reservoir Discharge Parameter
	KI	Interflow Reservoir Discharge Parameter

Figure 9 illustrates the CREST model physics including the runoff production as well as the consequent routing of the flow. The atmospheric forcing defines the top boundary condition and it includes the input rainfall and potential evapotranspiration. There is also no connection between the ground water and the soil moisture storage and processes thus defining a no flow boundary at the bottom of the model. The hydrological divide is defined by the extent of the digital elevation model both for the interflow and the surface flow runoff. Therefore, any cell can only contribute its flow into a downstream cell defined by the DEM.

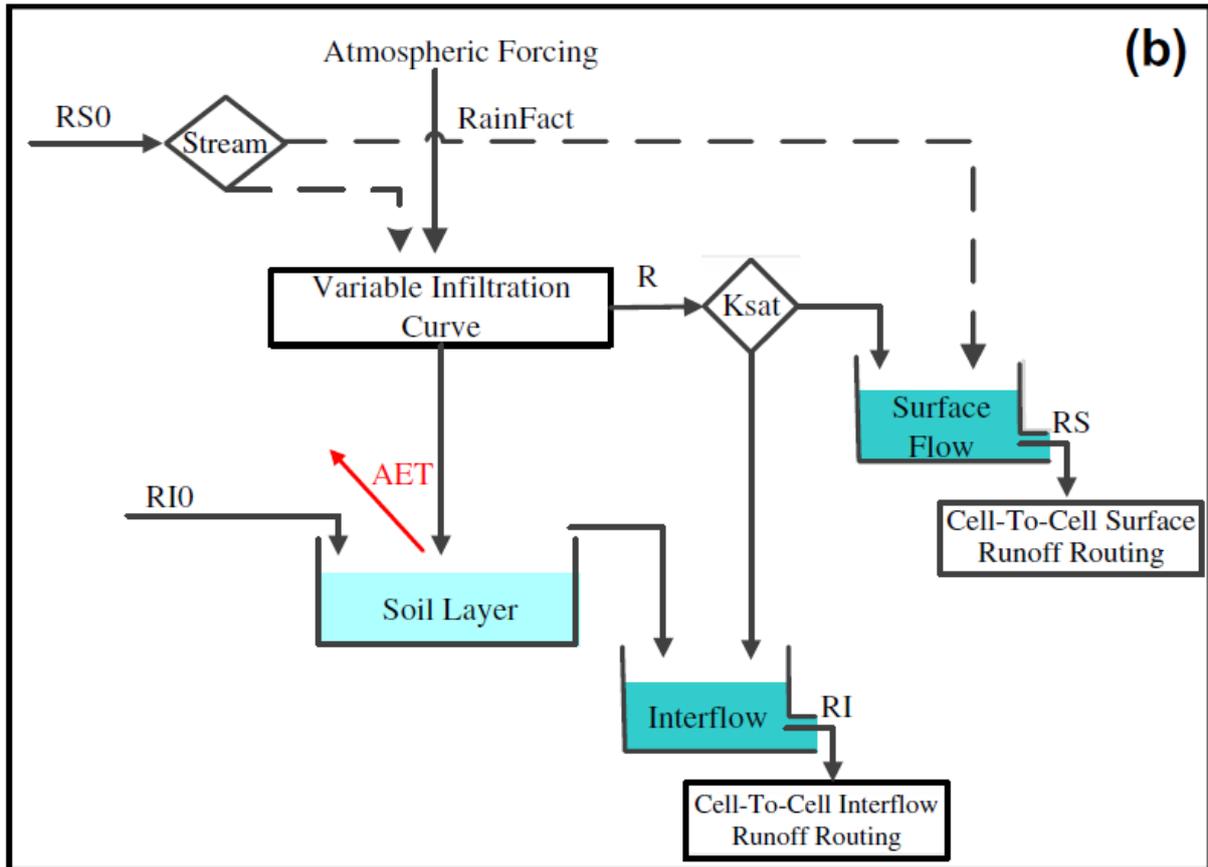


Figure 9: Schematic Representation of the CREST model v2.0 showing two vertical reservoirs, the variable infiltration curve and some of the processes involved in the model.

5.1.1. Runoff generation

The rainfall-runoff generation process is activated by precipitation P falling on the surface of a grid cell.

CREST model is forced by precipitation and potential evapotranspiration that can be specified for each grid cell and serve as the top boundary condition for solving the surface water budget at the level of each cell given. The water budget is solved for each time step by

$$\frac{W}{dt} = P - E_a + \sum R_{O,in} - R_{O,out} + \sum R_{I,in} - R_{I,out} \quad (\text{Eq 2})$$

Where W is the total cell water storage, including all the water stored in the canopy, the soil layer, overland and the interflow reservoirs. P is the input precipitation falling on the cell and is given by

$$P = \text{RainFact} \times \text{Rain} \quad (\text{Eq 3})$$

Where RainFact is a multiplier in the rain field

E_a is the actual evapotranspiration, which is derived from the potential evapotranspiration (E_{pot}) and the vertically-integrated water contents of the cell. E_{pot} is the local potential evapotranspiration obtained from the modification of the input potential evapotranspiration by the parameter KE .

$R_{O,in}$ and $R_{O,out}$ are the surface runoff into the soil surface of the cell from the upstream cell and out of the soils surface to the downstream cell respectively. On the other hand, $R_{I,in}$ and $R_{I,out}$ are the interflow runoff into the soil of the cell from the upstream cell and out of the soil layer of the cell to the downstream cell respectively.

Therefore the precipitation (P) and potential evapotranspiration (E_{pot}) triggers the rainfall-runoff generation process in two ways:

$P > E_{pot}$ (some precipitation reach the soil surface)

$P \leq E_{pot}$ (No precipitation reaches the soil surface)

Case 1: ($P > E_{pot}$), thus $E_a = E_{pot}$ (Eq 4)

The fraction of the precipitation reaching the soil surface (P_{soil}) is calculated as

$$P_{soil} = (P - E_{pot}) \times (1.0 - IM)$$
 (Eq 5)

Where IM is the impervious area ratio

When initial soil moisture ($W0$) is more than the soil maximum capacity (WM) then the soil is saturated at the beginning of the time step) and all the P_{soil} goes to excess rain (R). The new soil moisture (W) is thus equal to the initial soil moisture ($W0$) which is equal to WM . However, if $W0 < WM$, then WMM (an internal state variable representing the available capacity for water in the cell) is first determined as

$$WMM = WM \times (1.0 + B)$$
 (Eq 6)

Where B is the exponent of the variable infiltration curve (VIC)

Another internal variable A (representing the fractional area at which infiltration acts) is calculated as

$$A = WMM \times [1.0 - (1.0 - W0/WM)^{(1/(1+B))}]$$
 (Eq 7)

The incoming moisture is compared to the available capacity so as to determine the excess rain. Thus, if $P_{soil} + A \geq WMM$, then $R = P_{soil} - (WM - W0)$ and $W = WM$ (Eq 8)

For $P_{soil} + A < WMM$, $R = P_{soil} - WM \times [(1 - A/WMM)^{(1+B)} - (1 - \frac{A+P_{soil}}{WMM})^{(1+B)}]$ (Eq 9)

In that case, the excess rain can only occur if the calculated R is greater than zero. Moreover, the new soil moisture becomes:

$$W = W0 + P_{soil} - R$$
 (Eq 10)

In order to calculate the amount of water available for infiltration, a variable ($temX$) is defined that represent the maximum possible soil infiltration.

$$temX = [\frac{W0+W}{2}] \times K_{sat}/WM$$
 (Eq 11)

Where K_{sat} is the hydraulic conductivity

For ($R \leq temX$), $ExcI = R$ (Eq 12)

Where $ExcI$ is interflow excess rain in mm/hr

For ($R > temX$), $ExcI = temX$ (Eq 13)

In that case, $ExcS = R - ExcI + (P - E_{pot}) \times IM$ (Eq 14)

Where $ExcS$ is surface excess rain in mm/hr and IM is the ratio of the impervious area.

For $ExcS < 0$ then $ExcS = 0$ (Eq 15)

Case 2: $P \leq E_{pot}$, and $E_a = W0 - W$ (Eq 16)

In this case, $P \leq E_{pot}$ therefore, there is no excess rainfall ($P_{soil}=0$). Thus

$ExcS = 0$ (Eq 17)

$ExcI = 0$ (Eq 18)

and $temX$, an internal variable that in this case is representing the maximum possible loss of soil moisture to ET, is calculated as,

$$temX = (E_{pot} - P) \times W0/WM$$
 (Eq 19)

For ($temX < W0$), $W = W0 - temX$ (Eq 20)

For ($temX \geq W0$), $W = 0$ (Eq 21)

5.1.2. Runoff routing

The flow speed is calculated for every time step, with the current surface speed being the surface speed ($V(j,i)$) calculated by:

$$V(j,i) = coeM(j,i) \times V_s \times slope(j,i)^{expM(j,i)} \quad (Eq 22)$$

Where $V(j,i)$ is the flow in cell (j,i) in m/s, and $coeM$ is the overland's runoff velocity coefficient and conceptualizes the land surface roughness or hydraulic conductivity. Slope is obtained from the DEM as the differences in height between the cell (j,i) and the next cell, divided by the length between them.

A stream is formed when the number of cells draining into cell (j,i) is greater than a given threshold (TH) which represents the number of cells draining into cell (j,i). This TH is defined by the user in the file stream.def of the CREST model. Thus,

if($FAC(j,i) > TH(j,i)$), *then Stream is true*

For streamflow

$$V_R = V(j,i) \times coeR(j,i) \quad (Eq 23)$$

Where V_R is river flow velocity in m/sec and $coeR$ is the multiplier to convert the overland flow speed to channel flow speed.

For interflow

$$V_s = V(j,i) \times coeS(j,i) \quad (Eq 24)$$

Where V_s is speed of horizontal flow through the soil in m/s and $coeS$ is a multiplier to convert the overland flow speed to interflow speed.

The channel flow adds water to the initial surface storage ($SS0$) while interflow adds water to the initial interflow storage (SIO).

The surface storage per time step is calculated as:

$$SS0(t1) = [SS0(t0) + ExcS](1 - KS) + \sum R_{o,in} \quad (Eq 25)$$

On the other hand, the interflow storage per time step is calculated as

$$SIO(t1) = [SSI(t0) + Excl](1 - KI) + \sum R_{i,in} \quad (Eq 26)$$

Where $t1$ and $t0$ are the current and initial time steps respectively, KS and KI are the surface and interflow storages discharge parameters respectively

Surface discharge is determined as

$$RS = KS \times (SS0(t0) + ExcS) \quad (Eq 27)$$

And interflow discharge is calculated as

$$RI = KI \times (SIO(t0) + Excl) \quad (Eq 28)$$

The overall discharge is obtained as

$$Runoff = \left[\frac{(RS + RI)}{t} \right] \times GridArea/3.6 \quad (Eq 29)$$

5.2. Implementation

The data required by the CREST v2.0 model include topographical data from the HydroSHEDS (<http://hydrosheds.cr.usgs.gov/index.php>), data generated by the CREST model itself, and forcing data. The preparation of the data and other files for the CREST model V2.0 is as discussed in the following sections:

1. HydroSHEDS data

HydroSHEDS data include the topographical data obtained from the HydroSHEDS website including:

- SRTM 90m digital elevation model (DEM)
- Flow direction map (FDR) and
- Flow accumulation map (FAC).

All these topographic data were downloaded from the HydroSHEDS (a mapping product providing hydrological information data for global and regional applications in a consistent format). They were processed for the study area and converted to ASCII format before inputting to a folder “BASICS” of the CREST model.

The DEM in CREST contains a digital elevation model of the basin area with heights in meters. The DEM in this study was prepared in ASCII format using ARCGIS software. The DEM was georeferenced in WGS84. The other topographical data was the FDR file which contains the flow direction from each cell to its steepest downslope neighbor in the basin area. The final HydroSHEDS topographical data was the FAC file which contains the accumulation flow to each cell within the basin. Both the FDR file and the FAC file have the same coordinate system as the DEM file.

2. Files Generated by the CREST V2.0 model

This category of data includes the files that are generated by the first run of the CREST model and were placed in the Basic folder together with the topographical data. Although these files are optional, their inclusion in the Basics folder improves the speed of the subsequent model runs. The Mask file contains a mask of the basin, and it indicates which cells are inside the basin from the area of interest. The mask file has the same coordinate system as the DEM file. The slope file identifies the rate of maximum change in DEM file from each cell. If the slope file is missing, slope.def can be provided, which contains the threshold for calculating the slope by CREST v2.0. However, if both the Slope file and slope.def are missing, CREST V2.0 will calculate the slope automatically. The Stream.def file contains the threshold to determine which cells are streams and CREST v2.0 generates the stream file based on it. The stream definition file was prepared with a threshold of 2 and inputted to the Basics folder.

3. GridArea file:

This file contains the area of each cell in the basin. It is user defined and takes the same coordinate system as the DEM. The GridArea file was prepared, with the same dimensions as the DEM and inputted in the BASICS file. The grid spacing was done at $1\text{km} \times 1\text{km}$.

4. Forcing Data

The forcing data for the CREST model include both gridded rainfall and potential evapotranspiration. In this study, the forcing data used are different products developed from satellite observation. Two rainfall products were used i.e. the TRMM 3B42-v7 and the CMORPH8km-30 minutes. The daily potential Evapotranspiration was obtained from FEWSNET. All the forcing data was downloaded and processed for the study area in the ASCII format and put in their appropriate directories in the CREST model.

5. Other files

Other files prepared for the CREST model were the ICS, OBS, parameters, and states file. The ICS (initial conditions) file contains the initial conditions of the storages in the model system. In the CREST model v2-0 used in this study, the initial conditions included the initial values of the soil water (WU0), surface storage (SS0) and interflow storage (SS0). These values were estimated from crest models pre-runs and were later calibrated to optimize them. The initial conditions file was placed in the ICS directory. The other file that was prepared was the observations file that contains the observed discharges in m³/s. The file has two columns, the date and the discharges columns, with the date's column depicted by the model run options. In this study, the date column was in days as all the model runs were done on daily basis. This file was in the format:

Station Name+_Obs.csv

The Observations file was placed in the Observation folder (OBS). The other file that was prepared is the parameters file, containing the parameters for the CREST model and was placed in the Parameters folder.

6. Control file

The control file (also project file) takes the form; "ProjectName.Project". The control file has different sections as follows: The CREST's project file contains the following:

- Model Area section, which shows the area information as obtained from the digital elevation model
- Model runtime section, which dictates if the model is to run in simulations, calibration or real-time modes
- Model input and output directories as well as their format, which directs the model on where to get the input data and where to store the output data
- OutPix information, which states the information about the output pixel
- Outlet information
- Gridded states variables
- Specified date and time for the model runs as well as for the model output (see the CREST v2.0 manual (Xue & Hong, 2013), for information about the CREST V2.0 model project file.

The Project file was prepared for each of the three subcatchments in the Lake Naivasha basin. Likewise, apart from the topography data and the model forcing data, all the other data and files were prepared for each of the three subcatchments in the Naivasha basin.

Figure 10 shows a schematic representation of the organization of the CREST v2.0 files

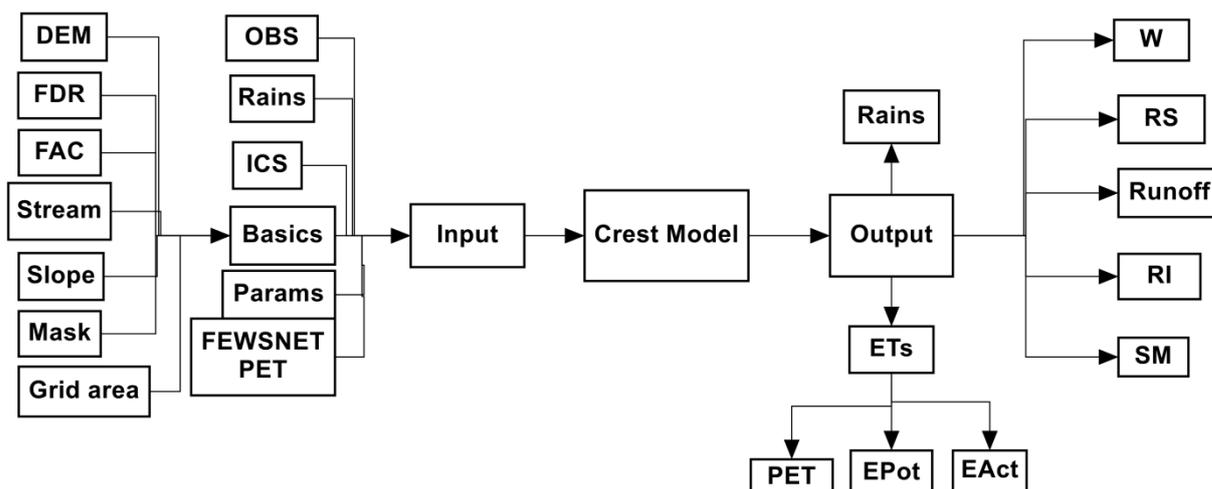


Figure 10: Schematic representation of the organization of files for the CREST V2.0 model as used in this study

5.3. Coupling PEST to CREST

PEST optimization procedure was used for calibrating the CREST model in this study. Therefore, the three different files required when using PEST for calibration, were prepared. These files are:

1. Template file: instructs PEST where to write the optimized model in the model input file.
2. Instruction file: instruct PEST from which position within the model output file the output should be read
3. Control file: defines the complete PEST setup, and is thus the most important file

These files were prepared for each of the three subcatchments in the Naivasha basin and put in their respective folders. The PEST procedure runs from the “outside” of the model as shown in figure 11 below.

