

**RESPONSES OF
HYDROLOGICAL PROCESSES
TO LAND COVER CHANGE OF
AN ENDORHEIC TROPICAL
BASIN: THE CASE OF LAKE
NAIVASHA BASIN**

CHENYANG ZHANG

[March, 2014]

SUPERVISORS:

Dr. Ing. T.H.M. Rientjes

Dr. Ir. C. van der Tol



RESPONSES OF HYDROLOGICAL PROCESSES TO LAND COVER CHANGE OF AN ENDORHEIC TROPICAL BASIN: THE CASE OF LAKE NAIVASHA BASIN

CHENYANG ZHANG

Enschede, The Netherlands, [March, 2014]

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

Specialization: Water Resource and Environment Management

SUPERVISORS:

Dr. Ing. T.H.M. Rientjes (1st)

Dr. Ir. C. van der Tol (2nd)

THESIS ASSESSMENT BOARD:

Dr.C.M.M. Mannaerts (Chair)

Dr. L.G.J. Boerboom (External Examiner, ITC – PGM department)

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

In the past decades, the land cover in Malewa catchment in Kenya has been severely changed that mainly result from the increasing population and economic development. These changes might affect the water balance in Malewa catchment. This research is set for the purpose of examining the effect of land cover change on the stream flows (discharge) in one of the sub-basin in Malewa catchment. A hydrologic model (SWAT) is employed to simulate the discharge. The input parameters, such as precipitation, potential evapotranspiration (PET) and land cover, are integrated with remote sensing and GIS techniques.

The result of this study shows that some components, e.g. the surface runoff, peak runoff and base flow, are affected by the land cover change, especially deforestation. Specifically, the deforestation results in the significant increase of peak flows and surface runoff, but leads to decreases of base flow. The SWAT model performance is validated by the observed discharge data, which shows that $NS=0.69$, $R^2=0.71$, $RMSE=2.71m^3/s$, and $RVE=8.9\%$.

Key words: Land cover change, discharge. Malewa basin, SWAT.

ACKNOWLEDGEMENTS

First and for most I would like to thank for ITC who provided me the necessary resources which were needed to finish my research.

I was very indebted to my two qualified supervisors Dr. Ing. T.H.M. Rientjes and Dr. Ir. C. van der Tol. From both of their encouragement, guidance and advises can I went through the tough time for this thesis. My supervisors are patient and experienced with this research. Their critical comments and valuable guidance supported and leaded me during the whole time for this thesis. I learned a lot from them, I never forget their treatment.

Also, I am grateful to Ir. Arno Van Lieshout who is the WERM program director. When I had trouble with the research and cannot finish the thesis on time, he kindly help me to apply the one month extension. This gave me more time to focus on my research and finish it.

I also gratitude goes to my Advisor V. O. Odongo, a PHD student in ITC who provided me the data that I needed for this research and helps me a lot with the modelling.

I would like to express my gratitude to all officers and staffs in ITC for their support in different ways to my research.

I also would like to thank my family and friends who always supported me and encourage me to the last minute of the study.

Last but not least, I would like to thank my department WERM gave me a lot of basic knowledge of my study.

Thanks a lot for everyone and everything.

TABLE OF CONTENTS

1. Introduction.....	1
1.1. RESEARCH PROBLEM.....	2
1.2. RESEARCH OBJECTIVE.....	3
1.3. RESEARCH QUESTIONS.....	3
1.4. THESIS OUTLINE.....	3
2. LITERATURE REVIEW.....	5
2.1. PREVIOUS STUDIES.....	5
2.2. HYDROLOGIC MODELING.....	6
3. STUDY AREA AND DATA.....	8
3.1. THE STUDY AREA.....	8
3.2. THE DATA MATERIAL.....	11
4. METHODOLOGY.....	17
4.1. PROCESSING OF DATA.....	17
4.2. THE FRAME WORK OF THE RESEARCH.....	17
4.3. THE MODEL SWAT.....	18
4.4. MODEL CALIBRATION AND VALIDATION.....	21
5. DATA PROCESSING AND ANALYSIS.....	23
5.1. PRECIPITATION DATA AND DISCHARGE DATA.....	23
5.2. POTENTIAL EVAPOTRANSPIRATION (PET).....	32
6. MODELLING SET-UP.....	34
6.1. INTRODUCTION.....	34
6.2. BASIN DELINEATION.....	34
6.3. THE LAND USE AND SOIL TYPE DEFINITION.....	35
6.4. THE INPUT FILE OF RAINFALL AND PET.....	37
7. RESULT AND DISCUSSION.....	39
7.1. LAND COVER CHANGE DETECTION.....	39
7.2. PET DATA ANALYSIS.....	41
7.3. RAINFALL AND DISCHARGE DATA IN DIFFERENT YEARS.....	44
7.4. THE MODELLING RESULT.....	46
8. CONCLUSIONS AND RECOMMENDATIONS.....	59
8.1. CONCLUSIONS.....	59
8.2. RECOMMENDATION.....	60
Appendix.....	63
APPENDIX 1.....	63
APPENDIX 2.....	70
APPENDIX 3.....	73

LIST OF FIGURES

Figure 1: The water cycle.....	1
Figure 2: The level of Lake Naivasha over the past 130 years (reconstructed from several sources) (Becht & Harper, 2002).....	2
Figure 3: Map of the Lake Naivasha basin(Meins, 2013).....	8
Figure 4: The Malewa catchment. The numbers 1 to 7 indicate the subbasins.....	10
Figure 5: The original land cover map for three years.....	15
Figure 6: The reclassification results of three land cover maps.....	16
Figure 7: The flowchart of the research method.....	18
Figure 8: The soil water storage schematization in the SWAT model (SWAT manual (Neitsch, 2002b))	19
Figure 9: The shallow aquifer in SWAT model (SWAT manual (Neitsch, 2002b))	20
Figure 10: The deep aquifer in SWAT model (SWAT model manual(Neitsch, 2002b))	20
Figure 11: Double-mass curve of station2GB01 vs. rainfall stations.....	24
Figure 12: Observed discharge vs. rainfall station.....	25
Figure 13: The three parts of catchment	27
Figure 14: Results of double mass curve analysis.....	28
Figure 15: The results of seven subbasins for $ \Delta P /\Delta Q$ ration during 2007/02/01-2010/02/01	32
Figure 16: The steps in ARCGIS software to achieve the PET data.....	33
Figure 17: The ARCSWAT result after the basin delineation.....	35
Figure 18: The reclassification result of land cover map in ARCSWAT.....	36
Figure 19: The reclassification result of soil map in ARCSWAT	36
Figure 20: The reclassification result of slop in ARCSWAT.....	37
Figure 21: The location of rainfall stations in the ARCSWAT model.....	38
Figure 22: The figure of land cover change in year 1973, 1986, 2011	40
Figure 23: The comparison for the MODIS data and weather station data.....	41
Figure 24: The consistent of MODIS16 and weather station.....	42
Figure 25: The PET from different land cover	43
Figure 26: The rainfall and runoff comparison for the wet season in different years.....	45
Figure 27: The first running result for ARCSWAT in subbasin 5.....	46
Figure 28: The effect of SOL_AWC.....	51
Figure 29: The effect of SOL_K.....	52
Figure 30: The effect of SOL_Z.....	53
Figure 31: The result of model calibration and validation.....	54
Figure 32: The comparison of discharge for different land cover simulation result.....	55
Figure 33: The comparison of base flow for different land cover simulation result.....	55

Figure 34: The comparison of surface runoff for different land cover simulation result	56
Figure 35: The comparison of two scenarios	57
Figure 36: The comparison of base flow for two scenarios	57
Figure 37: The comparison of surface runoff for two scenarios.....	58

LIST OF TABLES

Table 1: Land cover and use change research on other catchment (This study)	5
Table 2: Advantages and disadvantages of 6 different hydrological models(Meins, 2013)	7
Table 3: Comparison of the five different models with respect to hydrological processes(Deelstra et al., 2010).....	7
Table 4: The location of rainfall stations	12
Table 5: The location and elevation of the river gauging stations	13
Table 6: The river gauge stations & the rainfall stations.....	23
Table 7: The six input rainfall stations	29
Table 8: The average weight in each subbasin for each rainfall station.....	30
Table 9: The land cover table for year 1973, 1986, 2011 (km ²)	39
Table 10: The land cover table for year 1973, 1986, 2011 (%)	40
Table 11: The land cover for subbasin5 in year 1973, 1986, 2011	41
Table 12: The total rainfall, runoff and runoff coefficient	45
Table 13: The parameters were chosen for this research.....	47
Table 14: The test value used for the GW parameters.....	48
Table 15: The value for the three soil parameters in this research.....	48
Table 16: The results for the chosen parameters: L means the depth of layer, K means saturated hydraulic conductivity.....	49

1. INTRODUCTION

Water is indispensable resource for life. Water cycle, also called hydrologic cycle, is one of the big systems on the earth. It contains many components such as: precipitation, evapotranspiration, infiltration, stream flow, ground water flow, surface runoff (Figure 1). Some interactions between these components, such as runoff, evapotranspiration and infiltration, are affected by land cover change of the catchment (see Figure 1). Evapotranspiration in this respect includes evaporation of water from interception of the vegetation and transpiration as well as the water stored in the soil.

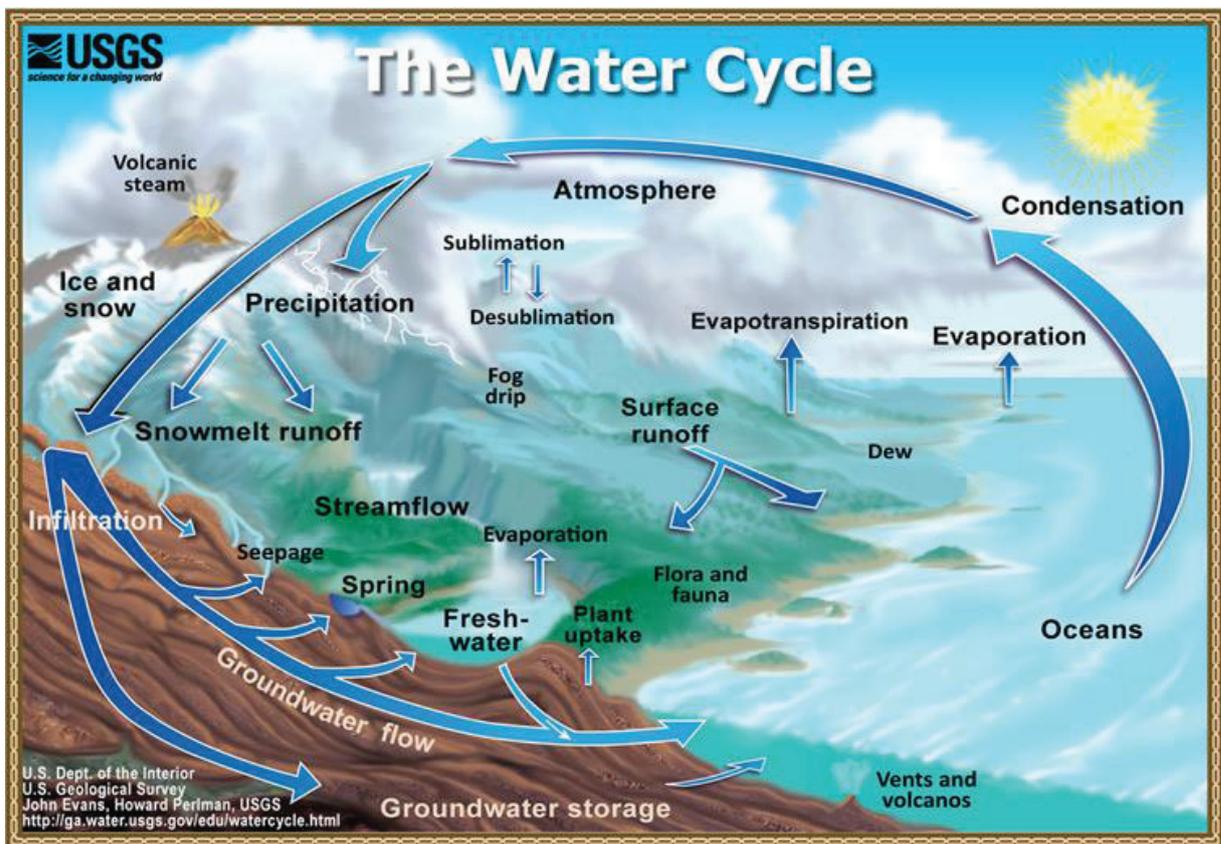


Figure 1: The water cycle

The Lake Naivasha basin is the second RAMSAR in Kenya, which is an endorheic tropical basin. This lake is very important because it serves as a major source of water for domestic uses and for irrigation, agriculture, fishing and tourism. In recent years, with the development of the economy and society, water demands to support irrigated horticulture largely increased. Moreover, a rising population in response to the increasing employment opportunities lead to (further) land cover change with deforestation and increased build-up area. These changes affect the hydrological cycle where increased economic activities raises the demand of water resources (van Oel, Mulatu et al. 2013). In the past 130 years, Lake Naivasha water levels show a

decreasing trend as indicated in Figure 2 (Becht & Harper, 2002). In view of expected and extended horticultural and industries activities it is beneficial for a sustainable development to assess impacts by land cover changes. Therefore, the research after hydrologic processes by land cover change in the Lake Naivasha catchment is necessary and effective to help the government to develop a water management plan.



Figure 2: The level of Lake Naivasha over the past 130 years (reconstructed from several sources) (Becht & Harper, 2002).

1.1. RESEARCH PROBLEM

As mentioned before, the lake Naivasha basin is facing a serious threat. The water level in the lake shows a decreasing trend in the past years. A number of studies investigated the impact of different aspects that may affect the water level such as human impact (Becht & Harper, 2002), and the effects of inflows by groundwater and surface water (van Oel et al., 2013). In Awange et al. (2013) it is described that future studies on the decline of water level of lake Naivasha basin should not only take into account the lake area but also include the fluctuations of the Malewa and Gilgil river systems. This since both have major contributions to the lake. Especially the Malewa drains approximately 80% of the surface inflows to the lake. The land cover and land use of the Lake Naivasha has changed for the past decades. How land cover and land use changes have affected the hydrologic processes is not clear in this catchment. Consequently, there is a strong need for hydrologic assessment on the influence of land cover and land use changes in a catchment hydrological behaviour so to better understand the hydrologic cycle and related water balance. Also, it is expected that water resource management in the basin can be improved.

The main problem to be researched is that hydrological impacts by land cover and land use changes in Malewa catchment are unknown. In this research, impact assessments aim at runoff that is produced in the catchment. Impact assessment is by hydrological modelling and use of satellite remote sensing.

1.2. RESEARCH OBJECTIVE

1.2.1. MAIN OBJECTIVE

The main objective of the research is to improve an understanding of the impact of land cover and land use changes on surface runoff, base flow, peak runoff and runoff volume on Malewa catchment through hydrological modelling.

1.2.2. SPECIFIC OBJECTIVE

The specific objectives defined are:

- (1) Analysis the rainfall and discharge data for the three years which had the land cover map and identify the trends of rainfall and discharge data of Malewa catchment.
- (2) Select rainfall and discharge data for calibration and validation by means of time series analysis.
- (3) Identify the differences in land cover maps of the study area to assess land cover changes.
- (4) Develop and parameterize a hydrologic model for Malewa catchment.
- (5) To assess effect of land cover and use change on:
 - Base flow
 - Peak runoff
 - Surface runoff
 - Runoff volume

1.3. RESEARCH QUESTIONS

- (1) Is there an increasing or decreasing trend in rainfall and discharge measurements in Malewa catchment since 1970?
- (2) How well do stream flow records relate to rainfall records?
- (3) How has land cover and land use changed in Malewa catchment since 1970?
- (4) How did the land cover and land use change in Malewa catchment affect the base flow, surface runoff, peak flow?

1.4. THESIS OUTLINE

The main chapters contains in this thesis are described as follows:

Chapter 1 gives an introduction to this study and describes the problem statement, research objectives and research questions. A literature review land cover and use changes is presented in Chapter 2. Results on impact assessments for various study areas are described in terms of base flow, surface runoff and evaporation. Also a brief description on hydrologic modelling is added in the same Chapter. In Chapter 3 the study area and available data are described, for example the location of the Malewa catchment, the environment of the Malewa catchment and the source of rainfall data, discharge data, PET and land cover maps. Chapter 4 introduces the methodology of the research includes the data processing, the flow chart of the research, and gives a brief description of the SWAT model that is used in this study. The detail methods for data processing and analysis of the processing result is described in Chapter 5. The Chapter 6 gives a brief introduction of the modelling setup step by step. In the Chapter 7, the results of this research are analysed and discussed which includes the land cover change detection, PET data analysis, the compare of the rainfall data and discharge data in year 1973, 1986 and 2011 and the modelling results. The Chapter 8 is the last chapter that gives the conclusions and recommendations of this research. Conclusions give the summary of the thesis and the recommendations give some advices for further studies.

2. LITERATURE REVIEW

2.1. PREVIOUS STUDIES

Land cover changes attracted researchers in recent years because of its influence on the hydrological processes of the water cycle. To name a few, Morán-Tejeda et al. (2012) found that changes in land cover affected the total water yield but also runoff production and stream flow distribution.. Moreover, results suggested that further increases in forest area would cause hydrological decline and highlights the significance of integrating land-cover information in water availability assessments. In addition, the land cover affects the annual evapotranspiration (ET) (Zhang et al., 2012). It was concluded that deforestation caused a decrease in the annual ET and that shrubland to forest transformation caused an increase of ET. The table below shows the research about land cover/use change in other catchments.

Table 1: Land cover and use change research on other catchment (This study)

Study area	land cover change	runoff	ET	Temperature	rainfall	Soil moisture
Duero basin(Spain) (1)	reforestation	-	no	+	no	no
Duoyingping(China) (2)	deforestation	+	-	no	no	+
	reforestation	-	+	no	no	-
Blue Nile Basin(Ethiopia) (3)	deforestation	low flow -	no	no	- & +	no
		high flow +				
		mean annual flow -				
Miyun reservoir(China) (4)	afforestation	-	no	+	-	no
Laohahe catchment(China) (5)	decreasing in grassland and water body	-	+	no	no	no
	increasing in forest land crop land					
Rhine basin(French-german) (6)	afforestation	lower peak runoff +	no	no	no	no
	increasing in urbanisation	peak runoff&runoff volume -				
Upper Du watershed(China) (7)	farm land increase	runoff +	no	no	no	no
	forest increase	runoff -				
	urban area increase	runoff +				

Where (+) means increase, (-) means decrease, no means not studied, & means both (increase in one season& decrease in another season).

(1)(Morán-Tejeda et al., 2012); (2)(Zhang et al., 2012); (3)(Rientjes et al., 2011); (4)(Tang et al., 2011); (5)(Liu et al., 2008);(6)(Hundechea & Bárdossy, 2004); (7)(Yan et al., 2013)

For Malewa catchment, it is often stated that the most significant part of the change in land cover is by deforestation. However, the effect of forest change on runoff is not clear in the table because results don't agree well for the various studies. For example, in the Blue Nile Basin in Ethiopia deforestation causes the decrease in low flow but increase in high flow. The same condition of the land cover changes does not give the same result in Duoyinping catchment in China.

2.2. HYDROLOGIC MODELING

2.2.1. INTRODUCTION

Hydrology and water resource engineering are important in people`s daily life and the environment. It plays significant role in society. Because of the huge effects and importance of the hydrology, human beings have been trying to understand and make the control of the water cycle processes for a long time. They found that the water storage is easily affected by the regional hydrological processes. And a hydrologic model can show the detail of the relations between these various components such as land cover, climate, and human impacts. Based on these relationships, a model can make a prediction and forecast of the future. It can be traced back to the halfway the nineteenth century that the first model can predicted the runoff from rainfall was developed (Mulvaney, 1851).

In order to make an understanding of hydrologic catchment behaviour, people created models to simulate them. The purposes of models vary depending on the interests of the research. Thus, some basic understanding of the system components or characters is required for any research. Simulation models are applied to make assessments of sensitivity and changes in hydrological processes by changes in the system such as human activities as well as land cover and use changes.

2.2.2. HYDROLOGICAL MODELS

A basin catchment is abstracted for modelling and for analysis by limitation of the global methods to assess hydrologic behaviour. Hydrologic processes may change by climatic impacts, human activities, land use changes, other natural changes (crustal movement, volcanic movement etc.). The links between water cycle processes and influencing factors are conceptualized and investigated by hydrological modelling frame works.

Hydrologic models serve to represent the real world. Rainfall-runoff(R-R) models aim to simulate the R-R transformation which includes many core processes of hydrology. Few hydrological models suitable for rainfall-runoff simulation are HBV and SWAT (Table 3). Both models are water balance based, continuous, stream flow models which simulate base flow, peak flows, surface and subsurface flows. Both models therefore are suitable candidates for application in this study. In Table 2, several models are described and compared with the advantages and disadvantages identified. For instance, the SWB model ignores small scale processes, Pitman model which is not well known, WEAP model which is not focus on the rainfall-runoff modelling and not available, and GR4J model which need the background of MATLAB. The SWAT model has the advantage that changes in land cover can directly be imposed and therefore, for this research, the SWAT model is more suitable than HBV and is described in Section 4.

Table 2: Advantages and disadvantages of 6 different hydrological models (Meins, 2013)

Model Name	Advantages	Disadvantages
SWB	+ Easy to understand and implement + Requires almost no computational time + Data at lumped scale is available	- Small scale processes are omitted - No information on the spatial distribution of water in the basin
HBV	+ The model structure is easy to understand + Requires little computational time + Model is freely available	- Requires manual basin delineation when dividing a larger basins in sub-basin - Interflows between basins are not accounted for
SWAT	+ Allows for several different basin delineations at different scales + Can easily incorporate changes in land use and land cover + The model is open source and is freely available + Multiple calibration methods are readily available	- Requires more computational time, especially during calibration - Requires more data (suggesting that more assumptions will have to be made)
Pitman	+ Has been successfully applied to semi-arid African basins before + The model structure is easy to understand	- Requires calibration of 24 parameters - The model is not well known
WEAP21	+ Can incorporate the effects of water management decisions easily	- Does not focus on rainfall-runoff modelling - The model is licensed and will be difficult to obtain
GR4J	+ The model structure is easy to understand + Requires calibration of only 4 parameters + Requires little computational time + The model code is readily available in MATLAB	- Requires manual basin delineation when dividing a larger basins in sub-basin

Table 3: Comparison of the five different models with respect to hydrological processes (Deelstra et al., 2010)

Model layer	Processes	DrainMod	Coup	HBV	INCA	SWAT
Above ground vegetation zone	Precipitation	Driving	Driving	Driving	Driving	Driving
	Snow dynamics/snowmelt	Calculated	Calculated	Calculated	Calculated	Calculated
Soil surface	Evapotranspiration	Calculated	Calculated	Calculated	Indirectly	Calculated
	Surface runoff	Calculated	Calculated	Calculated	Calculated	Indirectly
	Infiltration	Calculated	Calculated	Indirectly	Indirectly	Indirectly
	Bypass/ macropore flow	NO	Calculated	Indirectly	NO	Calculated
	Plant water uptake	Calculated	Calculated	Indirectly	Indirectly	Calculated
	Soil water redistribution	Calculated	Calculated	Calculated	NO	Uniform
	Water flow in frozen soil	Calculated	Calculated	Calculated	NO	at saturation
	Lateral flow to stream	NO	NO	Calculated	Calculated	Calculated
	Subsurface drainage flow	Calculated	Calculated	NO	Indirectly	Indirectly
	Percolation to sat. zone	Calculated	Calculated	Calculated	Calculated	Calculated
Unsaturated zone	Lateral inflow	Parameter	Parameter	NO	NO	NO
	Capillary rise to unsat. zone	Calculated	Calculated	Calculated	NO	Indirectly
	Recharge to deep aquifer	NO	NO	NO	NO	Calculated
	Base flow	NO	NO	Calculated	Calculated	Calculated

3. STUDY AREA AND DATA

3.1. THE STUDY AREA

This chapter describes the Lake Naivasha basin. Though the main research area is the Malewa river catchment, an overview of the whole Naivasha basin is made to better illustrate the research area (See Figure 3).

3.1.1. THE LAKE NAIVASHA BASIN

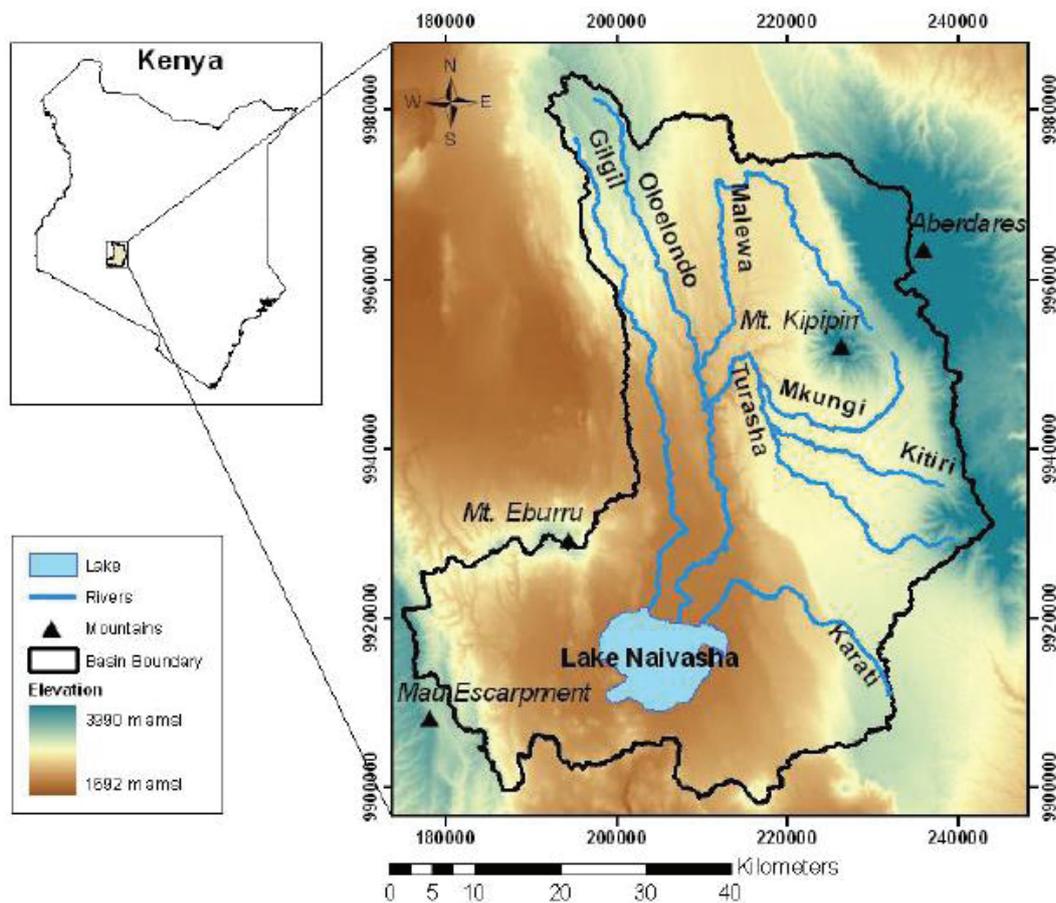


Figure 3: Map of the Lake Naivasha basin (Meins, 2013)

Lake Naivasha is located in the bottom of the Kenyan Rift Valley which is almost 80km north-west from the Kenya's capital, Nairobi. The Kenyan rift valley makes part of the East African Rift Valley. Many volcanoes are located in the research area. Lake Naivasha basin is located to the north of the lake which has elevation range from 1881 to 3989 m a.m.s.l. The elevation of Lake Naivasha is approximately 1890m above the mean sea level (a.m.s.l.) and is the highest lake of the East African Rift lakes. The lake area is

approximately 140km² which is smaller than the largest Lake Turkana in Kenya and become the second-largest fresh water lake in Kenya. (Stoof-Leichsenring et al., 2011)

The large topographic variation in the basin contributes to pronounce spatially variability of rainfall. For example, the average rainfall in the lake Naivasha is approximately 670mm, while in the Aberdare Mountains at higher altitude it can increase to 1350mm. Moreover, the Naivasha basin has two rainy seasons, one is from March to May called 'long rains' and the other is from October to November called 'short rains'. In the Figure 3, it is shown that the basin area is on the west of the Aberdare Mountains and east of the Mau Escarpemnt as well as south of Eburru Hills.