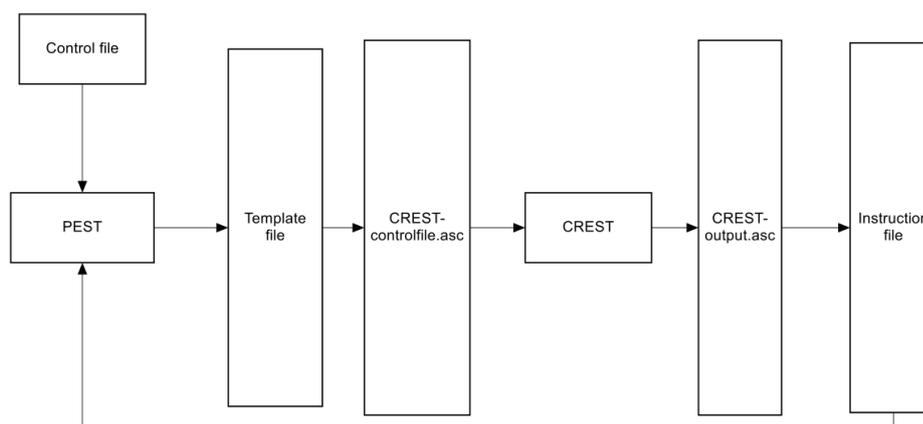


Figure 11: Schematic representation of the PEST calibration procedure showing how PEST encapsulates the CREST model

A multiple restarting procedure for the CREST model was designed. The multiple restarting was done for the CREST model in the simulation running mode. For each of the runs, the PEST procedure runs the GML algorithm changes the parameter sets for the next model run in the efforts of optimizing the objective function. The Multi-restarting procedures continues until a convergence point is attained, which in this case is a local optimum point. For this reason, the calibration process was done in a way that ensured most of the parameter space was utilized so as to increase the chances of obtaining the near global optimum point. First, for each of the model set-up, the physical parameters were optimized, followed by the conceptual parameters, then the physical parameters, and so on until a satisfactory value of the NSCE and bias was attained.

5.4. Calibration

The Lake Naivasha basin comprises of three rivers, Malewa, Gilgil, and Karati. Therefore, calibration and validation was done for each of the three river systems separately. The most downstream gauge stations with reliable streamflow data were used as the discharge point for each of the subcatchments. Notably, the CREST model masks the drainage area based on the pixels contributing to the flow at the targeted discharge point. Thus; the first runs for each of the subcatchments yielded the masks for each of the subcatchments as shown in figure 12.

The observed streamflow discharges were used as the ground truth and were compared to the resulting simulated flows from the model runs within the CREST model. The results of the model runs are written in a CSV file, and each model run overwrites the results of the preceding model run on this file. Several correlation test runs were done to gain insight on physical and modelling by observing differences between the simulations and the observations.

The objective functions obtained from the CREST model setting include the Nash-Sutcliffe coefficient of efficiency (NSCE), the relative bias, as well as the correlation coefficient (CC). NSCE helps in determining the simulation of the hydrograph's shape (in this case, the time to peak, and recession). The optimum value for the NSCE is 1. The relative bias is used for assessing the systematic bias of the runoff. The optimum value for the bias is 0, and the closer one gets to zero the better the simulation. The correlation coefficient (CC) is used to assess the agreement between simulated runoff and observed runoff. The optimum value for CC is 1.

$$NSCE = 1 - \frac{\sum_{i=1}^n (R_{obs,i} - R_{sim,i})^2}{\sum_{i=1}^n (R_{obs,i} - \overline{R_{obs}})^2} \quad (\text{Eq 30})$$

$$BIAS = \left[\frac{\sum_{i=1}^n R_{sim,i} - \sum_{i=1}^n R_{obs,i}}{\sum_{i=1}^n R_{obs,i}} \right] \times 100 \quad (\text{Eq 31})$$

$$CC = \frac{\sum_{i=1}^n (R_{obs,i} - \overline{R_{obs}})(R_{sim,i} - \overline{R_{sim}})}{\sqrt{\sum_{i=1}^n (R_{obs,i} - \overline{R_{obs}})^2 \sum_{i=1}^n (R_{sim,i} - \overline{R_{sim}})^2}} \quad (\text{Eq 32})$$

Where: n is number of observations R_{obs} is observation at time (i), and R_{sim} is the corresponding model prediction, $\overline{R_{obs}}$ and $\overline{R_{sim}}$ are the averages of the observations of simulations respectively.

The objective function for PEST is defined as $\phi(\vec{b})$ (which is the sum of squared deviations between simulated and observed values). This objective function is reduced to a minimum and it can be represented mathematically as:

$$\phi(\vec{b}) = \sum_{i=1}^n [W_i (R_{obs,i} - R_{sim,i})]^2 \quad (\text{Eq 33})$$

Where: ϕ , is the objective function, \vec{b} is the vector with fitting parameters, and W_i is weight associated with the measurement at the particular point.

PEST calibration procedure was done for the period 2002 to 2005 with the year 2001 as the warm-up period. After setting up the CREST model and the PEST calibration files, the model was run in the

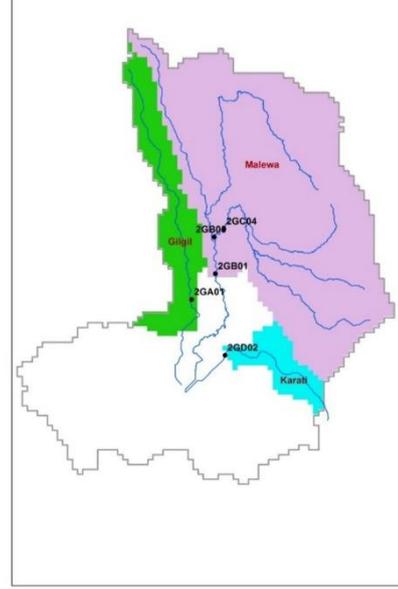


Figure 12: The Naivasha basin's sub-catchments as they were extracted using the CREST model based on the outlet points chosen depending on the most downstream available discharge data for each of the rivers in the basin

simulation mode. First, a sensitivity analysis was carried out to determine the parameters to target for the calibration.

5.4.1. Sensitivity analysis

The CREST model parameter sensitivity analysis was done using the setup for the Malewa River to gain an insight on the impact of each parameter to the model simulation. A local sensitivity analysis was done in that, the streamflow output responses were determined by sequentially varying each of the parameters and fixing all other parameters to nominal values. The nominal values selected for this study are the CREST model default parameter values. The NSCE was used as the objective function for the sensitivity analysis

The sensitivity results showed that the range of the various parameters had varying impact on the NSCE value. The order of magnitude by which the NSCE value changed due to the change of each parameter varied and as such, the parameters were grouped according to their sensitivity as is shown in the 4 plots in figure 13 below.

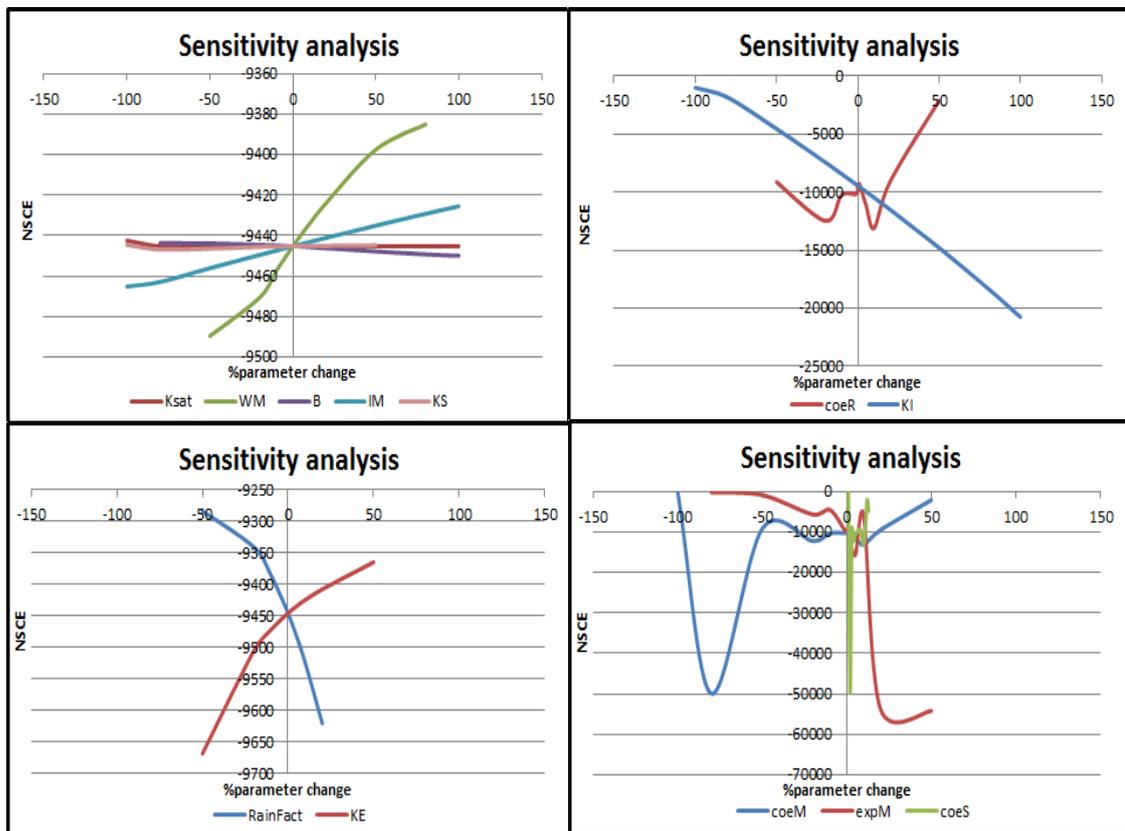


Figure 13: The CREST model parameter sensitivity analysis plots for Malewa sub-catchment

From the sensitivity analysis, the parameters were categorized according to their impacts on the simulation results as follows:

Category 1: Physical parameters with a low sensitivity in the parameter space. They included Ksat, WM, and IM. This category of parameters was assumed not to shift the optimum point and hence they required only one step of calibration.

Category 2: Physical parameters with a high sensitivity. Parameters in this category included B, RainFact, coeS, and KE.

Category 3: Routing parameters consisting of KI, KS, coeR, expM and coeM. These parameters were very sensitive and their alteration shift the optimum point to a different order of magnitude. These parameters were thus set to be optimized in a multiple approach so as to get as close as possible to the global optimum.

5.4.2. Sequence of calibration

The calibration was done following the criteria discussed below:

1. The first parameters to be optimized were the physical parameters including Ksat, WM, IM, RainFact and coeS. These parameters were found to have distinct optimum values for each model setup and therefore, they took the first priority for optimization after which they are defined. Also, actual values for these parameters can be obtained from fieldwork data.
2. The second parameters to be optimized were the second category of physical parameters including; B, KE, and coeM.
3. The third parameters to be optimized were the conceptual parameters including expM, coeR, KS, and KI.
4. The initial conditions were then optimized manually while considering their impact on the objective function, such that the adjustment of each of them stopped when the change on the NSCE was less than 0.001.
5. The process was then repeated again from step number 2, and then the step number 3 until a satisfactory accuracy was attained based on the NSCE value, the streamflow mass curve and hydrographs obtained.

Figure 14 shows a flow chart for the calibration strategy involved for the CREST model using the PEST calibration procedure for the Malewa, Gilgil, and Karati Subcatchments.

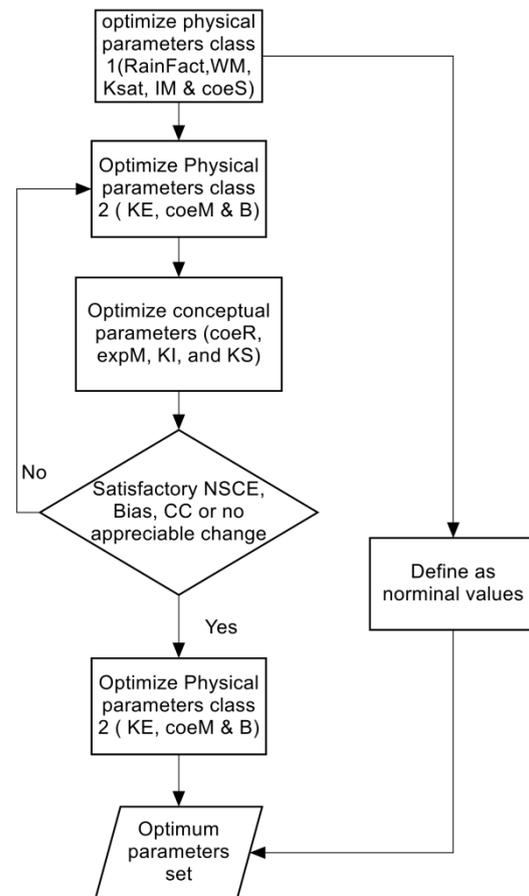


Figure 14: Flowchart for the steps followed in the calibration process for the 3 subcatchments in the Lake Naivasha basin

6. RAINFALL DATA ASSESSMENT

6.1. Daily rainfall assessment

The in-situ rainfall was assumed as the ground truth. As described earlier in section 3.2.1, five raingauge stations were selected as the continuity of their time series data for the study period was assumed to be satisfactory. The daily average values of these raingauge data were obtained and are thus referred as the daily gauge data in this study. The satellite representative data for the catchment was obtained by running the crest model with the input of the respective rainfall products and maintaining the RainFact parameter value as 1. Visual inspection was done to check on the agreement between the three rainfall products for the peaks and the lows as per the seasonal trends. (The three rainfall products are; gauge data, CMORPH8km and TRMM3B42-v7). All the rainfall products represented the seasonal dynamics of dry periods and rainfall periods.

Figure 15 shows the time series data trends of the daily gauge, TRMM3B42-V7 and CMORPH8km for 2001 and 2002 as this show the general seasonal dynamics. Also, the cumulative rainfall was obtained for the three rainfall products and their mass curves plotted to show the volumetric relationship between them. The mass curves were as shown in figure 16. Moreover, the scatterplots were made for the three rainfall products so as to gain an insight on the variations between them. Figure 17 show the scatter plots between gauge versus TRMM3B42-v7 and gauge versus CMORPH8km respectively.

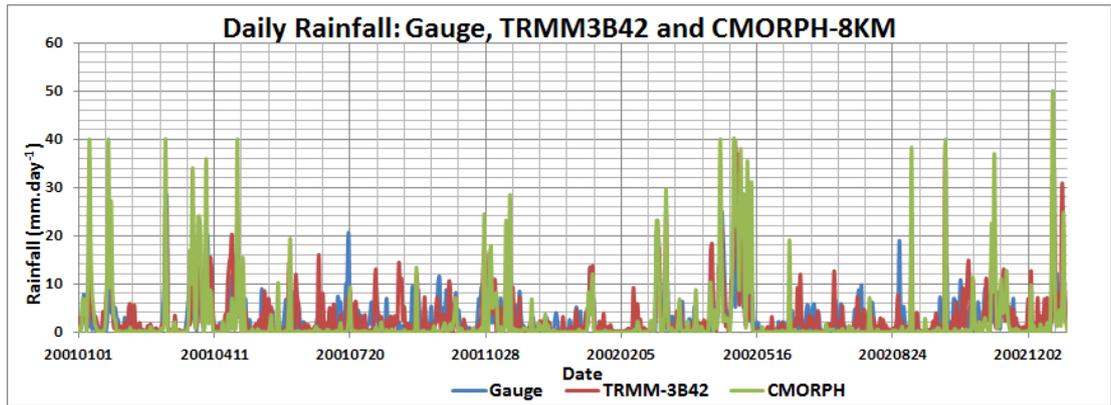


Figure 15: The time series data trends of the daily gauge, TRMM3B42-V7 and CMORPH8km for 2001 and 2002, over the Lake Naivasha Basin, showing the seasonal dynamics of rainfall

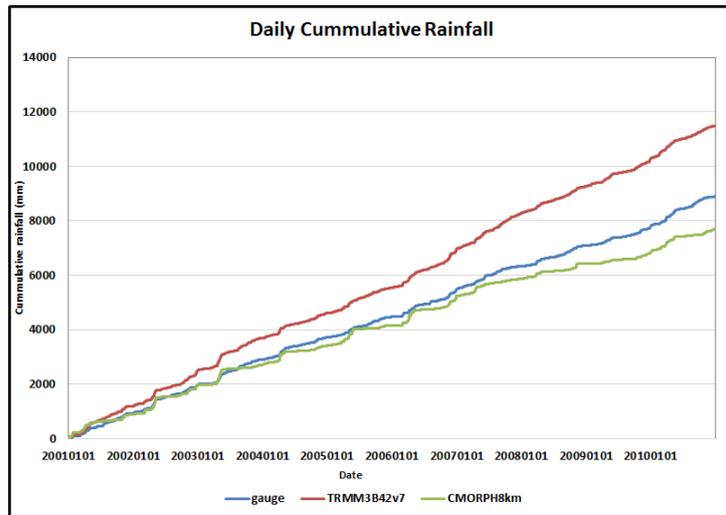


Figure 16: Cumulative rainfall plots for daily gauge, TRMM3B42V7 and CMORPH-8Km for the Lake Naivasha Basin from 2001 to 2010, showing volume relationship of the three rainfall products over the area

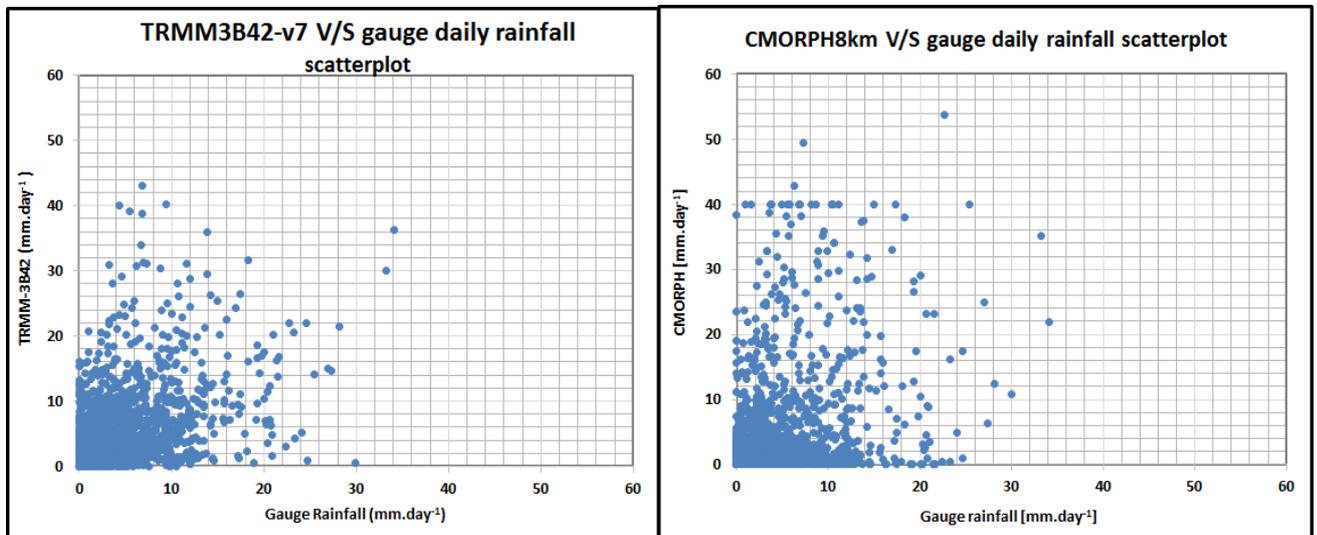


Figure 17: Scatter plots for daily; gauge versus TRMM3B42-v7 and gauge versus CMORPH8km over the Lake Naivasha Basin for 2001 to 2010

From the time series data trends of the daily gauge, TRMM3B42-V7 and CMORPH8km for 2001 and 2002, TRMM3B42V7 rainfall has more days with rainfall but the rainfall peaks are lower than those of the gauge rainfall in general. On the other hand, CMORPH-8km has fewer days with rainfall than the gauge rainfall, but the peaks are more exaggerated. From the cumulative rainfall plots, TRMM3B42-v7 produces the most rainfall, followed by gauge rainfall and CMORPH8km produces the least cumulative rainfall. Also, there is a closer volumetric coherence between the Cmorph8km and gauge rainfall. Furthermore, the volumetric difference between the three rainfall products increases with increasing time. The scatter plots for the daily values were very noisy, showing significant temporal and spatial variations between the satellite rainfall product and the gauge rainfall.

Various statistical analysis measures were done for the satellite rainfall products of TRMM3B42-v7 and CMORPH8km with the raingauge observations as the ground truth. The statistical analysis measures done in this case included: the bias, RMSE, and MAE.

$$\text{Bias} = \left[\frac{\sum_{i=1}^n P_{sat,t} - \sum_{i=1}^n P_{gauge,i}}{\sum_{i=1}^n P_{gauge,i}} \right] \quad (\text{Eq 34})$$

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (P_{gauge,i} - P_{sat,i})^2}{n}} \quad (\text{Eq 35})$$

$$\text{MAE} = \frac{\sum_{i=1}^n [\text{abs}(P_{gauge,i} - P_{sat,i})]}{n} \quad (\text{Eq 36})$$

Where n is the number of observations, P_{gauge} is gauge rainfall and P_{sat} is satellite rainfall

The results were as displayed in Table 6 below.

Statistical Measure	TRMM3B42-v7	CMORPH8km
Bias	-0.708	-0.334
RMSE(mm)	4.21	5.57
MAE (mm)	2.48	2.69

Table 6: Showing the statistical measures for the comparison of daily TRMM3B42-v7 and CMORPH8km rainfall products against the raingauge observations

The statistical measures show that TRMM3B42-v7 rainfall performs better in terms of RMSE and MAE than the CMORPH8km. On the other hand, the CMORPH8km rainfall performs better than the TRMM3B42-v7 in terms of bias.

6.2. Monthly Rainfall Assessment

The monthly average values of the raingauge data were obtained and are thus referred as the gauge monthly data in this study. Also, the monthly averages for the satellite rainfall were obtained. Visual inspection was done to check on the agreement between the three rainfall products for the peaks and the lows as per the seasonal trends. All the rainfall products represented the seasonal dynamics of dry periods and rainfall periods.

Figure 18 shows the time series data trends of the monthly gauge, TRMM3B42-V7 and CMORPH8km for 2001 and 2010. Also, the cumulative rainfall was obtained for the three rainfall products and their mass curves plotted to show the volumetric relationship between them. The mass curves were as shown in figure 19. Moreover, the scatterplots were made for the three rainfall products so as to gain an insight on

the temporal and spatial variations between them. Figure 20 show the scatter plots between gauge versus TRMM3B42-v7 and gauge versus CMORPH8km respectively.

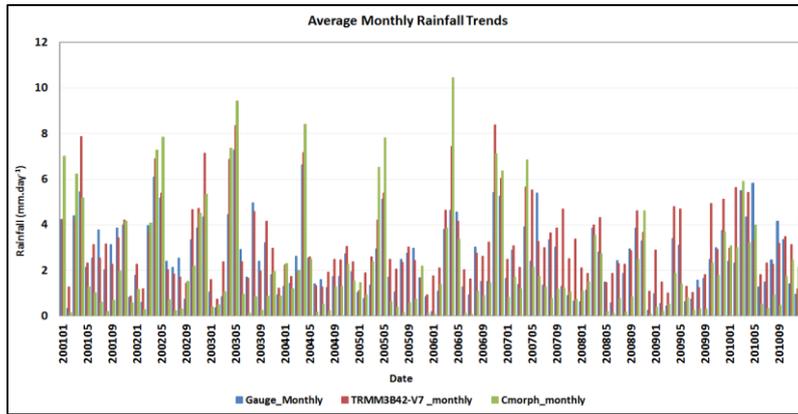


Figure 18: The time series data trends of the monthly gauge, TRMM3B42-V7 and CMORPH8km for 2001 and 2010

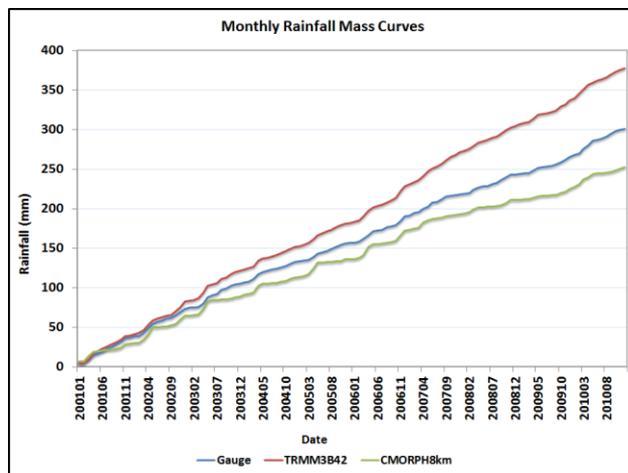


Figure 19: Cumulative rainfall plots for monthly rainfall of gauge, TRMM3B42V7 and CMORPH8Km

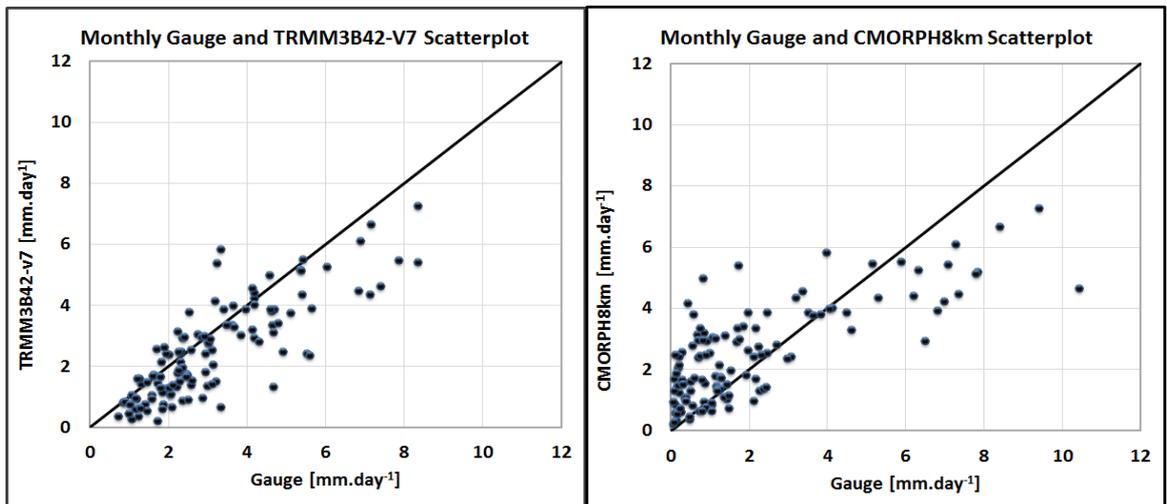


Figure 20: Scatter plots for monthly gauge versus TRMM3B42-v7 and gauge versus CMORPH8km

From the monthly average rainfall, the CMORPH-8km shows the highest peaks although the peak difference between the three different rainfall datasets reduces. There is therefore a better coherence between the monthly rainfall trends than in those of that of daily values. From the cumulative volumes there is a better volumetric relationship between the CMORPH-8km rainfall and the gauge rainfall than that of the TRMM3B42V7. Also, the relationship deteriorates with increase in time. The noise observed in the daily values scatterplots was reduced in the case of the monthly scatterplots as the time varying bias and errors were reduced.

The results of various statistical assessment of the monthly average rainfall values for the TRMM3B42-v7 and CMORPH 8km rainfall products in relation to the monthly average values of the raingauge observations were as shown in Table 7 below:

Table 7: Table showing the relationship between monthly average rainfall values for the TRMM3B42-v7 and CMORPH 8km rainfall products with the monthly average values of the raingauge observations.

Statistical measure	TRMM3B42-v7	CMORPH8km
Bias (%)	-0.6355	-0.4026
RMSE(mm)	1.77	3.12
MAE (mm)	0.883	2.1

From the statistical measures of accuracy, it is evident that the correlation between the two satellite rainfall products improves for the monthly values as compared to the daily values. Therefore, better representation of the rainfall would be expected when monthly values of TRMM3B42V7 and CMORPH-8km are used.

7. RESULTS AND DISCUSSION

7.1. Calibration and validation

The CREST model was run and calibrated using the TRMM3B42-v7 as well as the CMORPH 8km product. Since the streamflow model for the lake Naivasha in this study was important for water resources management in the Naivasha basin, it is important to relate the simulated streamflow hydrographs as well as the streamflow mass curves with those of the observed discharge. The NSCE, CC and the bias statistical measures of accuracy were also included in the performance analysis of calibration and validation for the three subcatchments.

7.2. Model calibration and validation while using TRMM3B42-V7 rainfall

Calibration was done for each sub-catchment for the period January 2002 to December 2005 followed by validation from January 2006 to December 2010 using TRMM3B42-v7 rainfall data. The simulated streamflow hydrographs in the calibration process, as well as, in the validation process were plotted together with the corresponding observed streamflow, for all the catchments. Figure 21 shows the calibration and validation streamflow hydrographs for the Malewa, Gilgil and Karati subcatchments respectively. Figure 22 shows the calibration and validation streamflow mass curves for the Malewa, Gilgil and Karati subcatchments respectively.

Since calibration aims at improving the relationship between the simulated stream flow and the observed streamflow to be as close as possible with the observed streamflow assumed to be the ground truth or having its uncertainties assessed, visual inspection of the calibrations results was done. The streamflow hydrographs were used for visual inspection of the success in simulating the streamflow dynamics of peaks and troughs. Therefore, for the streamflow hydrograph, we took note of the general shape entailing the peaks, time to rise and fall of the peaks, size of the peaks and the base flow.

Apart from the Gilgil River's case, the time to rise and fall of the peaks were simulated well. Most of the peaks in the observed streamflow were reproduced in the simulated hydrograph except for those that did not coincide with a causative rainfall event. The mismatch between the rainfall events and the streamflow was noted to be prominent in the case of the Gilgil River, hence the outstanding mismatches in the peaks between the observed and the simulated streamflows. Also, the base flow was well simulated for the Malewa River. Note that there was no base flow for the Karati River. Although the simulated base flows in both the Malewa and Gilgil continues to increase with time, the Gilgil case was observed to rise at a more prominent rate than that of the Malewa. A defined soil moisture reservoir and no-flow boundary condition to the ground water can be ascribed to cause the increasing base flow with time.

It is however worth noting that the sizes of the peaks were not well reproduced in all the cases as the observed streamflows had higher peaks than the simulated streamflows in all the three cases. The differences between the observed and the simulated streamflow peaks can be attributed to errors from the rainfall inputs or the observed discharges. Also, the coherence between the simulated hydrograph trend and that of the observed deteriorates with increasing simulation time. This time variant deteriorating simulation trend can be attributed to the cumulative spatial and temporal inaccuracies as well as weakness in the CREST model in reproducing the base flow.

Still, the contributions of the suspicious events described in the discharge assessment section can be seen in both the calibration and validation simulated hydrographs. Two examples of these suspicious high peak flow events are in the observed streamflows in the month of September 2003 (for Malewa and Gilgil) and May 2010 (for all the three subcatchments). Clearly, these high peak flow events do not coincide with any rainfall, whether from gauge, TRMM3B42-v7 or the CMORPH8km. Therefore, these events can be attributed to errors in the observation data. The simulated hydrographs did not produce the peaks in such events.