The climatic variation in the lake Naivasha basin is mainly because of the altitude and land forms. The climate at lake Naivasha is semi-arid and the upper basin is humid (R. Becht et al., 2011). The potential evaporation in the lake Naivasha is almost 1360 mm, and suggests that the runoff from the non-immediate catchment results in the water level of Lake Naivasha.

For the inflow of lake Naivasha, 80% of it is contributed by surface flow and 20% is from subsurface flow (Gaudet & Melack, 1981). The surface flow is mainly fed by two perennial river systems named Gilgil and Malewa respectively in the North (Meins, 2013). The origination of both rivers is larger than 2500m a.m.s.l. (Becht & Harper, 2002). The Malewa River offer almost 80% and the Gilgil River contains 10% and other streams from Karati and seasonal streams (Abiya, 1996).

The soils in the lake Naivasha basin are mainly developed by the high volcanic activity in the past years (Becht & Harper, 2002). This cause the higher permeable of water, in other word, the capacity of holding water is weak in the soils which especially happened around the lake. It is suggested that the surface flow is mainly caused by the rain in upper basin and the base flow is the only flow which is added in the lower area (Meins, 2013).

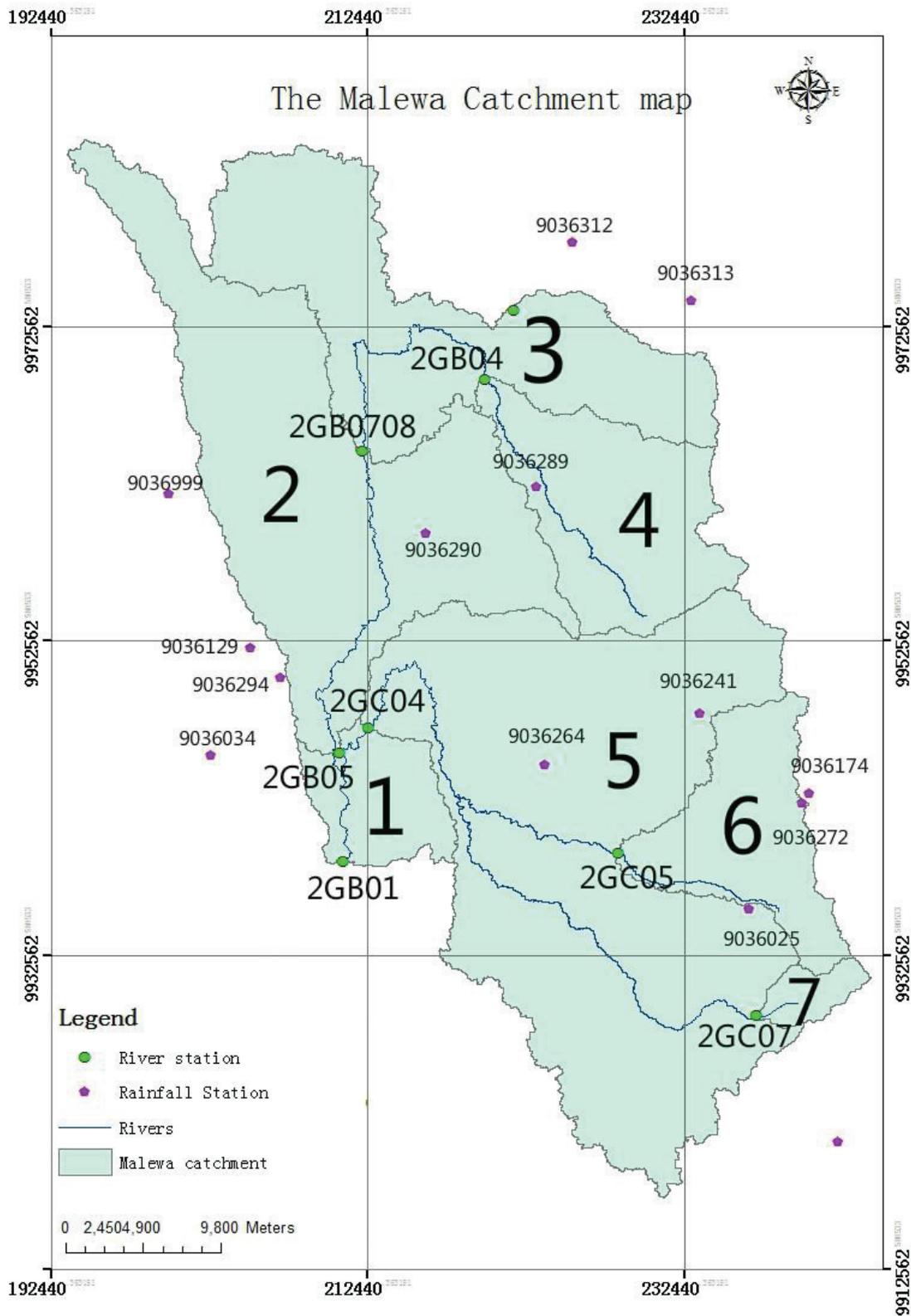


Figure 4: The Malewa catchment. The numbers 1 to 7 indicate the subbasins.

3.1.2. THE MALEWA CATCHMENT

The Malewa catchment is located in the Kenya Rift valley which is approximately 70km from Nairobi. The location of the Malewa is between latitude 0°09 to 0°55 south and longitude 36°09 to 36°24 east. Most of the basin lies within the Kinangop Plateau and has the latitude value between 2100-2700m. The highest point in the basin is about 3990 meter and the lowest is about 1980 meter above the mean sea level. The River Malewa catchment covers almost 1700 km². The rivers within this catchment are consisted by the main river Malewa and its tributaries such as the Turasha, Mkungi, and Kitiri. The vegetation in this area can be divided into 5 classes after reclassification of the original land cover maps: Forest, Grassland, Farm land, Shrub land and Wood land.

3.2. THE DATA MATERIAL

3.2.1. INTRODUCTION

The data used in this study is prepared and made available through the work by Meins (2013) with topic looked at “Evaluation of spatial scale alternatives for hydrological modelling of the lake Naivasha basin, Kenya”. The data is related to the lake side environment. However, the data of the surrounding area is obtained in the research as well. Although some stations are not located within the Malewa catchment, the distances of these stations are not far away from the research area. These stations also used in this research. In his research, much effort had to be made to fill in the gaps of the data.

3.2.2. THE HYDROLOGICAL DATA

3.2.2.1. Rainfall data

The rainfall data is from the Kenyan Meteorological Department (KMD). The data was collected in 2004 and contains data for 65 stations for the total area of Lake Naivasha basin. However, in this research, the main stations which are located within and close to the catchment were utilized. The locations of rainfall stations are shown in Figure 4 and elevation of the rainfall stations is shown the Table 4. As described before the data gaps has been filled by Meins (2013).

Table 4: The location of rainfall stations

station ID	X	Y	Elevation (a. m. s. l)
9036290	216155.3	9959398.7	2332
9036289	223136	9962352	2437
9036129	205018.6	9952093.3	2234
9036294	206913	9959398.7	2170
9036034	202500	9945300	1995
9036264	223622.8	9944687.8	2403
9036999	199826	9961909	2391
9036241	233394	9947946	2588
9036025	236545.6	9935511.5	2617
9036174	240329.2	9942813.5	3125
9036272	239876	9942226	3191
9036164	242120.7	9920692	2541
9036312	225396.1	9977875.4	2797
9036313	232858.9	9974226.1	3274

3.2.2.2. The potential Evapotranspiration (PET)

The data for calculating the potential evapotranspiration is limited. Since most of the weather stations in the Naivasha basin are close to the Naivasha Lake. The weather stations near the lake will be influenced by the water in the lake, which may cause the PET data affected. Besides, the weather stations in the Malewa catchment do not have sufficient data from year 2007-2010 (the time series has been chosen in this research). Thus, the PET data have been achieved through the remote sensing data (MODIS 16 products). The detail steps to achieve the data would be explained in section 5.2.

3.2.2.3. The stream flow data

The stream flow data were provided by observers who recorded the water levels (m) at the various gauge sites in the basin. As mentioned before, the data were processed by Meins (2013) and the water levels were converted to the discharge data (m^3/s) already. The gaps in the data were interpolated for the period 1960 to 2010. The locations of the river gauging stations in the Malewa catchment are shown in the Figure 4 and the elevation of these locations are described in Table 5.

Table 5: The location and elevation of the river gauging stations

Station	x-coordinate	y-coordinate	Elevation(m)
2GB01	210908	9938530.8	1951
2GB03	221632.3	9973620.2	2366
2GB04	219808.8	9969175.2	2334
2GB05	210688.5	9945446	1987
2GB0708	212081.6	9964640.5	2264
2GC04	212451.6	9964983.4	2000
2GC05	228295.2	9939060.7	2408
2GC07	236916.6	9928708.5	2708
2GC10	225447.7	9942224.9	2419

3.2.3. THE GEOGRAPHIC DATA

3.2.3.1. The GDEM data

For the study area, a Global Digital Elevation model (GDEM) was developed by the U.S. National Aeronautics and Space Administration (NASA) and Japan's Ministry of Economy, Trade, and Industry (METI) which is freely available in the internet (<http://asterweb.jpl.nasa.gov/gdem.asp>). The DEM uses the $1^{\circ} \times 1^{\circ}$ tiles. The Naivasha basin is covered by the 1°S and 36°E tile. For this research, the DEM data was pre-processed in the previous research and had resolution of $30\text{m} \times 30\text{m}$.

3.2.3.2. Land cover/land use maps

A land cover/land use (LULC) map was produced from a Landsat MSS image of the 21st of December 1973 by Odongo (2012). This map can represent the period before 1985. At that time, large scale flower farming was not practiced in the basin. For this study, LULC maps were also available for the years 1986 and 2011. These maps were utilized to make the definition of the different types of vegetation in the Lake Naivasha basin. All maps came from the remote sensing images and already processed with the classification depend on the different land covers. The accuracy of the maps are 0.73.

The original land cover maps are shown in Figure 5. When comparing these three maps, it is obvious that the forest area decreased from 1973 to 1986 as well as the bush land. The increase for the grass land is not so obvious in the maps, but the attribute tables for these three maps indicated that the grass land was increase from 1973 to 2011. The grey colour in these maps means this area maybe mixed with bush land, grassland and some other land cover classes. It also can be noticed that for the years 1973 to 2011, different land cover classes and classification are used. For example, in 1973 there are only 7 land cover classes in the map, while in 1986 the number of land cover classes increased to 8. The fallow land was the new land cover

class in 1986. Also, in 2011, there are 10 land cover classes in the map. The built up, horticulture were the new land cover classes when compare to the land cover map 1986.

Because there are too many land cover classes in difference maps. This could cause a confusion of the land cover effect with the water cycle. When do the comparison, all the land cover classes in each map should be the same. For these reasons, the land cover maps were reclassified by combining the classes might have the same reflectance for the solar radiation. When combing the bush land and fallow land to shrubland, and combing the built up and horticulture to farm land, the final results of the land cover reclassification were 6 classes: farmland, shrubland, forest, woodland, grassland and water. (See Figure 6)

3.2.3.3. Soil map

The soil map was provided by the Kenya Soil Survey (Sombroek, 1982). This map was used to determine the soil group of the Lake Naivasha basin.

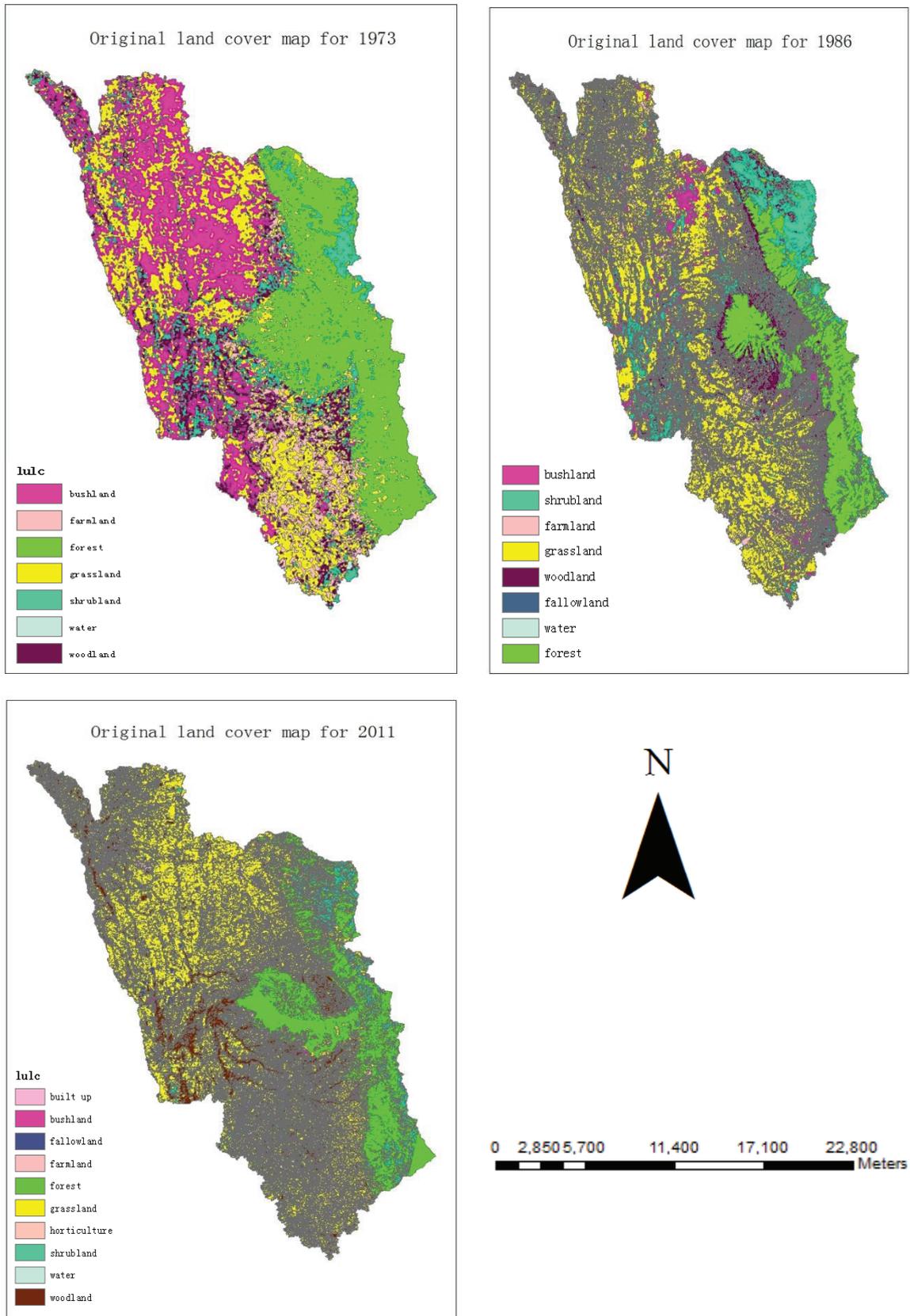


Figure 5: The original land cover map for three years

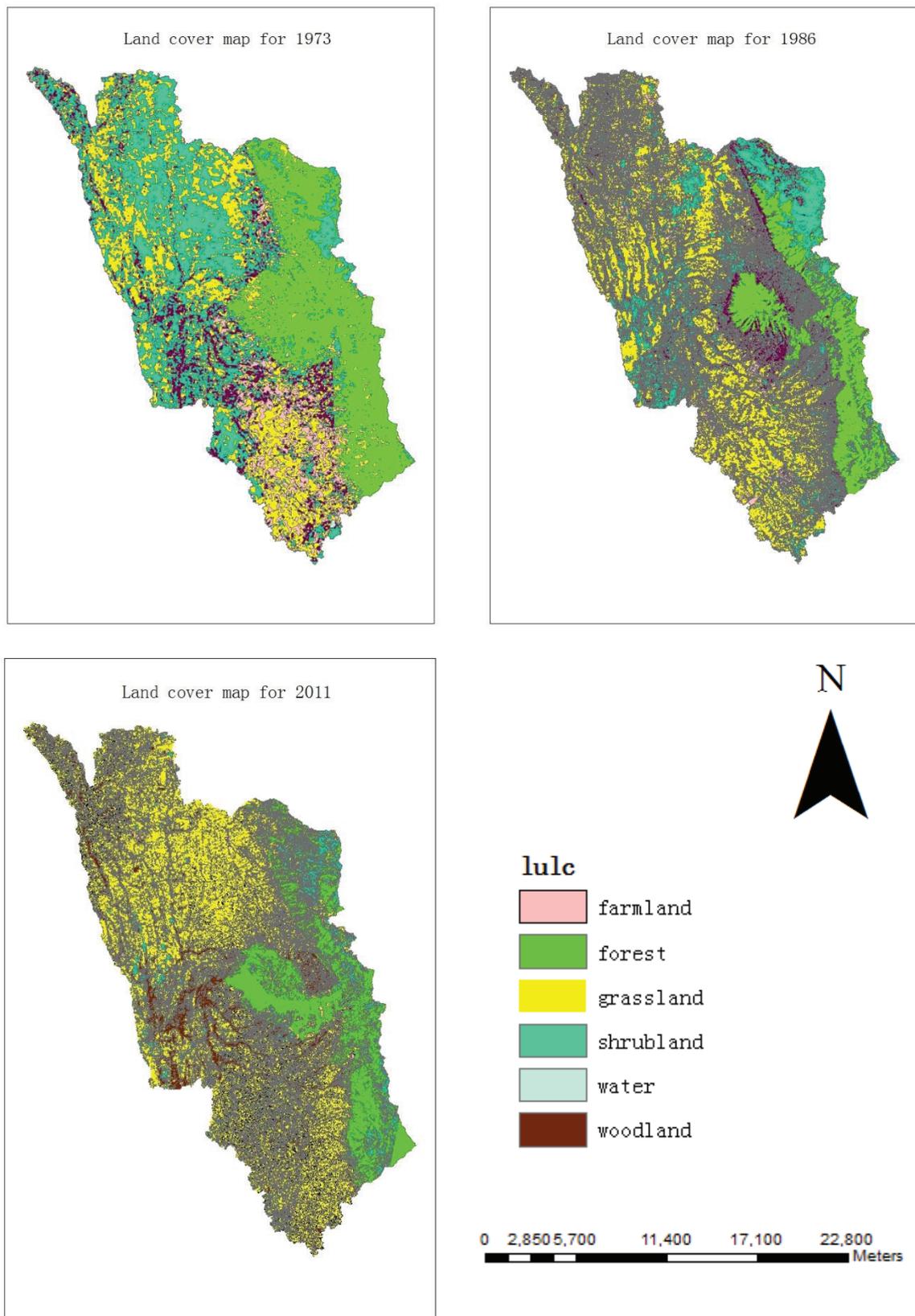


Figure 6: The reclassification results of three land cover maps

4. METHODOLOGY

4.1. PROCESSING OF DATA

The in-situ data which includes observed rainfall, and observed river discharge should be collected in this research. The discharge data was collected from the river stations at the outlets of the streams for this study area. The rainfall data came from the rainfall stations maintained by the government or research organizations in the Malewa catchment. The rainfall data and discharge data was analysed for the consistency. Results showed that rainfall data did not well relate to the discharge data. Because the data of rainfall was from the rainfall station, which means the data was coming from different locations of the area, it cannot present the data of the whole catchment. The interpolation of rainfall data from one point to area was an important step in the research. The source of the rainfall and discharge data are described in section 3.2.2. Results of data processing and analysis are shown and described in section 5.1. Also the geographic data for soils, land cover, and elevation were collected and prepared for this research. Available geographic data is described in section 3.2.3.

The potential evapotranspiration (PET) data is one of the most important inputs for this research. Whereas PET data cannot directly collected from the observation data in the weather stations, PET data was extracted from the remote sensing data from MODIS product 16. The source of PET is described in section 3.2.2. The processing steps of PET is described in section 5.2.

4.2. THE FRAME WORK OF THE RESEARCH

The Figure 7 shows the frame work of this research. In this figure, the main procedure and methodology applied in this research are illustrated.

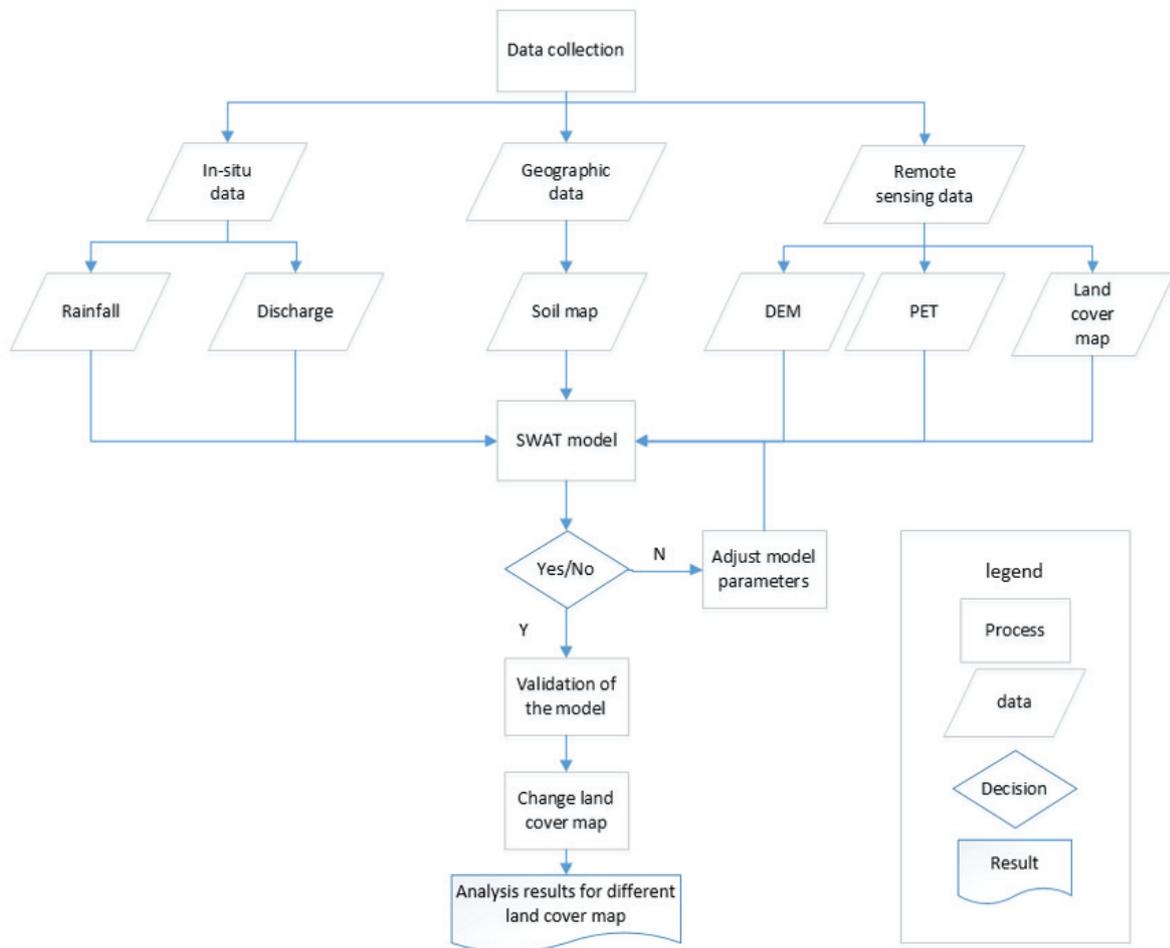


Figure 7: The flowchart of the research method

4.3. THE MODEL SWAT

4.3.1. INTRODUCTION

Soil and Water Assessment Tool (SWAT) was developed by Dr. Jeff Arnold for the USDA Agricultural Research Service (ARS). The model was developed to address the impact of land management practice on water, sediment and water quality by considering different soil types, land covers, and management conditions. This model relies on inputs on specific information that will affect the water response which is associated with the weather, soils, vegetation, and slope. Besides, the relative influence of alternative input data such as land cover, climate, human activities, can be quantified (Neitsch, 2002a).

4.3.2. THE BRIEF DESCRIPTION OF THE MODEL

The water balance equation used in soil profile in the SWAT model reads:

$$SW_t = SW_0 + \sum_{i=1}^t (R_{day} - Q_{surf} - E_a - w_{seep} - Q_{gw}) \quad (1)$$

SW_t = final soil water content (mm)

SW_0 = initial soil water content on day i (mm)

R_{day} = amount of rainfall on day i (mm)

Q_{surf} = amount of surface runoff on day i (mm)

E_a = amount of evapotranspiration on day i (mm)

w_{seep} = amount of water entering the deep aquifer from the soli profile on day i (mm)

Q_{gw} = amount of return flow on day i (mm)

Considering the climate of Malewa catchment, the surface runoff is only by rainfall and thus snowmelt is not considered. When rainfall infiltrates the soil, one part will become lateral flow and stay in the layer of the soil while the other part will percolate to the shallow and deep aquifers. However, if the soil layer stores the water until the soil field capacity is reached, the upper layer will become saturated. The infiltration and percolation of the rainfall will stop when the soil is saturated and overland flow is generated. (See Figure 8).