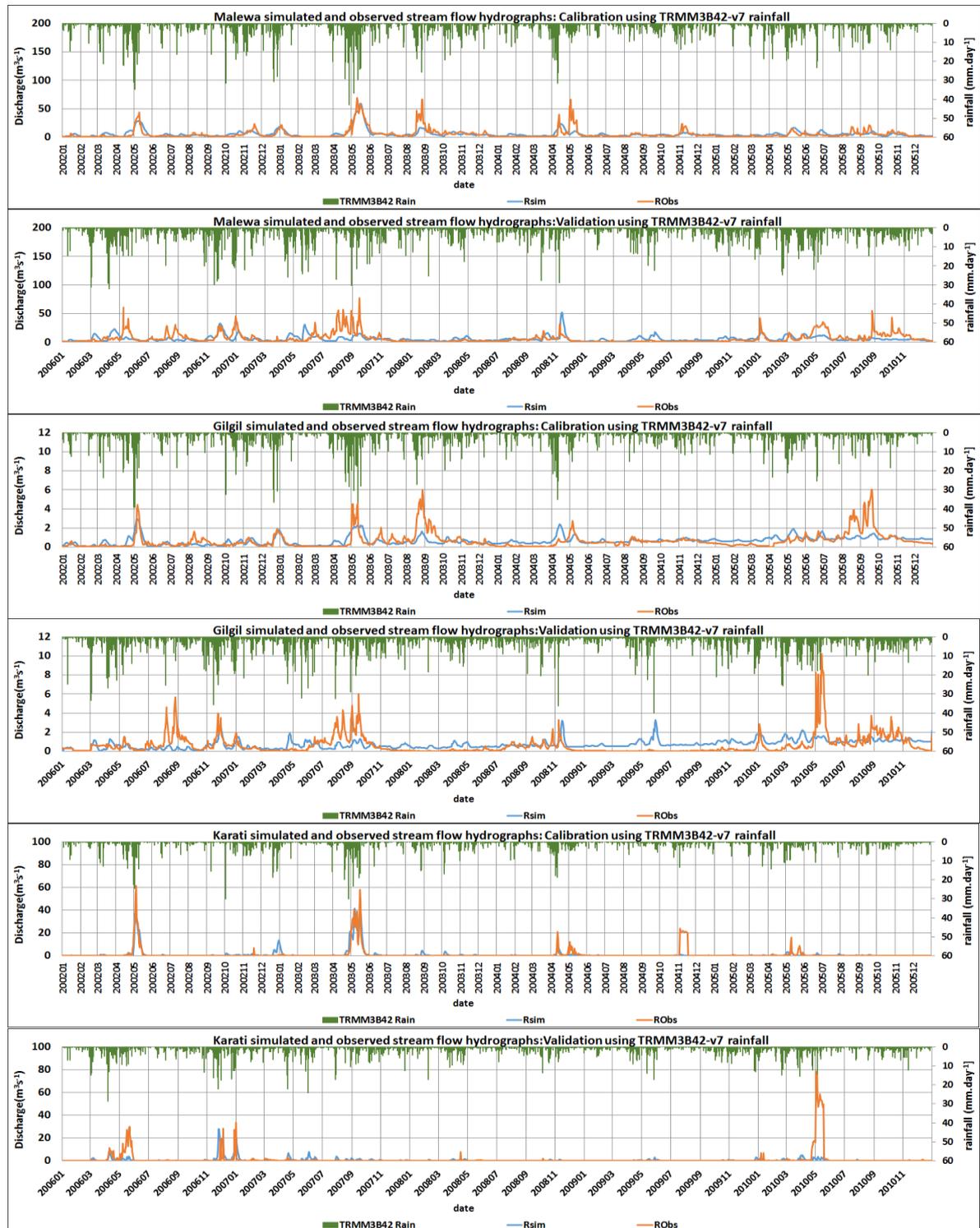


Figure 21: Streamflow hydrographs for calibration and validation of Malewa, Gilgil and Karati subcatchments using TRMM3B42-v7 rainfall

Similar observations were made for the validated streamflow hydrographs as those of the calibrated hydrographs. Since the validation process ascertains the calibration process, in this case therefore, the calibration can be said to have been successful as far as the streamflow hydrographs are concerned. It is however notable that the time variant mismatches between the observed and the simulated hydrographs were more prominent in the validation cases than in the calibration cases because of the longer validation period than calibration period.

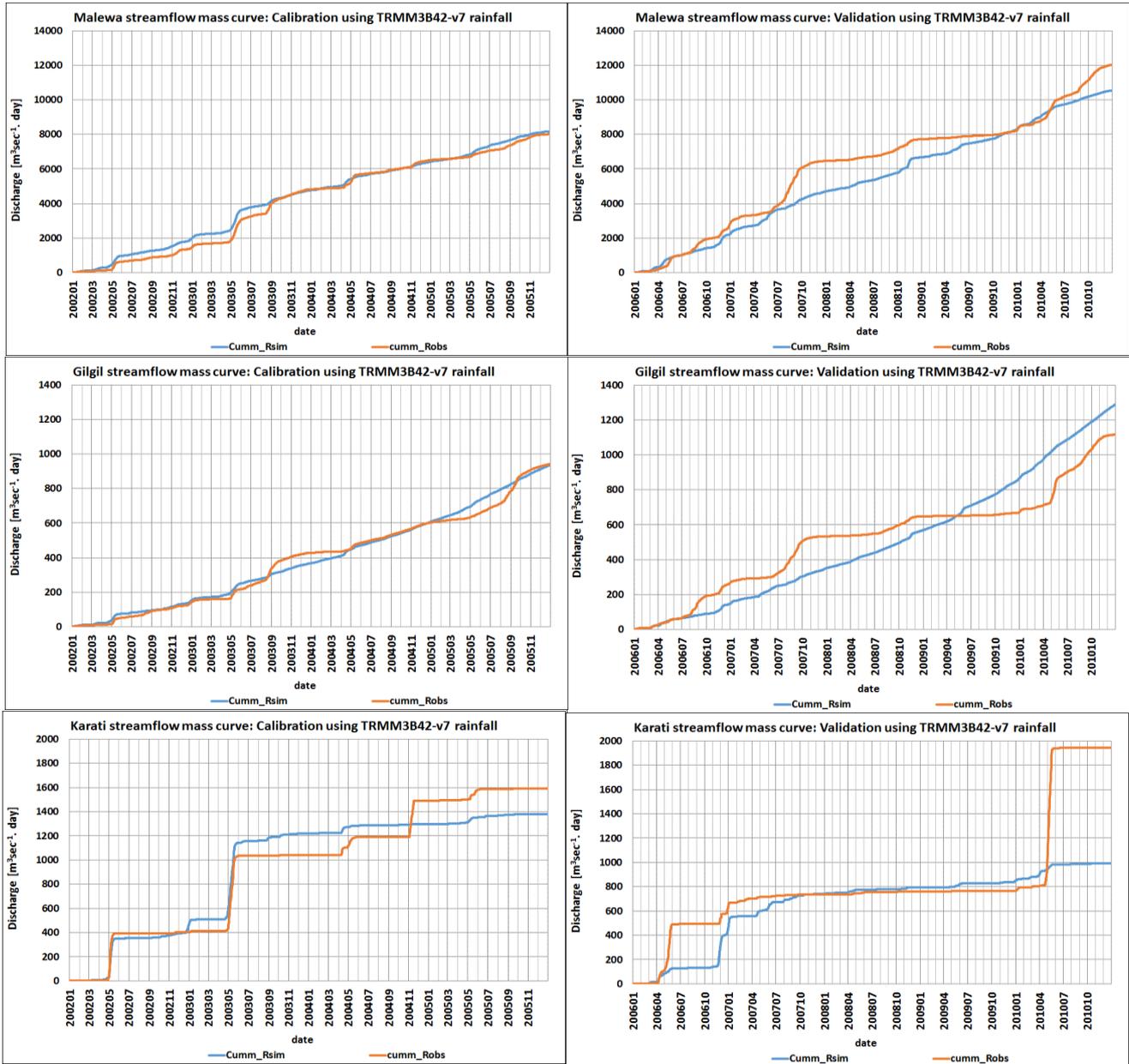


Figure 22: Streamflow mass curves for calibration and validation for Malewa, Gilgil and Karati subcatchments using TRMM3B42-v7 rainfall

The mass curves were also considered in the visual inspection of successful calibration and validation. The mass curves represent the cumulative volumes, which in-turn relates to the catchment yield. Considering that the reconstruction of the Naivasha basin streamflow is required for water management purposes, the

catchment yield is as such an important aspect. Therefore, there was need for the calibration to be successful in reproducing the observed streamflow mass curve. Through the mass curves, the impacts of unique as well as suspicious events in the observed streamflow as well as the gauge rainfall can be noted.

In all the cases, the initial parts of the mass curves were well simulated both for the calibration and the validation. Also, the simulation coherence deteriorated with time for all the cases although at varying extents. For the Malewa River, the observed streamflow mass curve was smoother for the calibration part than in the validation phase. This observation can be related to the streamflow hydrograph, whereby, the calibration phase has smoother trend than the validation phase. But this observation is also seen in the rainfall trend therefore it is as result of the rainfall input. The point to note here is that the discontinuities in the general trend of the observed streamflow in the validation phase was not well reproduced in the simulated validation mass curve, because the rainfall parameter had already been defined in the calibration process. Similar cases applied for both the Gilgil and Karati streamflow mass curves. With the exception of the discontinuities introduced by the rainfall input, the rest of the mass curve trends were well simulated in both the calibration and the validation process for all the three subcatchments. This signifies that the calibration and validation processes were successful for the mass curves simulations.

Statistical measures of accuracy were also used to assess the results of both the calibration and validation processes for the TRMM3B42-v7 rainfall. NSCE, relative bias and CC were used in this case, as they were produced by the CREST model runs. Table 8 below shows the obtained values of NSCE, bias and CC for the calibration and validation accuracy.

Table 8: Calibration and validation statistics for Malewa, Gilgil, and Karati while using TRMM3B42-V7 rainfall

Objective Function	Calibration			Validation		
	Malewa	Gilgil	Karati	Malewa	Gilgil	Karati
NSCE	0.59	0.27	0.63	2.88E-02	0.039	2.16E-02
Bias (%)	2.06	-10.88	-13.33	-12.329	15.55	-48.956
CC	0.771	0.526	0.797	0.326	0.277	0.207

The highest NSCE value for calibration was obtained from Karati (0.63) followed by Malewa (0.59) and the least NSCE value was from Gilgil (0.27). For the validation, the highest NSCE was from Gilgil (0.039), followed by Malewa (0.029) and the least was Karati (0.022). The highest calibration and validation bias was for Karati (-13.33 and -48.96) followed by Gilgil (-10.88 and 15.55) and the least was for the Malewa (2.06 and -12.33).

The NSCE, bias and CC deteriorated in all the validation cases from the values obtained in the calibration. The discrepancy can be expected especially, when it is considered that the calibration period was shorter than the validation period. Therefore, the cumulative effect of the errors between the observed and the simulated streamflow would be more pronounced in the validation phase than in the calibration phase. The cumulative errors are especially as a contribution of the base flow which tends to keep increasing indefinitely, thus compromising the formation of peaks by the quick flow.

The possible causes of the errors in the production of peak flow and base flows with increasing time are as follows:

First is the simplification of the CREST v2.0 model physics with regard to the surface and subsurface systems and processes through constant parameters. This makes it difficult to simulate the non-linear and non-continuous flow characteristics in the long run.

Second is the linear nature of the interflow in the CREST model simulation. Notably, interflow in the crest model is discharged from the variable dumping storage (SI) through the KI discharge parameter. SI in return receives its water from the $exaI$, which is derived from the excess rain through the variable $temX$. Also, $temX$ varies linearly from equation 11, and is determined only by the available capacity and the hydraulic conductivity. Therefore, the interflow in the CREST model takes a linear form. This is contrary to the concept of the soil water retention curve which dictates nonlinear soil water movement as shown in equation 37 (Rajkai, Kabos, & Van Genuchten, 2004).

$$W(\varphi) = W_r + \frac{WM - W_r}{[1 + (a|\varphi|)^n]^{1-1/n}} \quad (\text{Eq 37})$$

Where $W(\varphi)$ is the water retention curve (L^3/L^3), $|\varphi|$ is suction pressure (L), W_r is residual water content (L^3/L^3), a is a factor of inverse of air entry suction (L^{-1}), and n is pore-size distribution measure (no units).

Thirdly, from equation 8, the excess rain is the excess of the infiltration process. This implies, therefore, that Hortonian surface flow is suppressed in the CREST model as infiltration is given the first priority in the water distribution at the soil surface (see equation 8). When the overland flow production is suppressed and the prominent flow is linear, peak flow production is reduced in the simulation, thereby reducing the NSCE and the CC values. Therefore, the observed statistical errors can be attributed to the weakness in the CREST model physics.

The optimum parameter set for the three subcatchments were as presented in table 9. It is however worth noting that there are other possible combinations of parameters that can yield the same values of the objective function. This is because of parameter compensation within the CREST model set up. Also, the physical meaning of most of the parameters can be ascertained explicitly as most of the physical process are lumped within the parameters so as to minimize on the data requirements by the CREST v2.0 model.

Table 9: Optimum parameters for the Malewa, Gilgil, and Karati subcatchments while using TRMM3B42-v7 rainfall

Parameter	Malewa River	Gilgil River	Karati River
WU0 (mm)	0.0001	0.5	0.001
SS0 (mm)	0.1	1.0	0
SI0 (mm)	2500	1.68	0
RainFact (1)	0.92382249	0.93718459	0.53065301
Ksat (mm/d)	39.17937	44.07542	0.4713779
WM (mm)	26.6705	25.3862	59.8465
B (1)	0.1016145	1.449521	0.3359513
IM (1)	0.040736	0.051221	0.001
KE (1)	0.872333	1.02114	0.923609
coeM (1)	85.05867	79.11181	96.23827
expM (1)	0.7480409	0.81	0.8554739
coeR (1)	2.071327	2.0	2.936614
coeS (1)	0.0452626	0.1305285	0.5
KS (1)	0.514212	0.7	0.53259
KI (1)	0.001059	9.0E-4	0

The *RainFact* parameter was high for both Malewa and Gilgil catchments. It was however relatively low for the Karati catchment. The attained *RainFact* parameter value trends for the different subcatchments can reflect on their canopy and vegetation layer characteristics.

The *Ksat* parameter was highest in the Karati, followed by the Gilgil, but lowest in the Malewa, although the difference was not very significant. Nevertheless, the difference in this parameter can be attributed to the soil characteristics in the various catchments. From the Harmonized World Soil database, (HWSD) the dominant soil in the Malewa catchment is planosols with characteristic silt loam top soil and clay loam subsoil. Thus the obtained *Ksat* for Malewa is within the range provided for by the HWSD lookup table (Xue & Hong, 2013) (see Appendix F). The Gilgil catchment has soils ranging between Pheozems and Luvisols from the Harmonized World Soil database. Thus the characteristic of the Gilgil catchment top soil is clay loam and the subsoil is mostly sandy clay. From the HWSD lookup table, *Ksat* for the soil type such as that of Gilgil catchment is about 36mm/day. Therefore, the obtained value is acceptable. For Karati, the dominant soils are Phaeozems and Planosols with clay loam top soil and sandy loam subsoil. From the HWSD lookup table, *Ksat* for such a soil is 261mm/day. Therefore, the *Ksat* there was a huge discrepancy for the *Ksat* value in Karati subcatchments. However, the relative hydrologic inactivity of Karati was not expected give a good relationship between its soil and the hydraulic conductivity as presented in the HWSD lookup table.

The *WM* parameter was highest for Karati and least for Malewa subcatchments. *WM* is expected to be related with the soils' rooting depth for each catchment. The obtained values cannot however be relied on in stating the rooting depth of the subcatchments. Connected to parameter *WM* is the *B* parameter, which is an exponent of the infiltration curve. Although the values for parameter *B* varied in the three catchments, the physical meaning of this parameter was thought to influence the nonlinearity of the infiltration process.

The *IM* parameter representing the impervious area ratio and the higher the value is, the more impervious is the surface. The increasing imperviousness leads to increasing surface flow and less infiltration. This parameter can also be affected by the shape of the catchment, as it relates more with the peak flow. Gilgil catchment had the highest value of impervious ratio, and this can be attributed to its shape. Note that the Gilgil catchment is elongated, while both Karati and Malewa have more circular shape. Therefore, the increasing *IM* in Gilgil can be thought of as routing the water quickly so as to produce the observed high peaks. The small *IM* value in Karati catchment can be attributed to the small size of the catchment as well as the shape, meaning that water from most part of the catchment converge quickly and at almost the same time at the outlet point.

The parameter *KE* is highest in Gilgil and Karati catchments. This can be attributed to the high evapotranspiration rates in these areas due to higher temperatures than those of Malewa as they are at lower altitudes. Also, the humidity in Malewa catchment is higher than that from the Gilgil and Karati subcatchments.

The parameters *coeM*, *coeR* and *coeS* are routing coefficients and their increase causes the peaks to increase. These parameters are notably highest for the Karati River, which is justified by the high peaks and no base flow. The routing parameters for Malewa catchment are relatively higher compared to those of Gilgil catchment, meaning that routing takes place much faster in Malewa than in Gilgil. This observation can be justified by the Malewa catchment's steep terrain, the catchment's shape as well as the channel characteristics. On the other hand, *expM* is an exponential of the *coeM* and is, thus, expected to influence the shape of the flow.