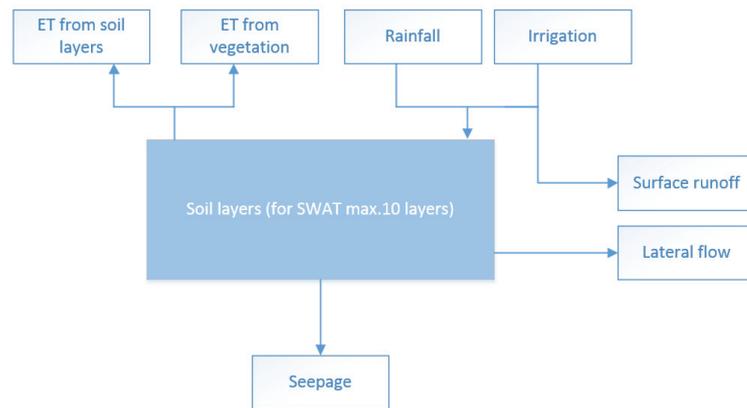


Figure 8: The soil water storage schematization in the SWAT model (SWAT manual (Neitsch, 2002b))

The water balance schematization of the shallow aquifer is show in Figure 9:

$$\frac{\Delta V_{aq,sh}}{\Delta t} = W_{rchr,sh} - Q_{gw} - W_{revap} - W_{pump,sh} \quad (2)$$

$\frac{\Delta V_{aq,sh}}{\Delta t}$ = the change of water content in the shallow aquifer divided by time step

$W_{rchr,sh}$ = the amount of water entered to the shallow aquifer (from w_{seep} and surface runoff losses during the transport)

Q_{gw} = the outflow of ground water to the channel (base flow)

W_{revap} = capillary rise

$W_{pump,sh}$ = pumping water from shallow aquifer

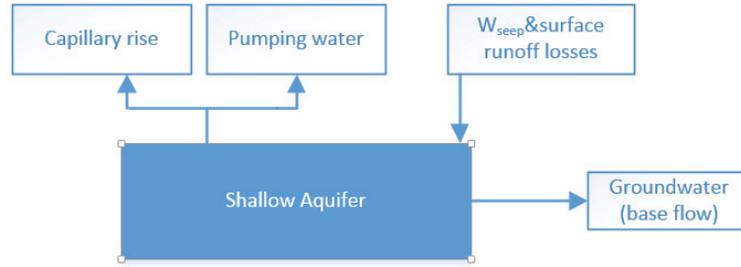


Figure 9: The shallow aquifer in SWAT model (SWAT manual (Neitsch, 2002b))

And the water balance of the deep aquifer:

$$\frac{\Delta V_{aq,dp}}{\Delta t} = W_{deep} - W_{pump,dp} \quad (3)$$

$\frac{\Delta V_{aq,dp}}{\Delta t}$ = the change of water content in deep aquifer over the time step

W_{deep} = the amount of water to the deep aquifer through percolation (provide by w_{seep})

$W_{pump,dp}$ = the amount of abstracted pump water from deep aquifer (Irrigation)

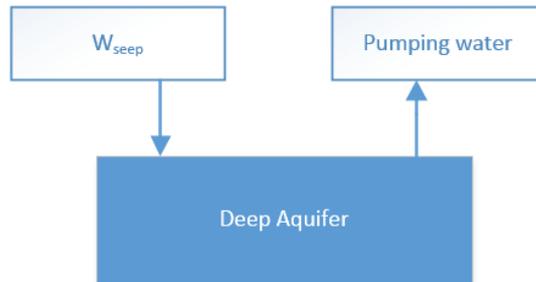


Figure 10: The deep aquifer in SWAT model (SWAT model manual(Neitsch, 2002b))

Water routing through the channel is simulated following Eq. (4) and reads:

$$\frac{\Delta V_{ch}}{\Delta t} = V_{in} - V_{out} - tloss - E_{ch} \pm div + V_{bnk} \quad (4)$$

$\frac{\Delta V_{ch}}{\Delta t}$ = the change of amount of water in the channel divided by time step

V_{in} = the water volume coming from upstream catchment and land phase

V_{out} = the amount of water flowing to the downstream catchment

$tloss$ = the water amount loss when the flow transmission through the channel

E_{ch} = the evaporation from the channel

div = the diversion of water (irrigation or dam)

V_{bnk} = the water stored in the bank

4.3.3. THE HRU OF SWAT MODEL

The SWAT model is a distributed model with spatial model domain based on Hydrology Response Units (HRU). Such units are created to adopt to the different conditions in a catchment by soil types, land cover, topography and climatic. By HRU-based catchment discretization, a catchment can be divided into a large number of HRUs thus making each HUR a unique spatial entity based on local conditions for soil type, land cover, slope, and climate. Environment condition in one HRU are assumed and considered homogenous. Then in the model SWAT, to each HRU the same model algorithms are applied to generate surface runoff, base flow and related hydrological processes. The routing equation is only applied on the subbasin output. The results calculated from equation 1, 2, 3 for all the HRUs in the subbasin are considered as the input of the equation 4 and converted from mm to m³. The only difference between the HRUs are the parameters used in the equation that reflect on the different local conditions (Meins, 2013). HRU is often considered to be the best choice to analysis for catchments with complex conditions.

4.4. MODEL CALIBRATION AND VALIDATION

Model calibration serves to fine tune the model so that simulated data matches observed data. In general, the first simulation result does not represent the 'real world' well since model parameters require optimization. The adjustment and optimization of the model input parameters aims to improve simulation results. Accurate fine-tuning may improve model reliability.

Model performance commonly is evaluated by use of objective functions which serve model calibration purposes. To validate the performance of a model, a second independent time series with model forcing data (precipitation and PET) was used as model input data. This data is called validation data. The result of model validation is also evaluated by objective functions.

In this research, the objective functions used for model calibration and validation are Nash-Sutcliffe (NS), coefficient of determination (R²), Root mean square error (RMSE) and the Relative Volume Error (RVE). The equations are shown below:

$$NS = 1 - \frac{F}{F_0} \quad (5)$$

Where

$$F = \sum_{n=1}^N (Q_{obs_n} - Q_{sim_n})^2 \quad (6)$$

$$F_0 = \sum_{n=1}^N (Q_{obs_n} - \overline{Q_{obs}})^2 \quad (7)$$

F_0 means the initial variance for discharge data. While F means the residual model variance. N is the total number of data. Q_{obs_n} and Q_{sim_n} are the observed and simulated data of discharge at nth time, $\overline{Q_{obs}}$ is the mean value of the whole observed discharge.

NS=1 means the model is perfect. NS between 0.9 and 1 means the model works extremely well. NS between 0.8 and 0.9 means the model works quite well. NS between 0.6 and 0.8 means the model performs reasonably well. The negative value of NS indicates that the observed value is better than the simulated one. A second objective function is the coefficient of determination (R^2) which reads:

$$R^2 = \frac{S_{reg}}{S_{tot}} \quad (8)$$

Where

$$S_{reg} = \sum_n (Q_{sim_n} - \overline{Q_{obs}})^2 \quad (9)$$

$$S_{tot} = \sum_n (Q_{obs_n} - \overline{Q_{obs}})^2 \quad (10)$$

S_{reg} means the regression sum of squares, S_{tot} means the total sum of squares. The other variables were indicated in the previous equations. The value of R^2 close to 1 means extremely high correlation, between 0.4 and 0.7 means quite high correlation, lower than 0.4 means low correlation.

The third objective function is the Root Mean Square Error (RMSE) function which reads:

$$RMSE = \left[\frac{1}{n} \sum_{i=1}^n (Q_{obs} - Q_{sim})^2 \right]^{0.5} \quad (11)$$

The variables were indicated before. The smaller the value the better the result of the simulation value.

Another objective function named RVE is used to check the quantity difference (mass balance error) between the model simulation result and observed result.

The equation for the RVE is shown:

$$RVE = \frac{\sum_{i=1}^n (Q_{sim} - Q_{obs})}{\sum_{i=1}^n Q_{obs}} \times 100 \quad (12)$$

The value of RVE can be any value between $+\infty$ and $-\infty$. In general, the value 0 means the accumulated difference between the simulated result and observed results is zero, which indicated a perfect result. However, this condition hardly happens in modelling. A relative volume error between $\pm 5\%$ indicate well performance and $\pm 5\%$ to $\pm 10\%$ indicate reasonable performance. While, the data for this research was bad (see Chapter 5), the value of RVE between $\pm 5\%$ to $\pm 20\%$ in this research also indicate the reasonable performance of the model.

5. DATA PROCESSING AND ANALYSIS

This chapter addresses the data collection and describes applied data analyses. During preparation of the data, quality of the input data for the modelling is analysed, and an overview of the whole model input is prepared.

5.1. PRECIPITATION DATA AND DISCHARGE DATA

5.1.1. INTRODUCTION

Precipitation data is one of the most important input data in hydrological models. Analysis on time series of precipitation could indicate flood or drought conditions in a catchment. For this research, time series of precipitation are selected from 2007-2010 to adopt the land cover map from the year 2011 that served as a reference to land cover change assessment. Because the wet season in Malewa start from March to May and October to November. However, in order to avoid the effect of the wet season for the stream flow, the data from 2007/02/01 to 2010/02/01 was chosen. And the data was separated to: 2007/02/01/-2009/02/01 for calibration of the model and 2009/02/01-2010/02/01 for the validation of the model.

5.1.2. RAINFALL VS. DISCHARGE

However, before preparation of the input data, it is necessary to check the data quality of the precipitation and discharge. The double-mass curve is a method used to check the relationship between the rainfall and discharge data(Sotomayor Maldonado, 2011).

In order to relate the rainfall stations and river gauge station, Table 6 shows the list of the river gauge stations and the rainfall stations which may contribute to stream flows as observed by respective gauging stations.

Table 6: The river gauge stations & the rainfall stations

River gauge station(2007/02/01-2010/02/01)	Rainfall stations(2007/02/01-2010/02/01)
2GB04	9036290&9036289
2GB01	9036129&9036294&9036034&9036264
2GB05	9036999&9036129&9036294&9036034&9036290
2GC04	9036264&9036241
2GC05	9036025&9036174&9036272
2GC07	9036025&9036164
2GB0708	9036290&9036289&9036312&9036313

The first step is to calculate the average rainfall data for the specific area that contribute to a river gauging station. And to compare the accumulated values with accumulated discharge data.

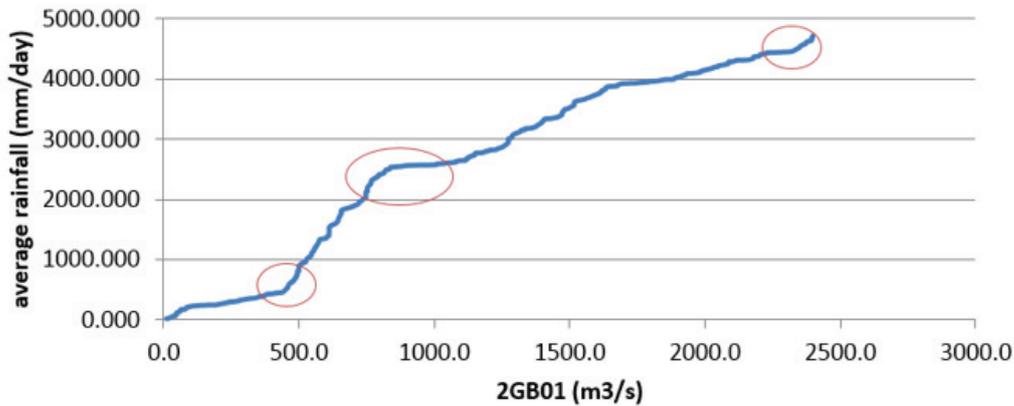


Figure 11: Double-mass curve of station 2GB01 vs. rainfall stations

Figure 11 shows an example result for the station 2GB01: The small slope changes indicate inconsistencies in the relation between rainfall and observed discharge. In general, the slope of the double-mass curve should almost be similar for the wet seasons and for the dry seasons separately. In the Figure 11, the slope shows almost four different results. This result illustrates that the relation between the rainfall data and discharge data is not good. Furthermore, it means there are errors for the rainfall stations which contribute to the river gauging station 2GB01 or that stream flow time series are unreliable.

The rainfall rate from the rainfall stations (9036129, 9036294, 9036034, and 9036264) and the hydrograph received from the discharge station 2GB01 are shown in Figure 12.

This figure illustrates some parts that rainfall data and discharge data are unrelated. The first black circle shows that at this time period high frequency rainfall occurs but only low stream flow is indicated. However, similar rainfall for the time period from 500 to 600 shows much higher stream flow. The response of the stream does not show the high stream flow when compare to the time period 500 to 600. The second black circle shows the period from 750 to 1000 with high rainfall but increases in stream flow only are small. This suggests that recorded data is not always correct.

Results for the other river gauge stations are show in the Appendix 1 and show similar inconsistencies.

The results of the double mass curve shows poor relation between the rainfall data and discharge data but to conclude which data is correct is difficult. Thus further analysis on data consistency was needed and is described in section 5.1.6.

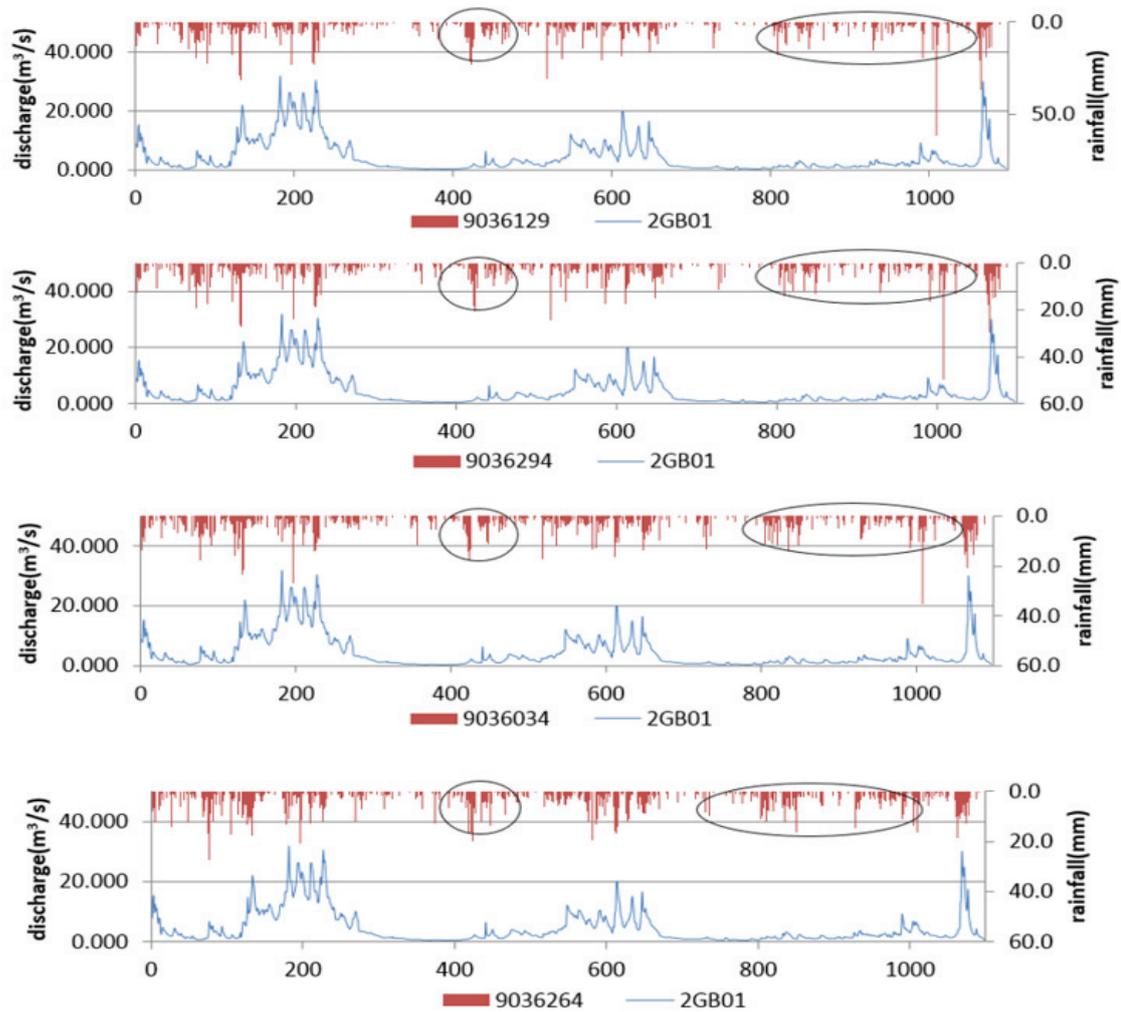


Figure 12: Observed discharge vs. rainfall station

5.1.3. RAINFALL STATIONS COMPARISON

A logical next step is to check whether the data of rainfall station is a copy from each other. Such could happen because the records of rainfall station sometimes are made by hand and since gaps in time series are simply filled with data from nearby stations. The previous section illustrates the rainfall data did not well related to discharge data. Checking whether the data of rainfall is copy from other stations can remove some stations which are not reliable. Thus, the quality of data can be improved.

During the assessment of similarity of each rainfall station, the Excel is a useful tool. The function used in this assessment is:

IF (AND ($A_i=0$, $B_i=0$), "n", A_i-B_i)

A_i = the rainfall value of the day i in station A

B_i = the rainfall value of the day i in station B

n = the day both station have no rainfall value

The function means if the day that both stations record the 0 value of rainfall, that day is a dry day. It will not be counted into the similarity comparison of the data. The difference between these values shows an “ n ”. Only when the value stands for the difference between the two stations not equal to 0, the results have been counted in the calculation. The result is 0 means the data is covered.

Then use the function “COUNTIF” to calculate how many days in these two stations gets the result of value 0 by using the first equation.

However, by comparing all the rainfall stations in the research, it is shown that the data of each station is independent and results illustrate that all the data of the rainfall station do not show signals of copying.

5.1.4. RAINFALL STATION DOUBLE-MASS CURVE

For representing spatially distributed rainfall, the whole catchment was divided into three parts (A, B, C) based on the runoff contributing area distribution for the respective gauging stations. In Figure 13 the rainfall stations within one circle are located in upstream of the river gauging station in the same circle.

To evaluate quality and reliability of rainfall time series, double curve analysis are performed. In such analysis records of a single station are compared to the averages of records of an ensemble of stations in close proximity. The rainfall station 9036025 in part A, 9036241 in part B and 9036264 in part C were used in previous study(Lukman, 2003) (with red rectangle in the Figure 13). Therefore, the data in these stations are assumed correct and regarded as the meta (reference) stations. The average rainfall data was calculated by the other rainfall stations in the same circle. In the double-mass curve, the accumulated of average rainfall and accumulated of meta station in the same circle was compared. Then, the relation of rainfall stations in the same part can be illustrated through the double-mass curve.

Results in Figure 14 indicate that rainfall records are consistent since each of the slope lines does not change much for the study period. Also, results indicate that rainfall data in the circled areas is consistent. The comparison figures between the average rainfall and discharge stations in the same circle are shown in the Appendix 2.

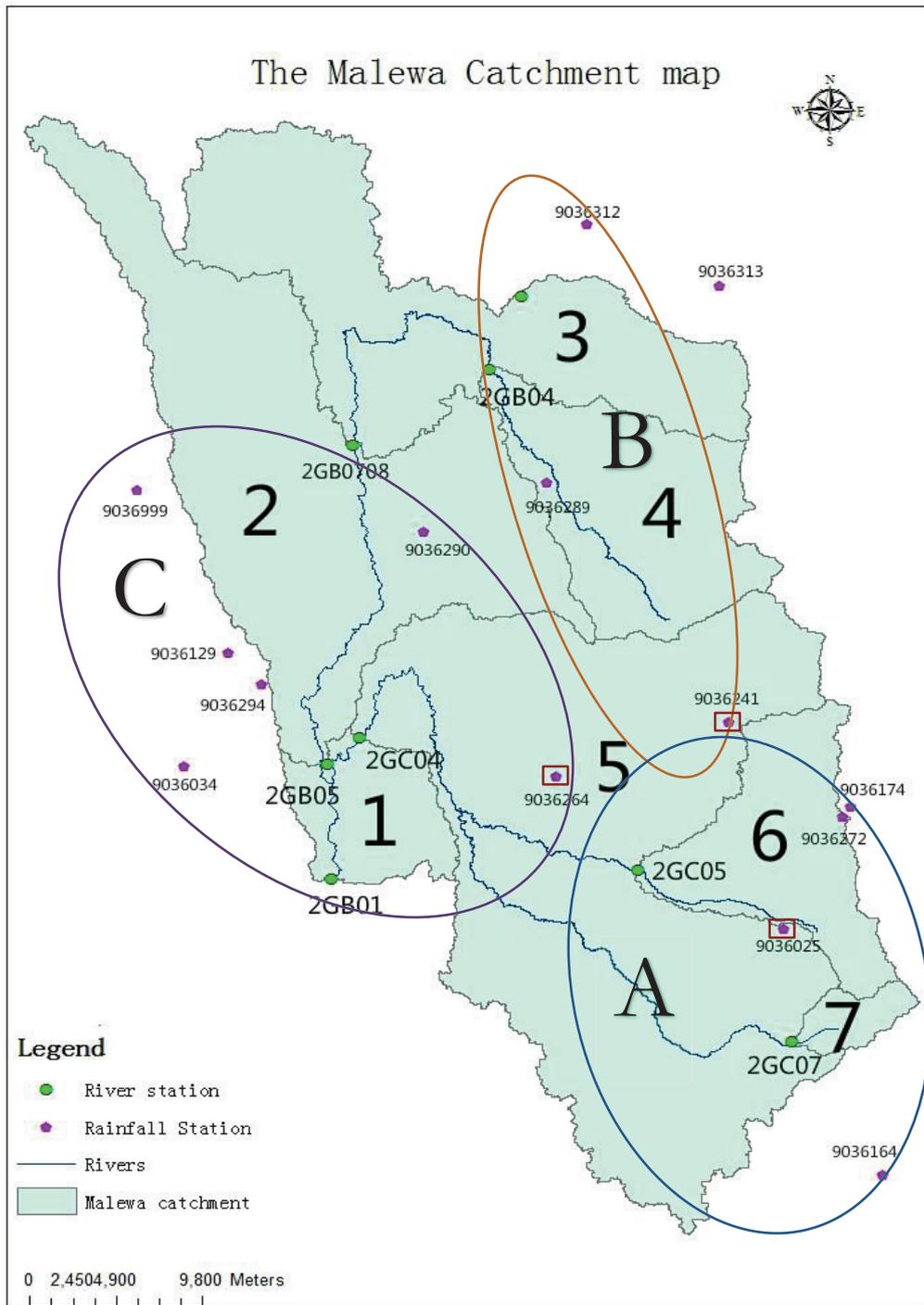


Figure 13: The three parts of catchment

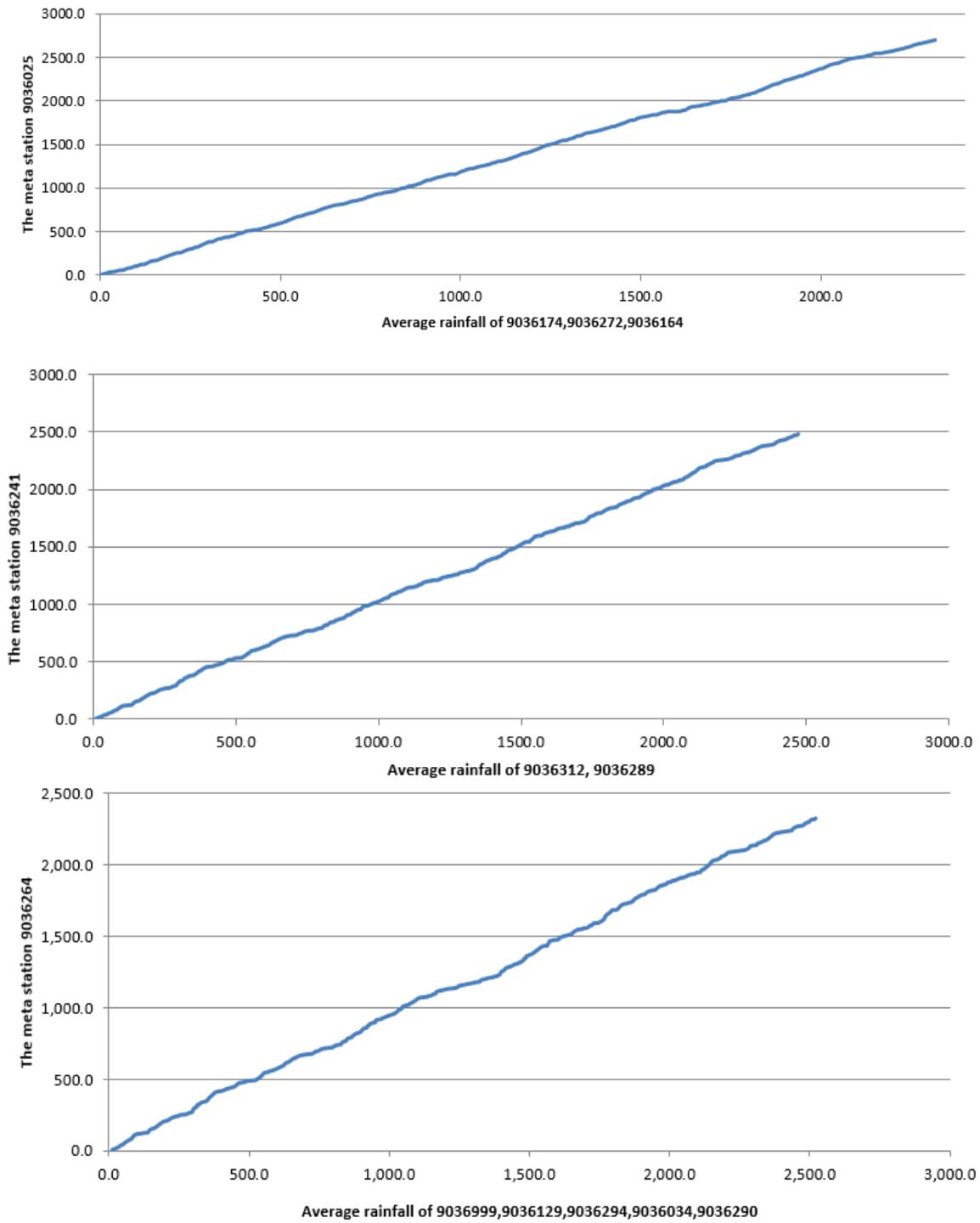


Figure 14: Results of double mass curve analysis.

5.1.5. THE INVERSE DISTANCE METHOD

Table 7 below shows the selection of the six rainfall stations used for this study. The selection includes the three stations which served as reference station in the above step.

Table 7: The six input rainfall stations

Station ID	X coordinate	Y coordinate	Elevation(m)
9036312	225396.1	9977875.4	2797
9036290	216155.3	9959398.7	2332
9036294	206913	9950213.1	2170
9036241	233394	9947946	2588
9036264	223622.8	9944687.8	2403
9036025	236545.6	9935511.5	2617

To represent rainfall in space, observations from the rainfall stations must be interpolated. Since values of rainfall stations only represent single points in the catchment. Therefore, it is necessary to interpolate the value of the six stations across each subbasin and for the whole catchment.

The inverse distance method is one of the most commonly used methods to do the interpolation of the point station to an area. The premise of this method is that the rainfall station which is nearer to the interpolation point has larger weight than other stations which are far away. In this study, the catchment is separated into seven subbasins. Since the inverse distance method weighs the contribution of each station to the interpolated values, this method is used for this research.

The equation of inverse distance method to calculate the spatial rainfall is as follows:

$$\bar{P} = \frac{\frac{1}{di^n} P_i}{\sum_{i=1}^k \frac{1}{di^n}} \quad (13)$$

\bar{P} = The average rainfall in each raster

P_i = The measured value for each station

di = The distance from the station to the raster point

n = Power parameter/weight (in this case the power parameter is 2)

k = The whole number of stations

i = The number of station

In this research, the ArcGIS is utilized to interpolate the rainfall stations to each subbasin by use the function “IDW” (Inverse distance weighting) directly and use the function “Raster calculator” to calculate the average weight of each subbasin to get one weight value presents whole subbasin.