Finally, KS and KI parameters were observed to vary significantly in the three subcatchments. Gilgil subcatchments had the highest KS parameter, but its KI was significantly low. This signifies that most of the flow in the Gilgil sub-catchment comes from the surface and the subsurface is not significantly hydrological active. For Karati subcatchments, the flow only comes from the surface as KI is equal to zero. This can be justified by the seasonal nature of the Karati River. The Malewa sub-catchment has significantly high values of KI and KS , implying that the streamflow originates from both the surface and the subsurface.

7.3. Model calibration and validation while using CMORPH8km rainfall

Calibration was also done for each sub-catchment for the period January 2002 to December 2005 followed by validation from January 2006 to December 2010 using CMORPH8km rainfall data. The simulated streamflow hydrographs in the calibration process as well as in the validation process were plotted together with the corresponding observed streamflow for all the catchments. Figure 23 shows the calibration and validation streamflow hydrographs for the Malewa, Gilgil and Karati subcatchments respectively. Figure 24 shows the calibration and validation streamflow mass curves for the Malewa, Gilgil and Karati subcatchments respectively.

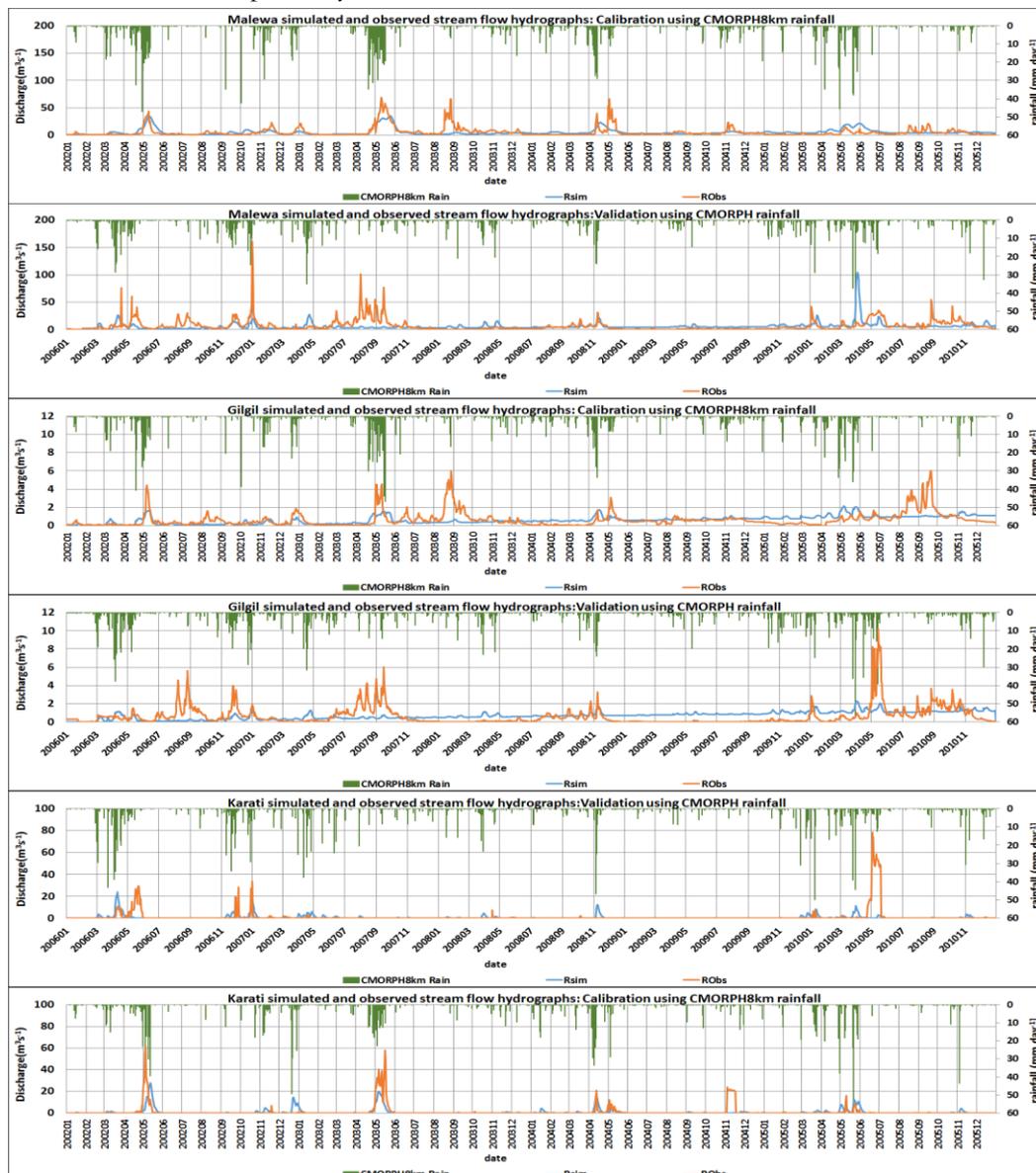


Figure 23: Streamflow hydrographs for calibration and validation for Malewa, Gilgil and Karati using CMORPH8km rainfall

The general trend in the flow dynamics was well simulated in all the catchments, both for calibration and validation. The accuracy of the simulation was especially good at the initial phases of the simulation, but deteriorated with time in all the cases. Just like the case of the TRMM3B42, most of the peaks in the observed streamflow were reproduced in the simulated hydrographs apart from those that did not coincide with a causative rainfall event. Since CMORPH8km rainfall is sparingly distributed over time but has high peaks, there was a poor match between the peaks in the observed streamflow and the high rainfall events. This mismatch affected the time to rise and fall of the peak flow as well as the heights of the peaks. Therefore, the simulated streamflow hydrographs had lower peaks than those of the observed streamflow hydrographs. The peaks for the Malewa River simulations were however observed to increase with time. Also, the impacts of the suspicious discharge events in the observed streamflow can be seen in the calibration and validation simulated streamflows.

The base flow for all the three subcatchments were well simulated for the initial phases. The simulation however deteriorated with increasing time as the base flow tended to increase continuously in all the cases. Just like in the cases of TRMM3B42-v7 the time variant deteriorating simulation trend can be attributed to the cumulative spatial and temporal inaccuracies as well as weakness in the CREST model in reproducing the base flow.

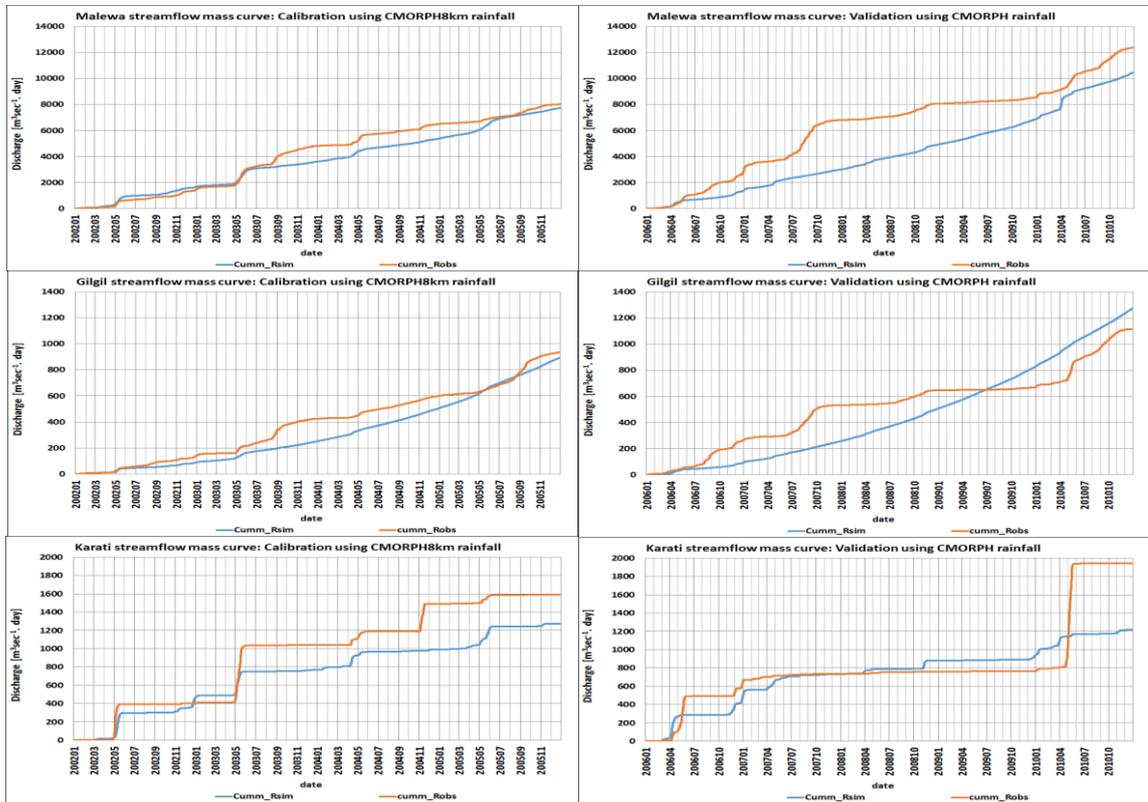


Figure 24: Streamflow mass curves for calibration and validation for Malewa, Gilgil and Karati subcatchments using CMORPH8km rainfall

In all the cases, the initial parts of the mass curves were well simulated both for the calibration and the validation. The streamflow mass curves for the calibrations show good coherence in the trend of the cumulative streamflow for all the three subcatchments. The Malewa and Karati subcatchments show lower simulated cumulative streamflow than the observed streamflow for the calibration phase. However, the Gilgil sub-catchment shows almost equal cumulative streamflow for observed and simulation over the calibration phase. Also, the simulation coherence deteriorated with time for all the cases although at varying extents. The differences in the cumulative streamflow trends for the calibration cases become

more prominent where the observed streamflows shows high flow peak events (from the suspicious flow events).

Statistical measures of accuracy were also used to assess the results of both the calibration and validation of processes for the TRMM3B42-v7 rainfall. NSCE, relative bias and CC were used in this case, as they were produced by the CREST model runs. Table 10 below shows the obtained values of NSCE, bias and CC for the calibration and validation accuracy

Table 10: Calibration and validation statistics for Malewa, Gilgil, and Karati while using CMORPH8km rainfall

Objective Function	Calibration			Validation		
	Malewa	Gilgil	Karati	Malewa	Gilgil	Karati
NSCE	0.4131	0.082	0.27	-0.25	-0.0133	-0.0233
Bias (%)	0.844	-4.268	-19.97	-15.5	14.226	-37.219
CC	0.643	0.335	0.522	0.158	0.195	0.139

The highest NSCE value for calibration was obtained from Malewa (0.41) followed by Karati (0.27) and the least NSCE value was from Gilgil (0.082). For the validation, the NSCE values were all negative with the least value being from Malewa (-0.25), Karati (-0.0233), and Gilgil (-0.0133). The highest calibration and validation bias was for Karati (-19.97 and -37.219) followed by Gilgil (-4.26 and 14.22) and the least was for the Malewa (0.844 and -15.5). The statistics shows discrepancies between the calibration and validation phases. The NSCE, bias, and CC for the validation deteriorated in all the cases.

The resulting optimum parameter values for the three subcatchments' calibration were as shown in table 11 below:

Table 11: Optimum parameters for Malewa, Gilgil and Karati while using CMORPH-8km rainfall

Parameter	Malewa River	Gilgil River	Karati River
WU0 (mm)	0.0001	0.5	0.001
SS0 (mm)	0.1	0.1	0
SI0 (mm)	2500	40	0
RainFact (1)	0.63166829	0.61	0.60738736
Ksat (mm/d)	58.2538	90	0
WM (mm)	100	50	42.9341
B (1)	0.3638892	0.4	1.206034
IM (1)	0.053388	0.05	0.1
KE (1)	1.4377	0.35	0.815153
coeM (1)	118.2883	103.0272	91.39401
expM (1)	0.7030533	0.8560200	0.476334
coeR (1)	1.114139	1.987218	1
coeS (1)	0.1633333	0.1975694	0.3
KS (1)	0.399889	0.8	0.363427
KI (1)	0.00115	0.001	0

There were notable changes in the optimum parameter values for CMORPH8km rainfall as compared to those for TRMM3B42-v7. Since by changing the rainfall product we are changing an input into the model, it was expected that only the RainFact parameter would change. However, this is not the case as there is no notable relationship between optimum parameters for the CREST model between the two satellite rainfall products. The differences evidenced in the optimum parameters values can be attributed to parameter compensations within the CREST model.

7.4. Streamflow Flow into Lake Naivasha

The streamflow into Lake Naivasha mainly results from the overall contribution of the three rivers as mentioned earlier. Thus, the summation of the daily streamflow for each of the three rivers was done. It is however important to note that some areas in the basin were not simulated since the gauge stations considered in this study are at considerable distance upstream of the pour point into the lake. Also, according to the digital elevation model of the basin, the converging point of the three rivers lies inside the lake; therefore, the flow cannot be modelled as a unit system. Nevertheless, the streamflow into the lake was determined by combining the daily streamflow contribution by each river for each of the rainfall product.

Figure 25 shows the comparison of streamflow hydrographs into lake Naivasha as simulated by the CREST model using the TRMM3B42-v7 and CMORPH8km satellite rainfall products. The hydrographs are plotted together with the observed gauge discharge. From the plots, the general trends of high peaks in the rainy seasons and low peaks in the dry seasons were simulated appropriately for all the rainfall products. There is generally a good agreement for the time to peak as the accession and recession phases coincide in most cases for all the streamflows.

The notable differences are in the base flows, the peak heights and the streamflow events described as suspicious in previous sections. The overall observed discharges have the highest peaks in most high flow events followed by the flow simulated using TRMM3B42-v7 rainfall.

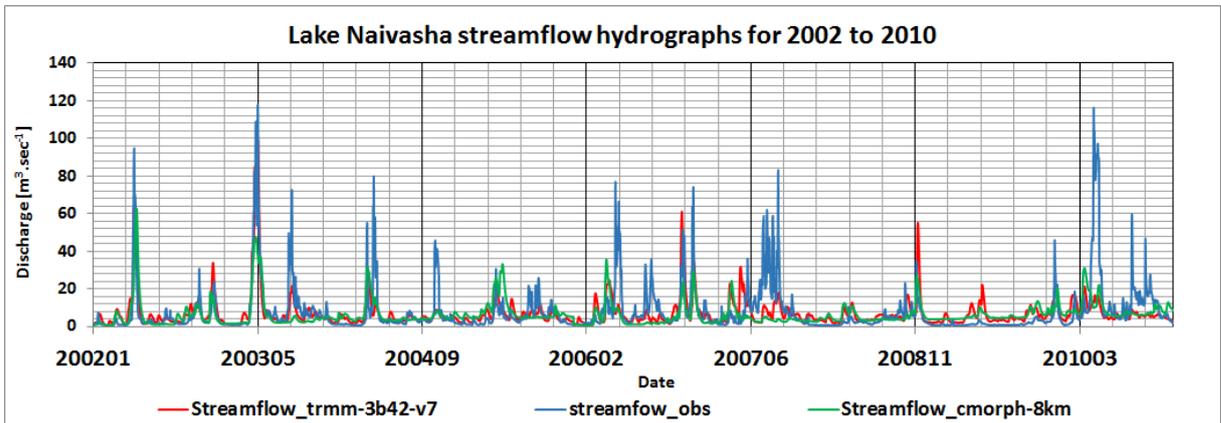


Figure 25: Streamflow hydrographs for the flow into Lake Naivasha as modelled using the CREST v2.0 distributed model and TRMM3B42-v7 and CMORPH8km rainfall

Figure 26 shows the comparison of the streamflow into lake Naivasha mass curves as simulated by the CREST model using the TRMM3B42-v7 and CMORPH8km rainfalls. The mass curves are also plotted

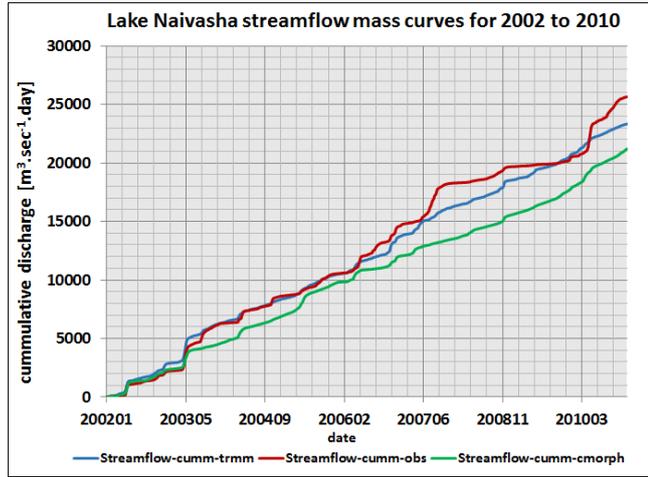


Figure 26: Streamflow mass curves for the flow into Lake Naivasha as modelled using the CREST v2.0 distributed model and TRMM3B42-v7 and CMORPH8km rainfall

together with the observed gauge flow mass curve. The observed gauge flow mass curve produces the highest overall volume. The cumulative volume from TRMM3B42-v7 simulations is in a closer relation with that of the observation. CMORPH8km rainfall produced the least cumulative rainfall.