The weight of each subbasin shows in Table 8:

Table 8: The average weight in each subbasin for each rainfall station

	9036321	9036290	9036294	9036241	9036264	9036025	SUM
subbasin1	0.02	0.121	0.394	0.081	0.323	0.061	1
subbasin2	0.101	0.475	0.273	0.051	0.071	0.03	1
subbasin3	0.438	0.302	0.104	0.073	0.073	0.01	1
subbasin4	0.16	0.3	0.07	0.26	0.15	0.06	1
subbasin5	0.02	0.08	0.07	0.23	0.33	0.27	1
subbasin6	0.01	0.02	0.02	0.33	0.09	0.53	1
subbasin7	0.01	0.02	0.02	0.101	0.071	0.778	1

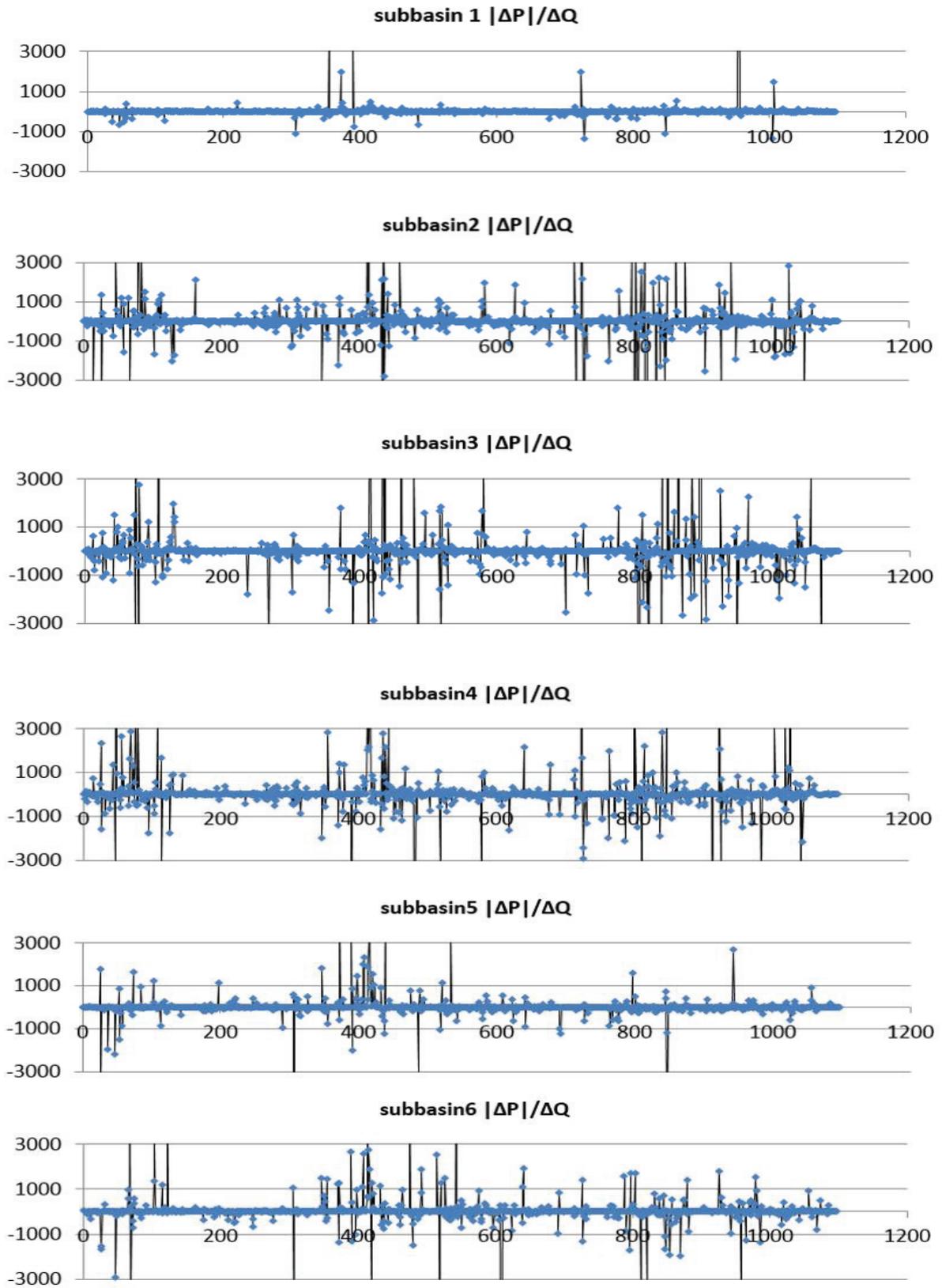
5.1.6. DATA CONSISTENCY ANALYSIS

The incremental difference method was applied to detect unreliable and fake data in the observation data of discharge. The equations of this method were described below:

$$|\Delta P| = P_t - P_{t-1} \quad (14)$$

$$\Delta Q = Q_t - Q_{t-1} \quad (15)$$

The $|\Delta P|$ and ΔQ mean the incremental difference between every consecutive time step of rainfall and discharge data. Then the $\frac{|\Delta P|}{\Delta Q}$ was calculated and plotted at daily time step for the period 2007/02/01 to 2010/02/01. Results for the respective subbasin are shown in Figure 15.



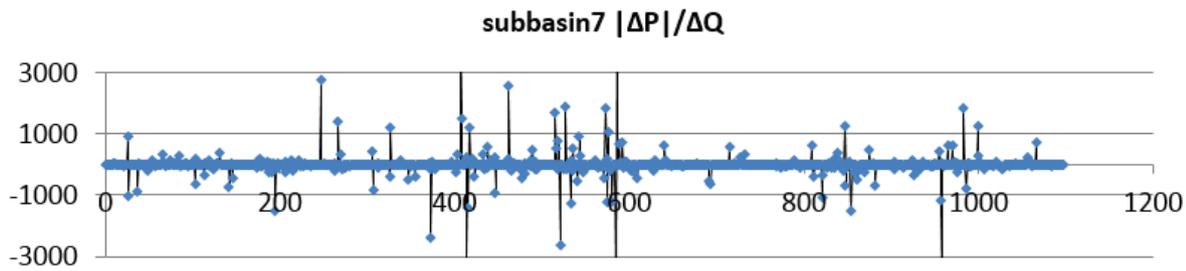


Figure 15: The results of seven subbasins for $|\Delta P|/\Delta Q$ ration during 2007/02/01-2010/02/01

The majority value for $\frac{|\Delta P|}{\Delta Q}$ was with certain interval value which can be switched depend on the different subbasin, the ratio values are usually close to zero, while the points of outliers come out when the ratio values are far from zero. Results for all stations indicate large errors in the runoff time series since ratio values generally are large. The results illustrated for subbasin 1 and subbasin 5 show best corresponding data of precipitation and discharge. Since the subbasin 1 only has 64.1 km² this subbasin 5 was finally selected for this research.

5.2. POTENTIAL EVAPOTRANSPIRATION (PET)

Potential evaporation (PET) is the maximum amount of water which normally only occurs when soils are fully saturated. Sufficient in-situ weather station data to estimate PET was not available for the years used in the research, so the data of PET is achieved from the MODIS16 products from year 2007-2010 (<http://www.ntsg.umt.edu/project/mod16>). The product MODIS16 Global Terrestrial Evapotranspiration data set is a part of NASA/EOS project to estimate the evapotranspiration from earth land surface through the remote sensing image.

The MODIS16 image includes PET datasets which has 1km×1km resolution and is estimated 8-day, monthly and annual intervals. Because the data of PET in SWAT is required for the daily value, the 8-day image was adopted in the research. After downloading the images of MODIS16, it is important to import the images into ArcGIS so to convert the projection of the images (Sinusoidal Projection) to the projection of the DEM map, land cover maps and soil type map (ARC_1960_UTM_Zone_37S). The conversion details of the projection from MODIS data to ArcGIS data was explained in the Marine Geospatial Ecology Tools website (<https://code.env.duke.edu/projects/mget/wiki/SinusoidalMODIS>).

Following the steps shown in Figure 16, the average PET in each subbasin was calculated in ArcGIS by using the function “Raster Calculator” in the time series images. In the SWAT model, the PET can be provided by four patterns: the Penman-Monteith, Priestly-Taylor, Hargreaves and the data already calculated by users. Thus, depending on the input of this research, the fourth method is the best choice.

Since the forth method of PET data in SWAT model acquire only one record for the whole study area, then the average of each subbasin should be calculated to one value which can present for the whole catchment. The equation of the calculation is shown below:

$$PET = \frac{\sum_{i=1}^7 PET_i * SUB_i}{\sum_{i=1}^7 SUB_i} \quad (16)$$

Where

PET = the value present whole area

i = the number of subbasin

PET_i = The PET value for subbasin i

SUB_i = the subbasin i

The final values of PET can represent the whole area with the unit mm/8days. While the SWAT model needs the PET data for daily value. It is necessary to calculate the value as an average for each day and finalises the processing of the PET input data for the SWAT model.

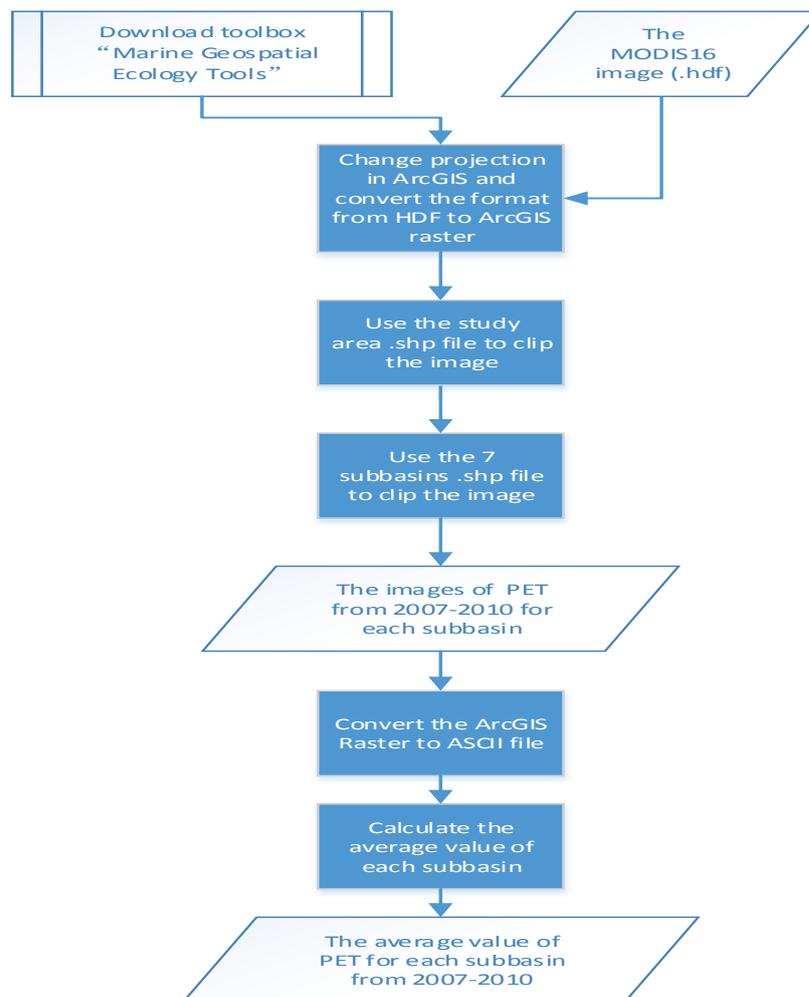


Figure 16: The steps in ARCGIS software to achieve the PET data

6. MODELLING SET-UP

6.1. INTRODUCTION

There are various specific catchment characteristics which influence the volume and generating of runoff such as slope, soil type, basin size as well as land use. In the SWAT model, the whole study area can be divided into several subbasins in order to simulate specific conditions. This is a good way to find out how land uses or soil types affect the hydrologic catchment behaviour.

The subbasin also can be further divided to the hydrologic response unit (HRU). The HRUs are lumped areas in the subbasin which have unique land cover, soil type, slope and management combinations.

To set up the model and create the input tables for the model there are several steps to be done:

1. The basin delineation
2. The land use and soil type definition (including the classification of the slope)
3. The input files of rainfall and PET.

6.2. BASIN DELINEATION

The subbasin is created to make the details of the research area are more specific. In this research, the delineation of the subbasin is through setting the river gauges which contain the observed data at the outlet of the subbasins. Which means the basin delineation is generated by available river gauges (See Figure 17). The maps of subbasin was converted from the DEM map which is already prepared by previous study(Meins, 2013). The SWAT model can recognize the .shp file and converted to the model language automatically. After the definition of basin delineation in the SWAT model, a report which contains the information of the research area such as the number of subbasins, the percentage of area at different elevation, was prepared by the model.