7.5. Lake Naivasha Water balance

Water balance of the lake is given by:

$$\frac{\Delta S}{\Delta t} = \text{input} - \text{output} \quad (\text{Eq 38})$$

where $\Delta S/\Delta t$ is change in storage over time (mm/day).

The input and output of the lake are calculated while ignoring the ground water contributions as well as abstraction for other purposes (such as irrigation) from the streams and the lake. Thus

Input into the lake is given by

$$\text{input} = \text{streamflow (mm/day)} + \text{rainfall on lake surface (mm/day)} \quad (\text{Eq 39})$$

$$\text{output} = \text{open water evaporation (mm/day)} \quad (\text{Eq 40})$$

Output from the lake is only through evaporation from the open water surface (mm/day). Thus the change daily change in the lake storage was calculated for three cases:

1. Using in-situ data of streamflow, rainfall, and pan evapotranspiration
2. Using simulated streamflow from the TRMM3B42-v7 rainfall, TRMM3B42-v7 rainfall over the Lake's surface and FEWSNET PET.
3. Using simulated streamflow from the CMORPH8km rainfall, CMORPH8km rainfall over the Lake's surface and FEWSNET PET.

Figure 27 below shows the daily lake levels calculated using the streamflow into the lake as simulated by the CREST model. The calculation of the lake balance was as shown in equation 34 and it considered the processes stated in equations 35 and 36.

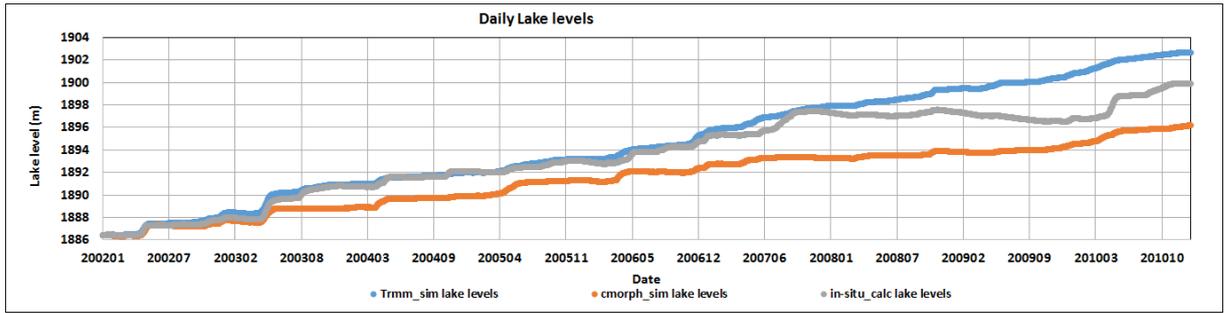


Figure 27: Plots for the daily lake level as calculated from the stream inputs using in-situ data, TRMM3B42-v7 simulations and CMORPH8km simulations. The plots shows a good coherence between the Lake levels calculated using TRMM3B42-v7 simulations and that calculated using the in-situ data

In general, the lake level was found to be in an increasing trend owing to the increasing trend of streamflow into the lake for all the three cases. The lake levels obtained from the TRMM3B42-v7 simulations have a good coherence with the lake levels obtained from the in-situ data. The difference between the lake levels from TRMM3B42-v7 simulation and that calculated from the in-situ data can be attributed to the errors in simulations as well as the abstraction of water from the rivers for irrigation and other purposes. The results of the lake levels from the CMORPH8km simulations were lower than those calculated from the in-situ data. This difference is as a result of the general poor simulations of the streamflow while using the CMORPH8km for this study as well as scale errors due to scale differences between the in-situ and the satellite based products. The relationships of the three cases of the streamflow contributions to the lake level were also assessed for the bias, mean absolute error and the root mean square error as shown in table 12 below.

Table 12: The statistical measures of accuracy for the calculated Lake levels using the TRMM 3B42-v7 simulations as well as using the CMORPH8km simulations against the calculated streamflow input into the lake using the in-situ data

Lake balance		
Statistical Measure	TRMM3B42-v7 simulations	CMORPH8km simulations
Bias	0.162	0.133
MAE	5.705	6.154
RMSE	9.568	10.894

Also, the calculated lake level changes from the three cases were assessed against the daily observed lake levels and a factor representing the contribution by the ground water, the abstraction of water from the lake and mismatch between utilized satellite products and in-situ rainfall and evapotranspiration was introduced as (K). Thus, equation 36 changed to:

$$\text{output} = \text{open water evaporation (mm/day)} + K(\text{mm/day}) \quad (\text{Eq 41}).$$

The factor K was thus determined for each of the three cases by optimising the relative error between the calculated lake level and the observed lake level so as to gain an insight on the level of magnitude of the processes not accounted for in this study. The factor K thus represented an average of the daily discrepancies between the actual lake level and those obtained in this study. Results for factor k were as follows:

Table 13: Values for the factor representing the water abstracted form the lake as well as the ground water contribution to the lake level in (mm/day).

Case	K (mm/day)
Case 1	4.113
Case 2	4.951
Case 3	2.980

From the k values above, the case of TRMM3B42-v7 (case 2) and that of the in-situ data (case 1) shows almost equal values for the ground water contribution and water abstraction. A good relationship for the TRMM3B42-v7 was also observed in the plots for the daily lake levels. The plots in figure 29 were obtained for the comparison of the different lake level values.

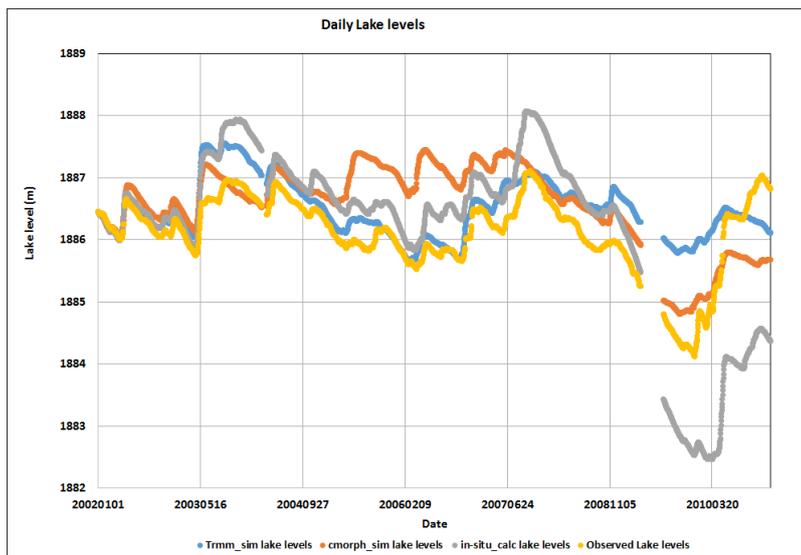


Figure 28: Plots for the Lake Naivasha daily levels from 2002 to 2010, showing the comparison between the obtained lake levels from the streamflow simulations and calculations from in-situ data against the observed lake levels

From the curves, the lake levels obtained from the TRMM3B42-v7 show the best relation with the observed lake level, apart from the last phase of the plots. There is also a notable rapid decline in lake level from late 2008 to early 2010. This decline can be attributed to decline in streamflow as it is seen in all the three cases. The decline is however more exaggerated in the lake level calculations using the in-situ data, implying that there was low measured discharge than the simulated but this was not investigated further. Also, the contribution of some of the Naivasha Basin's area to the streamflow were not considered as the discharge stations used in the study are upstream of the pour point into the lake. Therefore, the volume of streamflow used is less than the actual volume contributed by the whole basin.

The statistics for the comparison of the lake levels as obtained from the 3 cases of streamflow corrected for the factor K, were obtained against the observed lake levels. The various statistical measures done were in agreement with the preceding assessment trends. To elaborate this further, the TRMM3B42-v7 case had better results for the RMSE, and MAE, just it is in the various assessments. On the same note, the CMORPH8km continued the trend of yielding better results for the bias. Also included in this assessment is the relative volume error (RVE) so as to assess for the overall relationship between the volumes. The RVE is obtained as:

$$RVE = \left[\frac{\sum_{i=1}^n (Q_{sim} - Q_{obs})}{\sum_{i=1}^n Q_{obs}} \right] \times 100 \quad (\text{Eq 42})$$

Where Q_{sim} and Q_{obs} are simulated and observed discharges in mm/day.

Table 14: Statistical measures of accuracy for the relationship between the calculated lake levels from the three cases of streamflow against the lake levels observed in the lake level station by WARMA-Naivasha.

Lake level statistics			
Statistical Measure	TRMM3B42-v7 simulations	CMORPH8km simulations	in-situ
Bias	-4.823E-16	-2.186E-16	4E-16
MAE (m)	0.419	0.595	0.775
RMSE (m)	0.56	0.723	2.155
Averages (m)	1886.494	1886.572	1886.201
RVE	0.115	0.119	0.099

From the table 14 above, it can be seen that the lake levels calculated using the in-situ data (i.e. case 1) had a good relationship with the lake level observed by WARMA-Naivasha. Also, the lake levels obtained from the TRMM3B42-v7 calculations (i.e. case 2) are more reliable than those obtained from CMORPH8km (i.e. case 3).

8. FINAL REMARKS

8.1. Conclusions

This study shows that the streamflow into Lake Naivasha can be reconstructed using the CREST model and rainfall inputs of TRMM3B42-v7 and CMORPH8km. In general, this study entailed the acquisition of input data and their consequent pre-processing and quality assessment, CREST model setup, calibration and validation, and assessment of the overall streamflow input to the Lake Naivasha water balance. The streamflow into Lake Naivasha was reconstructed for the period 2001 to 2010 by simulating the streamflow for the three main rivers draining into Lake Naivasha namely, Malewa, Karati and Gilgil using the CREST distributed hydrologic model. The PEST calibration procedure was used in this study as it improved the efficiency in the number of the model for the calibration, thereby saving time.

The reliability assessment of both the TRMM-3B42-v7 and the CMORPH8km rainfall products was done by comparing their seasonal trends, cumulative volumes and scatterplots with the in-situ rainfall data from averages of 5 selected raingauges within the basin for a period of 10 years. The assessment was done for both the daily averages as well as the monthly averages to gain some insights on the temporal variation of errors in the data. The reliability assessment of the rainfall products showed that TRMM3B42-v7 produces the most frequent rain events of the three rainfall datasets, while CMOPRH8km produces the least rainfall events but with the highest peaks. Also TRMM3B42-v7 produced the highest cumulative rainfall. The daily scatterplots were very noisy depicting the presence of temporal and spatial errors. This was further affirmed by the monthly scatterplots that were smooth and not very biased as the daily scatterplots, showing that the spatial and temporal errors were smoothed in the long run. TRMM3B42-v7 performed better for the cases of the Root Mean Square Error (daily RMSE=4.21mm, monthly= 1.77mm) and the Mean Absolute Error (Daily MAE=2.48mm, monthly= 0.883mm), than the CMORPH8km (daily RMSE=5.57mm, monthly= 1.77mm) and (Daily MAE=2.69mm, monthly= 2.1mm). On the other hand, CMORPH8km had a less bias (daily Bias= -0.334, monthly= -0.403) than the TRMM3B42-v7 (daily Bias=-0.708, monthly=-0.635).

The measured discharge data for Malewa, Gilgil, and Karati were assessed for their consistency with the in-situ rainfall data. Generally, there were mismatches between the seasonal dynamics in the discharge and the rainfall data signifying presence of errors in the observation data. The study shows that the CREST model was able to reproduce the general streamflow dynamics, as well as the volumes of the streamflow from the three subcatchments in the Naivasha basin. The results from the calibration and validation of the three rivers in the Naivasha basin were better while using the TRMM3B42-v7 rainfall than when using the CMORPH8km in terms of the seasonal flow dynamics as well as the volumes. Also, most of the optimum parameter values while using the PEST calibration procedure, relates closely with those obtained from literature. However, the base flow produced by the CREST model increased continuously, which is not the case with the observed discharge.

The streamflow simulated from the satellite derived rainfall were assessed for their contribution to the lake level change using the water balance for the lake against the in-situ observations. The change in storage as the change in lake level, streamflow and rainfall on the lake surface as the input and evaporation from the lake surface as the output. Simulations from the TRMM3B42-v7 yielded better RMSE (9.57mm) and MAE (5.70mm) than those of CMORPH (RMSE=10.89 and MAE=6.15). There was observable rapid decline in the lake level from the late 2008 to early 2010, after which the lake level started to recover. Although the region experienced some droughts over this period, the rapid decline can be attributed

largely to the expansion in the floriculture farming both around the lake and upstream of the lake (World Wildlife Fund, 2010). The recovery of the lake level from early 2010 can be associated with the deliberate efforts to safeguard the lake, including the WWF.

8.2. Recommendations

There is the need to investigate and assess the uncertainties in the in-situ discharge data. Since the calibration targeted to relate the simulated discharge against the observed discharge, the outstanding mismatches between the rainfall events and the stream dynamics made it difficult to simulate the base flow properly. Also, the sizes of the simulated peaks and mismatches between the simulated streamflow peaks and the in-situ streamflow peaks contributed to the low values of NSCE and CC. Since most of the original discharge data had over 70% gaps, the interpolation procedure for filling the gaps need to consider the in-situ rainfall data that has relatively higher continuity than the discharge data.

It is recommended that the CMORPH8km rainfall be calibrated for the Naivasha basin using raingauge data before using with the CREST model. The calibration of the CMORPH8km for the Naivasha basin will enable the future modelling to utilize the advantage of the observed good performance in bias as compared to the in-situ rainfall. Moreover, the CMORPH8km has a higher spatial and temporal resolution than the TRMM3B42-v7.

The other recommendations are with regard to the CREST V2.0 model physics. First, the model simplifies the surface and subsurface systems and processes by use of constant parameters. The model should allow input of distributed maps, such as those of land cover and soil properties. Secondly, the CREST model should be improved to simulate the soil moisture storage, by incorporating the capillary rise and connecting the soil moisture storage and the interflow storage. It is expected that this can solve the experienced challenge of continuously increasing base flow in the Naivasha basin. If the soil moisture can be allowed to rise and add to the moisture storage in the upper layers, it can contribute to production of delayed peak flows such as those evidenced in the Naivasha basin, especially in the case of the Gilgil river. Also, the infiltration process should take into consideration the soil moisture retention curve concept so as to represent the interflow process more effectively.

LIST OF REFERENCES

- Al, J. E. T. (2004). CMORPH: A Method that Produces Global Precipitation Estimates from Passive Microwave and Infrared Data at High Spatial and Temporal Resolution.
- Artan, G., Gadain, H., Smith, J. L., Asante, K., Bandaragoda, C. J., & Verdin, J. P. (2007). Adequacy of satellite derived rainfall data for stream flow modeling. *Natural Hazards*, 43(2), 167–185. doi:10.1007/s11069-007-9121-6
- Bajracharya, S. R., Palash, W., Shrestha, M. S., Khadgi, V. R., Duo, C., Das, P. J., & Dorji, C. (2014). Systematic Evaluation of Satellite-Based Rainfall Products over the Brahmaputra Basin for Hydrological Applications.
- Becht, R., & Harper, D. M. (2002). Towards an understanding of human impact upon the hydrology of Lake Naivasha, Kenya. In *Hydrobiologia* (Vol. 488, pp. 1–11).
- Becht, R., Science, G., & Itc, E. O. (2011). *Lake Naivasha Experience and Lessons Learned Brief*.
- Bergner, A. , Trauth, M., & Bookhagen, B. (2003). Paleoprecipitation estimates for the Lake Naivasha basin (Kenya) during the last 175 k.y. using a lake-balance model. *Global and Planetary Change*, 36(1-2), 117–136. doi:10.1016/S0921-8181(02)00178-9
- Blasone, R.-S., Madsen, H., & Rosbjerg, D. (2007). Parameter estimation in distributed hydrological modelling: comparison of global and local optimisation techniques. *Nordic Hydrology*, 38(4-5), 451. doi:10.2166/nh.2007.024
- Chen, W., Jin, R., & Sudjianto, A. (2005). Analytical Variance-Based Global Sensitivity Analysis in Simulation-Based Design Under Uncertainty. *Journal of Mechanical Design*, 127(5), 875. doi:10.1115/1.1904642
- Climate Prediction Center Internet Team. (2013). NOAA Center for Weather and Climate Prediction. Retrieved August 15, 2014, from http://www.cpc.ncep.noaa.gov/products/janowiak/cmorph_description.html
- Deng, Y., Bowman, K. P., & Jackson, C. (2007). Differences in rain rate intensities between TRMM observations and community atmosphere model simulations. *Geophysical Research Letters*, 34(1), L01808. doi:10.1029/2006GL027246
- Doherty, J. (1994). *PEST: Model-Independent Parameter Estimation*.
- Druyan, L., & Fulakeza, M. (2013). Downscaling reanalysis over continental Africa with a regional model: NCEP versus ERA Interim forcing. *Climate Research*, 56(3), 181–196. doi:10.3354/cr01152
- Duan, Q., Sorooshian, S., & Gupta, V. K. (1994). Optimal use of the SCE-UA global optimization method for calibrating watershed models. *Journal of Hydrology*, 158(158), 265–284.
- Everard, M., & Harper, D. M. (2002). Towards the sustainability of the Lake Naivasha Ramsar site and its catchment. *Hydrobiologia*, 488, 191–203.
- Everard, M., Vale, J. A., Harper, D. M., & Tarras-wahlberg, H. (2002). The physical attributes of the Lake Naivasha catchment rivers. *Hydrobiologia*, 488(1998), 13–25.

- Gathenya, J. M. (2012). *Flood disaster management in Lake Victoria basin , Kenya*.
- Gupta, S, K. (2011). *Modern Hydrology and Sustainable Water Development* (1st ed., p. 472). Hoboken: Wiley. Retrieved from <http://ezproxy.utwente.nl:2163/patron/FullRecord.aspx?p=792631>
- Harper, D., & Mavuti, K. (2004). Lake Naivasha , Kenya : Ecohydrology to guide the management of a tropical protected area. *Ecology and Hydrobiology*, 4(3), 287–305.
- Joyce, R. J., Janowiak, J. E., Arkin, P. A., Xie, P., & Comph, A. (2004). CMORPH 3: A method that produces global precipitation estimates from passive microwave and infrared data at high spatial temporal resolution. *Hydromet*.
- Khan, Sadiq, I., Adhikari, P., Hong, Y., Vergara, H., Grout, T., Adler, R. F., ... Okello, L. (2010). Observed and simulated hydroclimatology using distributed hydrologic model from in-situ and multi-satellite remote sensing datasets in Lake Victoria region in East Africa. *Hydrology and Earth System Sciences Discussions*, 7(4), 4785–4816. doi:10.5194/hessd-7-4785-2010
- Khan, Sadiq, I., Hong, Y., Wang, J., Yilmaz, K. K., Gourley, J. J., Adler, R. F., ... Irwin, D. (2011). Satellite Remote Sensing and Hydrologic Modeling for Flood Inundation Mapping in Lake Victoria Basin: Implications for Hydrologic Prediction in Ungauged Basins. *IEEE Transactions on Geoscience and Remote Sensing*, 49(1), 85–95. doi:10.1109/TGRS.2010.2057513
- Khan, Sadiq, I., Vergara, H. J., Gourley, J. J., Brakenridge, G. R., De Groeve, T., Flamig, Z. L., & Policelli, F. (2012). Microwave Satellite Data for Hydrologic Modeling in Ungauged Basins. *IEEE Geoscience and Remote Sensing Letters*, 9(4), 663–667. doi:10.1109/LGRS.2011.2177807
- Li, X.-H., Zhang, Q., & Xu, C.-Y. (2012). Suitability of the TRMM satellite rainfalls in driving a distributed hydrological model for water balance computations in Xinjiang catchment, Poyang lake basin. *Journal of Hydrology*, 426-427, 28–38. doi:10.1016/j.jhydrol.2012.01.013
- Limaye, A., Gitau, J., & Kabuchanga, E. (2011). *Hydrologic Early Warning System*.
- Meins, F. M. (2013). *Evaluation of spatial scale alternatives for hydrological modelling of the Lake Naivasha basin, Kenya*. ITC-Faculty of Geo-Information Science and Earth Observation.
- Meng, J., Li, L., Hao, Z., Wang, J., & Shao, Q. (2014). Suitability of TRMM satellite rainfall in driving a distributed hydrological model in the source region of Yellow River. *Journal of Hydrology*, 509, 320–332. doi:10.1016/j.jhydrol.2013.11.049
- Mishra, S. (2009). Uncertainty and sensitivity analysis techniques for hydrologic modeling. *Journal of Hydroinformatics*, 11(3–4), 282. doi:10.2166/hydro.2009.048
- Montzka, C., Canty, M., Kunkel, R., Menz, G., Vereecken, H., & Wendland, F. (2008). Modelling the water balance of a mesoscale catchment basin using remotely sensed land cover data. *Journal of Hydrology*, 353(3-4), 322–334. doi:10.1016/j.jhydrol.2008.02.018
- Muleta, M. K., & Nicklow, J. W. (2005). Sensitivity and uncertainty analysis coupled with automatic calibration for a distributed watershed model. *Journal of Hydrology*, 306(1-4), 127–145. doi:10.1016/j.jhydrol.2004.09.005
- Muthuwatta, L. P. (2004). *Long term rainfall-runoff-lake level modelling of the Lake Naivasha Basin, Kenya*. ITC.

- National Center for Atmospheric Research Staff (Eds). (2013). The Climate Data Guide: PERSIANN: Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks. Retrieved September 08, 2014, from <https://climatedataguide.ucar.edu/climate-data/persiann-precipitation-estimation-remotely-sensed-information-using-artificial-neural>
- NOAA. (2014). Climate Prediction Center. Retrieved September 05, 2014, from http://www.cpc.ncep.noaa.gov/products/janowiak/cmorph_description.html
- Obiero, J. P. O. (2011). Modelling of Streamflow of a Catchment in Kenya. *Journal of Water Resource and Protection*, 03(09), 667–677. doi:10.4236/jwarp.2011.39077
- Odongo, V. O., Mulatu, D. W., Muthoni, F. K., van Oel, P. R., Meins, F. M., van der Tol, C., ... van der Veen, A. (2014). Coupling socio-economic factors and eco-hydrological processes using a cascade-modeling approach. *Journal of Hydrology*. doi:10.1016/j.jhydrol.2014.01.012
- Odongo, V. O., Onyando, J. O., Mutua, B. M., van Oel, P. R., & Becht, R. (2013). Sensitivity analysis and calibration of the Modified Universal Soil Loss Equation (MUSLE) for the upper Malewa Catchment, Kenya. *International Journal of Sediment Research*, 28(3), 368–383. doi:10.1016/S1001-6279(13)60047-5
- Prakash, S., Mahesh, C., & Gairola, R. M. (2013). Comparison of TRMM Multi-satellite Precipitation Analysis (TMPA)-3B43 version 6 and 7 products with rain gauge data from ocean buoys. *Remote Sensing Letters*, 4(7), 677–685. doi:10.1080/2150704X.2013.783248
- Rajkai, K., Kabos, S., & Van Genuchten, M. T. (2004). Estimating the water retention curve from soil properties: Comparison of linear, nonlinear and concomitant variable methods. *Soil and Tillage Research*, 79, 145–152. doi:10.1016/j.still.2004.07.003
- Skahill, B. E., & Doherty, J. (2006). Efficient accommodation of local minima in watershed model calibration. *Journal of Hydrology*, 329(1-2), 122–139. doi:10.1016/j.jhydrol.2006.02.005
- Sorooshian, S., Hsu, K., Imam, B., & Hong, Y. (2005). Global Precipitation Estimation from Satellite Image Using Artificial Neural Networks. *Journal of Applied Meteorology*, 36, 1176–1190. Retrieved from http://www.g-wadi.org/shortcourses/chapters/Sorooshian_L2.pdf
- Tiruneh, B. A. (2004). *Modelling Water Quality Using Soil and Water Assessment Tool (SWAT); A Case Study in Lake Naivasha Basin, Kenya*.
- Trucano, T. G., Swiler, L. P., Igusa, T., Oberkampf, W. L., & Pilch, M. (2006). Calibration, validation, and sensitivity analysis: What's what. *Reliability Engineering & System Safety*, 91(10-11), 1331–1357. doi:10.1016/j.ress.2005.11.031
- USGS. (2012). USGS. Retrieved August 09, 2014, from <http://earlywarning.usgs.gov/fews/global/web/readme.php?symbol=pt>
- USGS HydroSHEDS. (2013). Retrieved May 23, 2014, from <http://hydrosheds.cr.usgs.gov/index.php>
- Wainwright, H. M., Finsterle, S., Jung, Y., Zhou, Q., & Birkholzer, J. T. (2014). Making sense of global sensitivity analyses. *Computers & Geosciences*, 65, 84–94. doi:10.1016/j.cageo.2013.06.006
- Wang, J., Hong, Y., Li, L., Gourley, J. J., Khan, S. I., Yilmaz, K. K., ... Okello, L. (2011a). The coupled routing and excess storage (CREST) distributed hydrological model. *Hydrological Sciences Journal*, 56(1), 84–98. doi:10.1080/02626667.2010.543087

- Wang, J., Hong, Y., Li, L., Gourley, J. J., Khan, S. I., Yilmaz, K. K., ... Okello, L. (2011b). The coupled routing and excess storage (CREST) distributed hydrological model. *Hydrological Sciences Journal*. doi:10.1080/02626667.2010.543087
- WARMA. (2009). *Integrated Water Resources Management and Water Efficiency Plan for Kenya August 2009*.
- WARMA. (2010). *Water Allocation Plan - Naivasha Basin*.
- World Wildlife Fund. (2010). *Shared risk and opportunity in water resources : Seeking a sustainable future for Lake Naivasha* (p. 47).
- Xue, X., & Hong, Y. (2013). CREST Coupled Routing and Excess Storage User Manual.
- Xue, X., Hong, Y., Limaye, A. S., Gourley, J. J., Huffman, G. J., Khan, S. I., ... Chen, S. (2013). Statistical and hydrological evaluation of TRMM-based Multi-satellite Precipitation Analysis over the Wangchu Basin of Bhutan: Are the latest satellite precipitation products 3B42V7 ready for use in ungauged basins? *Journal of Hydrology*, 499, 91–99. doi:10.1016/j.jhydrol.2013.06.042
- You, Y., Liu, G., Wang, Y., & Cao, J. (2011). On the sensitivity of Tropical Rainfall Measuring Mission (TRMM) Microwave Imager channels to overland rainfall. *Journal of Geophysical Research*, 116(D12), D12203. doi:10.1029/2010JD015345
- Zabinsky, Z. B. (2003). *Stochastic Adaptive Search for Global Optimization* (p. 224). London: Kluwer Academic Publishers. doi:10.1007/978-1-4419-9182-9
- Zhan, C., Song, X., Xia, J., & Tong, C. (2013). An efficient integrated approach for global sensitivity analysis of hydrological model parameters. *Environmental Modelling & Software*, 41, 39–52. doi:10.1016/j.envsoft.2012.10.009

9. APPENDICES

Appendix A: The procedure followed in obtaining streamflow data (Meins, 2013)

Stream flow data interpolation

Stream flows are measured by reading water levels from gauging staffs at a number of locations in the basin. These water levels are then converted to stream flows by using rating curves. However, the locations of these gauging staffs in the Naivasha basin were not very well documented and different sources provided different locations. A field work expedition was organised to identify the exact locations of all stations and to obtain all the stream flow data that is available for the stations. Digital data was collected and cross-checked with hard copy data for the period between 1960 and 2010. The gauging locations (stations) that have data for this period are shown in Table 1. The 2GA stations are located in the Gilgil River, the 2GB stations are located in the Malewa and Wanjohi Rivers, the 2GC stations are located in the Turasha River and its tributaries and the 2GD stations in the Karati River.

Table 1: Characteristics of river gauging stations in the Naivasha basin

Station	River	WRUA	X-coordinate	Y-coordinate	Elevation [m]
2GA01	GILGIL	LOWER GILGIL	206516.6	9933672.7	1920
2GA03	GILGIL	LOWER GILGIL	204407.4	9945597.2	1996
2GA05	GILGIL	LOWER GILGIL	-	-	-
2GA06	LITTLE GILGIL	LOWER GILGIL	206495.0	9944747.3	2013
2GB01	MALEWA	LOWER MALEWA	210908.0	9938530.8	1951
2GB03	MALEWA	UPPER MALEWA	221632.3	9973620.2	2366
2GB04	WANJOHI	WANJOHI	219808.8	9969175.2	2334
2GB05	MALEWA	LOWER MALEWA	210688.5	9945446.0	1987
2GB0708	MALEWA	MIDDLE MALEWA	212081.6	9964640.5	2264
2GC04	TURASHA	LOWER MALEWA	212451.6	9946983.4	2000
2GC05	KITIRI	MKUNGI KITIRI	228295.2	9939060.7	2408
2GC07	TURASHA	UPPER TURASHA	236961.6	9928708.5	2708
2GC10	MUKUNGI	MKUNGI KITIRI	225447.7	9942224.9	2419
2GD02	KARATI	LANAWRUA	212710.1	9923164.6	1896
2GD07	KARATI	KARATI LONGONOT	225922.4	9919576.0	2506

Maps with locations of the gauging stations are shown in Appendix A and B. The exact location of the currently inactive 2GA05 station could not be verified, but it was supposedly located not too far downstream of the current 2GA01 station. 2GB07 and 2GB08 are basically the same stations, but on a slightly different location. In 1997 the 2GB07 station was demolished and in 1998 a new station named 2GB08 was installed approximately 200 meters downstream. These two stations are considered as one station named 2GB0708, which of course does have two different rating curves as is explained below.

The available data consist of water levels measured daily and in some cases sub-daily. In order to be able to use the data for calibration and validation complete daily stream flow series are required. However, due to various circumstances, such as broken staffs and unreliable readers, not all gauging stations contain a complete time series with water levels from 1960 till

2010. To obtain complete daily stream flow series the water levels are converted to stream flows first and then interpolated as is explained in the next two sections. An overview of the steps taken to achieve this is shown in Figure 1.

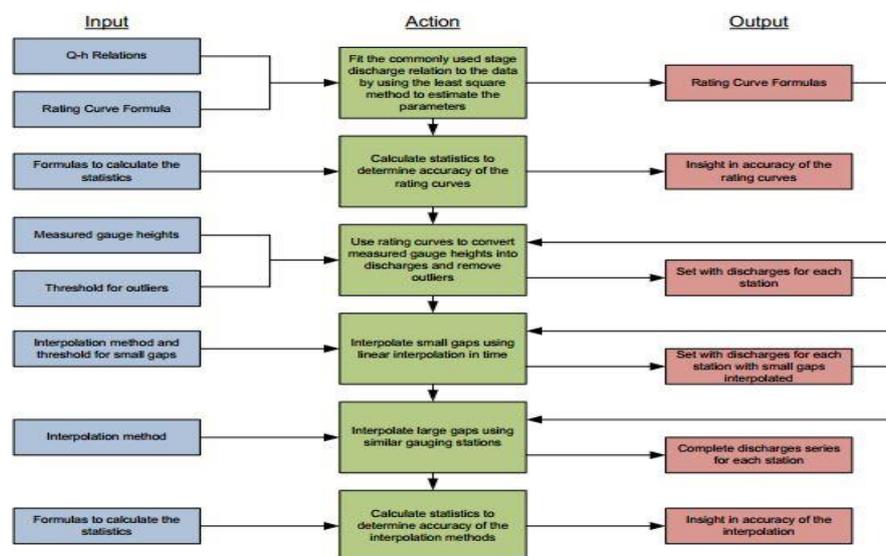


Figure 1: Stream flow interpolation scheme used in this study to fill the gaps in the available data series

Rating curves

The first step towards obtaining stream flow series (in m^3/s) is to convert measured water levels to stream flows. This is done using rating curves which define the characteristic relationship between water level and stream flow. A number of studies has been done to obtain this relationship for some gauging stations in the Naivasha basin (De Jong, 2011a; Podder, 1998), but none of them developed rating curves for all stations consistently. Therefore new rating curves are developed for each station. The rating curves are obtained by using measured stream flows at several points in time for different water levels and then fitting a function through these points. It is common practise to use a function derived from Chezy's Law for a simple rectangular river profile, which is as follows;

$$Q = C(H - H_0)^n \quad \text{if } H > H_0$$

$$Q = 0 \quad \text{if } H \leq H_0$$
Eq. 1

where Q is the stream flow (m^3/s), H is the measured water depth (m), H_0 is the threshold water depth at which the water starts flowing (m) and C and n are coefficients. Using the least squares method the coefficients C and n as well as the initial water level H_0 can be solved using the MS Excel Solver for each river gauging station. The Q-H data used consists of a number of stream flow values with their corresponding water levels measured at random moments in time between 1960 and 2010, where Q is measured using the velocity-area method. By using this method it is assumed that the channel geometry and thus the stage discharge relationship has not changed over time. Also possible effects of hysteresis and backwater curves are not accounted for. Nonetheless for each river gauging station a rating curve was developed based on the data, the results of the rating curve analysis are provided in Appendix C.

Stream flow interpolation

Once water levels are converted to stream flows they need to be interpolated in order to fill the many gaps that are contained within the data. Outliers in the data are removed first using a specific threshold for each station. Then the data are aggregated to daily values because in some instances there are two measurements per day while SWAT requires only one value per day, this is done by simple averaging. The next step is to interpolate the smaller gaps of less than 7 days; these gaps are interpolated using linear interpolation;

$$Q_{i,j} = \frac{Q_{\text{end},j} - Q_{\text{start},j}}{N_j} i + Q_{\text{start},j}$$
Eq. 2

where Q_i is the stream flow on day i in gap j , N_j is the length of gap j (number of days), $Q_{\text{start},j}$ is the stream flow at the start of gap j and $Q_{\text{end},j}$ is the stream flow at the end of gap j .

To interpolate the larger gaps of 7 days or more a method developed by Hughes & Smakhtin (1996) is used. This method aims to interpolate the missing data by using other stations that do have data upstream or downstream in the same river system. In Figure 2 an example of this method is given where one sample is interpolated. First a selection of river gauging stations is made that will act as source stations to interpolate one destination station. For both source stations and destination station flow duration curves (DTQ) are developed. 17 exceedence percentages i (0.01 0.10 1 5 10 20 30 40 50 60 70 80 90 95 99 99.90 99.99) are calculated to avoid the complication of fitting a flow duration curve to the data points. For every day that the destination station contains a gap the stream flow value (Q_S) from each of the source stations is selected for that day. Then using logarithmic interpolation the exceedence percentage is determined for each source station (DP) by interpolating between the two nearest DTQ values;

$$DP_j = \frac{\log(Q_S) - \log(DTQ_{j,i})}{\log(DTQ_{j,i-1}) - \log(DTQ_{j,i})} (DP_{j,i-1} - DP_{j,i}) + DP_{j,i}$$
Eq. 3

where the $DTQ_{j,i}$ and $DTQ_{j,i-1}$ are the closest stream flow values of the flow duration set of source station j below and above the stream flow value Q_S , $DP_{j,i}$ and $DP_{j,i-1}$ are the exceedence

percentages related to $DTQ_{j,i}$ and $DTQ_{j,i-1}$. When there is more than one source station ($j > 1$) the average exceedence percentages is taken;

$$DP_S = \frac{1}{n} \sum_{j=1}^n DP_j$$
Eq. 4

where n is the number of stations and DP_j the average exceedence percentage.

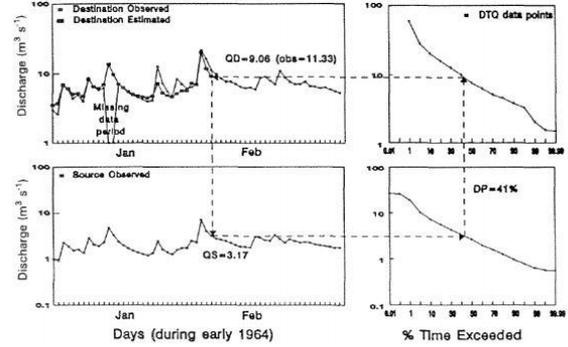


Figure 2: Stream flow interpolation method, Hughes & Smakhtin (1996)

To calculate the new stream flow value of the destination station a similar approach is used where the new stream flow value is interpolated using logarithmic interpolation and the flow duration curve of the destination station;

$$QD_S = \exp \left[\frac{DP_S - DP_{j,i}}{DP_{j,i-1} - DP_{j,i}} (\log(DTQ_{S,i-1}) - \log(DTQ_{S,i})) + \log(DTQ_{S,i}) \right]$$
Eq. 5

where QD_S is the new stream flow value for the destination station and $DTQ_{j,i}$ and $DTQ_{S,i-1}$ are the flow duration curve values at $DP_{j,i}$ and $DP_{j,i-1}$. To complete the interpolation scheme two conditions at the beginning and end of the flow duration curve need to be applied;

$$QD_S = QS \frac{DTQ_{S,1}}{DTQ_{j,1}} \quad \text{if } i = 1$$
Eq. 6

Appendix B: The ILWIS batch process for downloading and processing yearly daily FEWSNET PET

```
@echo off
echo Extract FEWS PET global daily coverage
echo File contains on yearly basis the daily data
echo File name format pet_2001.tar.gz
echo Output is days per year
echo.
echo.
set longfilename=%1
set shortfilename1=%longfilename:~0,6%
set shortfilename2=%longfilename:~0,4%
set InputDrive=%2
set InputDir=%3
set OutputDrive=%4
set OutputDir=%5
set gdalDir=%6
set IlwDir=%7
set UtilDir=%8
cd\
%OutputDrive%
cd %OutputDir%
echo your current working directory = %OutputDrive%\%OutputDir%
```

```

echo.
echo.
echo set shortfilename=%1>inputparam.bat
echo set OutputDrive=%4>>inputparam.bat
echo set OutputDir=%5>>inputparam.bat
echo set IhwDir=%7>>inputparam.bat
echo set UtilDir=%8>>inputparam.bat
echo set shortfilename2=%shortfilename2%>>inputparam.bat
"%UtilDir%\wget\wget.exe"
http://earlywarning.usgs.gov/ftp2/bulkdailydata/global/pet/years/pet_%shortfilename2%.tar.gz
echo off
if not exist "%OutputDrive%\%OutputDir%\pet_%shortfilename2%.tar.gz" goto MESSAGE
echo The file(s) "pet_%shortfilename2%.tar.gz" has been retrieved and transferred to your current working
directory
echo.
echo.
GOTO START
:MESSAGE
echo The input file was not found in the archive.
echo Check your Date stamp settings
echo Your current Date stamp used is %shortfilename1%
echo Data is retrieved from http://earlywarning.usgs.gov/ftp2/bulkdailydata/global/pet/years/
echo Check also manually if the data exists on ftp archive (see link above)
echo Internet has a transient nature - addresses might not be valid or has changed after some time!
pause
GOTO END
:START
"%UtilDir%\7z.exe" e pet_%shortfilename2%.tar.gz
"%UtilDir%\7z.exe" e pet_%shortfilename2%.tar
copy %IhwDir%\Extensions\ISOD-Toolbox\toolbox_batchroutines\fews_pet_import_loop.bat
for %%j in (et*.tar.gz) do cmd /c fews_pet_import_loop.bat %%j
del pet_%shortfilename2%.tar
del pet_%shortfilename2%.tar.gz
del fews_pet_import_loop.bat
del inputparam.bat
rd pet_%shortfilename1%
:END
"%IhwDir%\ilwis.exe" -C closeall

```

Appendix C: Python Script to convert NetCDF files to ASCII format.

```

import os, arcpy
srcPath="E:/Trmm_wrkSpace/src/" # insert the path to the directory of interest
dstPath="E:/Trmm_wrkSpace/rasters/" # insert the path to the directory where raster files will be saved
dstAscPath="E:/Trmm_wrkSpace/asciifiles/" # insert the path to the directory where ASCII files will be
saved
arcpy.env.overwriteOutput= "true"

```

```

def _net2ascii( fname ):
    print fname
    Input_NetCDF=srcPath + fname
    namee=fname.strip(".nc")
    theName=namee.replace(".7", "_v7")
    fnamee=theName.replace(".", "")
    ffnamee=fnamee.replace("_v7", "")
    Output_Raster= ffnamee.replace("3B42_daily", "TRMM_")
    print Output_Raster
    out_file=dstPath + Output_Raster
    out_asc=dstAscPath + Output_Raster + ".asc"
    # Process: Make NetCDF Raster Layer
    arcpy.MakeNetCDFRasterLayer_md(Input_NetCDF, "r", "longitude","latitude", Output_Raster, "", "",
    "By_VALUE")
    # Process: Copy Raster
    arcpy.CopyRaster_management(Output_Raster, out_file, "", "", "", "NONE", "NONE", "")
    arcpy.RasterToASCII_conversion(out_file, out_asc)
    # Process: Calculate Statistics
    arcpy.CalculateStatistics_management(out_asc, "1", "1", "")
    return
    dirList=os.listdir(srcPath)
for fname in dirList:
    _net2ascii(fname)

```

Appendix D: ILWIS batch process for downloading and processing CMORPH8km-30min

```

@echo off
echo.
echo.
rem file name input string example is CMORPH_V1.0_ADJ_8km-30min_1998020602
call inputparam.bat
set longfilename=%1
set shortfilename1=%longfilename:~26,10%
echo.
echo Processing Hour Timestep %shortfilename1%
echo.
echo.
ren CMORPH_V1.0_ADJ_8km-30min_%shortfilename1% tcmorph8km30min_%shortfilename1%
rem import as maplist and mirror rotate
"%IhwDir%\ilwis.exe" -C
%OutputDrive%\%OutputDir%\tcmorph8km30min_f%shortfilename1%.mpl:=maplist(%OutputDrive%\%
OutputDir%\tcmorph8km30min_%shortfilename1%',genras,Convert,4948,2,0,BSQ,Real,4,NoSwap,CreateM
pr)
"%IhwDir%\ilwis.exe" -C
%OutputDrive%\%OutputDir%\tcmorph8km30min_%shortfilename1%.mpl:=MapListApplic(%OutputDri
ve%\%OutputDir%\tcmorph8km30min_f%shortfilename1%',MapMirrorRotate(##,MirrHor))
rem create subset maps

```

```

"%IhwDir%\ihwis.exe" -C
%OutputDrive%\%OutputDir%\tcmorph8km30min_%shortfilename1%1_east.mpr:=MapSubMap(%OutputDrive%\%OutputDir%\tcmorph8km30min_%shortfilename1%_1',1,1,1649,2474)
"%IhwDir%\ihwis.exe" -C
%OutputDrive%\%OutputDir%\tcmorph8km30min_%shortfilename1%1_west.mpr:=MapSubMap(%OutputDrive%\%OutputDir%\tcmorph8km30min_%shortfilename1%_1',1,2475,1649,2474)
"%IhwDir%\ihwis.exe" -C
%OutputDrive%\%OutputDir%\tcmorph8km30min_%shortfilename1%2_east.mpr:=MapSubMap(%OutputDrive%\%OutputDir%\tcmorph8km30min_%shortfilename1%_2',1,1,1649,2474)
"%IhwDir%\ihwis.exe" -C
%OutputDrive%\%OutputDir%\tcmorph8km30min_%shortfilename1%2_west.mpr:=MapSubMap(%OutputDrive%\%OutputDir%\tcmorph8km30min_%shortfilename1%_2',1,2475,1649,2474)
rem add submap georef
"%IhwDir%\ihwis.exe" -C setgrf
%OutputDrive%\%OutputDir%\tcmorph8km30min_%shortfilename1%1_east.mpr
%UtilDir%\cmorph_east
"%IhwDir%\ihwis.exe" -C setgrf
%OutputDrive%\%OutputDir%\tcmorph8km30min_%shortfilename1%1_west.mpr
%UtilDir%\cmorph_west
"%IhwDir%\ihwis.exe" -C setgrf
%OutputDrive%\%OutputDir%\tcmorph8km30min_%shortfilename1%2_east.mpr
%UtilDir%\cmorph_east
"%IhwDir%\ihwis.exe" -C setgrf
%OutputDrive%\%OutputDir%\tcmorph8km30min_%shortfilename1%2_west.mpr
%UtilDir%\cmorph_west
rem glue maps using georef global
"%IhwDir%\ihwis.exe" -C
%OutputDrive%\%OutputDir%\tcmorph8km30min_%shortfilename1%1_recomposed.mpr:=MapGlue(%UtilDir%\cmorph_8km'.grf,%OutputDrive%\%OutputDir%\tcmorph8km30min_%shortfilename1%1_west,%OutputDrive%\%OutputDir%\tcmorph8km30min_%shortfilename1%1_east,replace)
"%IhwDir%\ihwis.exe" -C
%OutputDrive%\%OutputDir%\tcmorph8km30min_%shortfilename1%2_recomposed.mpr:=MapGlue(%UtilDir%\cmorph_8km'.grf,%OutputDrive%\%OutputDir%\tcmorph8km30min_%shortfilename1%2_west,%OutputDrive%\%OutputDir%\tcmorph8km30min_%shortfilename1%2_east,replace)
"%IhwDir%\ihwis.exe" -C
%OutputDrive%\%OutputDir%\cmorph1hr_%shortfilename1%:=iff(%OutputDrive%\%OutputDir%\tcmorph8km30min_%shortfilename1%1_recomposed+%OutputDrive%\%OutputDir%\tcmorph8km30min_%shortfilename1%2_recomposed ge 0,
(%OutputDrive%\%OutputDir%\tcmorph8km30min_%shortfilename1%1_recomposed/2)+(%OutputDrive%\%OutputDir%\tcmorph8km30min_%shortfilename1%2_recomposed/2),0)
rem delete obsolete objects
del tcmorph8km30min_%shortfilename1%
del tcmorph8km30min*.mp*

```

Appendix E: The batch process used to convert the hourly CMORPH8km-1hr data to daily.

I. Template to maplist hourly images in a each day

```
[Ihwis]
Description=
Time=1416494054
Version=3.1
Class=Map List
Type=MapList
[MapList]
GeoRef='C:\maatbuis\Ihwis372_July2014\Extensions\ISOD-Toolbox\util\cmorph_east'.grf
Size=1649 2474
Maps=24
BandPreFix=
Offset=0
Map0=cmorph1hr_2003090200.mpr
Map1=cmorph1hr_2003090201.mpr
Map2=cmorph1hr_2003090202.mpr
Map3=cmorph1hr_2003090203.mpr
Map4=cmorph1hr_2003090204.mpr
Map5=cmorph1hr_2003090205.mpr
Map6=cmorph1hr_2003090206.mpr
Map7=cmorph1hr_2003090207.mpr
Map8=cmorph1hr_2003090208.mpr
Map9=cmorph1hr_2003090209.mpr
Map10=cmorph1hr_2003090210.mpr
Map11=cmorph1hr_2003090211.mpr
Map12=cmorph1hr_2003090212.mpr
Map13=cmorph1hr_2003090213.mpr
Map14=cmorph1hr_2003090214.mpr
Map15=cmorph1hr_2003090215.mpr
Map16=cmorph1hr_2003090216.mpr
Map17=cmorph1hr_2003090217.mpr
Map18=cmorph1hr_2003090218.mpr
Map19=cmorph1hr_2003090219.mpr
Map20=cmorph1hr_2003090220.mpr
Map21=cmorph1hr_2003090221.mpr
Map22=cmorph1hr_2003090222.mpr
Map23=cmorph1hr_2003090223.mpr
[MultiBandStat]
VarCov_Size=0
Correlation_Size=0
Mean_Size=1 1
Mean_Row0=0.000000
StandardDev_Size=1 1
StandardDev_Row0=0.000000
```

II. Process to aggregate the hourly images in the daily maplists to daily values in the ILWIS software.

```
cmd /c mpl_substitutename.bat ______%1 template_cmorph1.mpl > cmorph_day_%1.mpl
C:\maathuis\Ihvis372_July2014\ihvis.exe -C
cmorph_sum_%1.mpr:=MapMaplist.Statistics(E:\CMORPH\cmorph_day_%1.mpl,Sum)
```

Appendix F: HWSD soil texture lookup table (Xue & Hong, 2013)

Table 10-2 Look-up Table for HWSD Soil Texture

Code	Texture	Abbr.	Fild Capacity $\theta_{fc}(m^3/m^3)$	Permanent Wilting Point θ_{pw} (m^3/m^3)	Hydraulic conductivity K_{sat} (cm/h)
0	No_Soil	NS	0	0	0.00001
1	Clay(heavy)	CH	0.36	0.21	0.03
2	Silty Clay	SIC	0.36	0.21	0.05
3	Clay	C	0.36	0.21	0.075
4	Silty Clay Loam	SICL	0.34	0.19	0.1
5	Clay Loam	CL	0.34	0.21	0.1
6	Silt	SI	0.32	0.165	0.495
7	Silt Loam	SIL	0.3	0.15	0.65
8	Sandy Clay	SC	0.31	0.23	0.15
9	Loam	L	0.26	0.12	0.34
10	Sandy Clay Loam	SCL	0.33	0.175	0.15
11	Sandy Loam	SL	0.23	0.1	1.09
12	Loamy Sand	LS	0.14	0.06	2.99
13	Sand	S	0.12	0.04	11.78