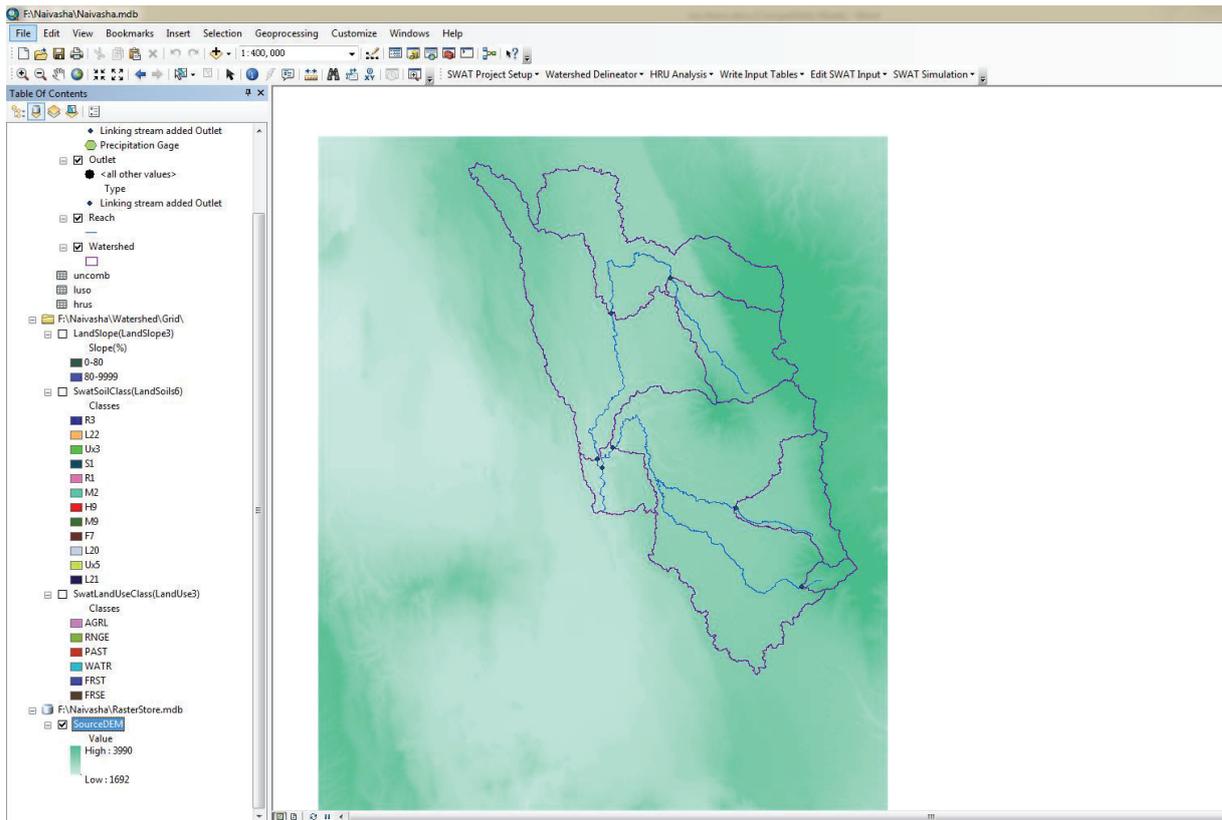


Figure 17: The ARCSWAT result after the basin delineation

6.3. THE LAND USE AND SOIL TYPE DEFINITION

When the basin delineation is finished the HRUs (hydrologic response unit) in each subbasin should be defined. The HRU has the unique land cover/soil type attributes. In this research, there are 6 land cover classes, 12 soil classes and 2 slope classes (by using the 80% threshold). Which means the catchment could have 144 HRUs in theory. However, because not every land cover classes is corresponding to every soil type and slope class. The result number of HRUs should be less than 144. Through the overlay of the map for the land cover in 2011 and soil map provide by previous study, and by choosing multiple HRUs by using 5% in land cover and 20% in soil types the final numbers of result for the definition of the HRUs is 87. After the processing, the model also provides reports with details of each HRU such as the elevation of the HRU, the subbasin it is located, the land cover type, the soil type and slope of the HRU. The report can prove that each HRU has unique land cover, soil type and slope.

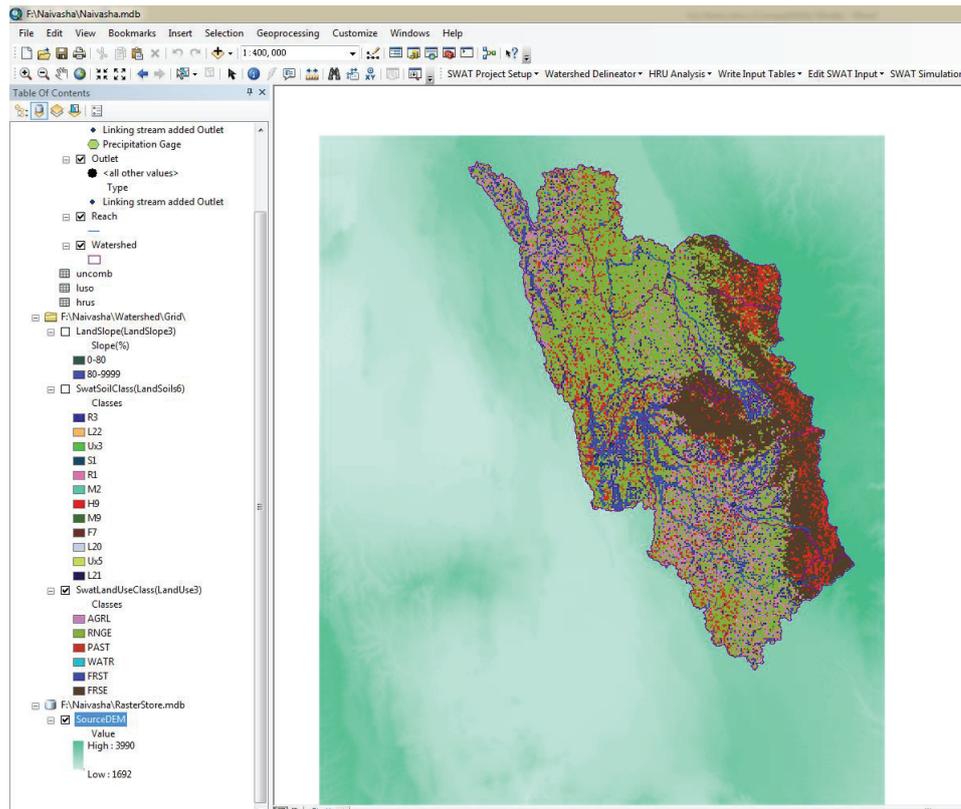


Figure 18: The reclassification result of land cover map in ARCSWAT

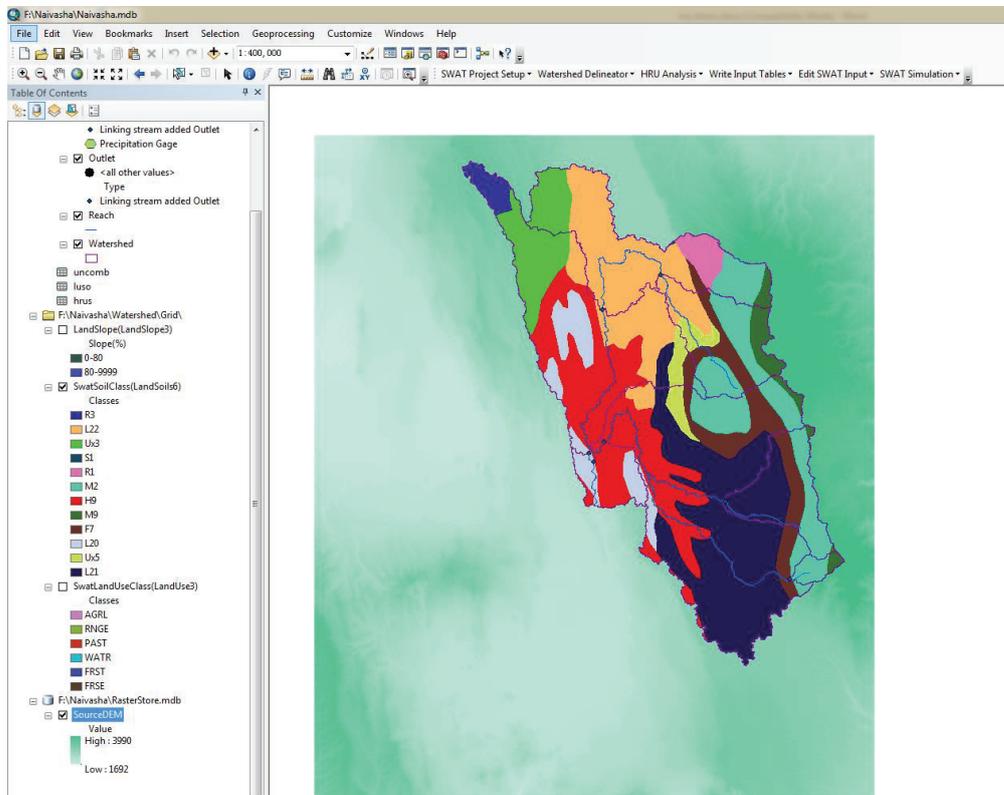


Figure 19: The reclassification result of soil map in ARCSWAT

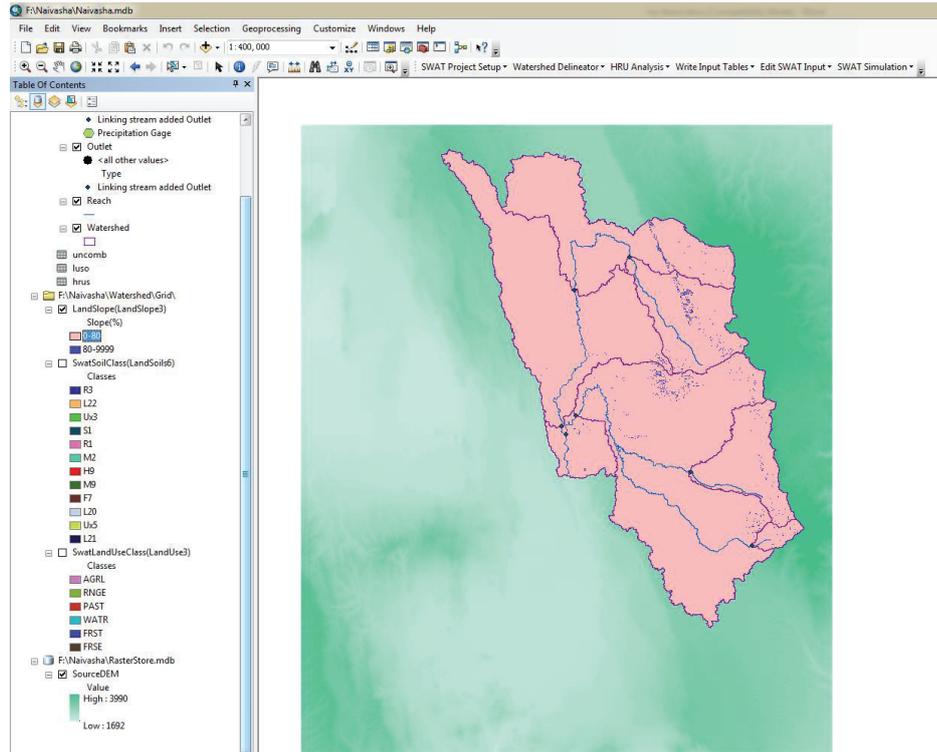


Figure 20: The reclassification result of slop in ARCSWAT

6.4. THE INPUT FILE OF RAINFALL AND PET

The input table should be created before the model can be executed. The model reads the file which contains location of the rainfall station (.bdf) and the station data files (.txt) which should have the same name in the location file for the research. When the model starts to compile the input tables, the rainfall data will be chosen by the model according to the name automatically. Besides, the model also will create the location of the stations in the ArcGIS map.

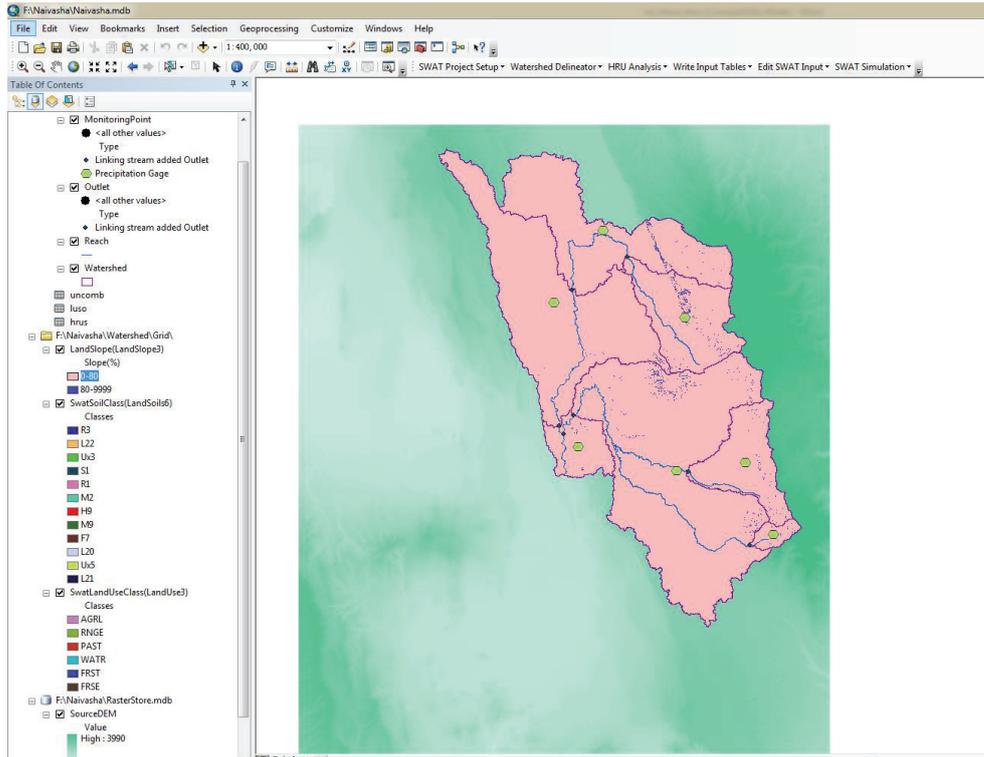


Figure 21: The location of rainfall stations in the ARCSWAT model

When compiling of the input table is finished, the model can be run and simulation results are produced.

7. RESULT AND DISCUSSION

7.1. LAND COVER CHANGE DETECTION

7.1.1. Land cover types

The aim of this research is to assess how the land cover changes affect hydrologic responses. The land cover types in this study area are defined into 6 different classes:

Grass land: the area covered by grasses for a long period.

Forest land: the areas have a high density of evergreen trees.

Wood land: the area covered by high density of mixed forest (deciduous trees and evergreen trees).

Shrub land: areas are made of shrubs, small trees as well as bushes also contains a few grasses.

Water: the area covered by water and remains for year which contains the rivers and the lakes.

Farm land: the area for crop cultivation, flowers, and scattered rural settlements.

7.1.2. Land cover change results for whole study area

In this research, there are three satellite based land cover to assess land cover changes. Maps are available for the years 1973, 1986, and 2011. The results of changes of land cover for these years are described below (see Table 9&Table 10&Figure 22).

Table 9: The land cover table for year 1973, 1986, 2011 (km²)

Year	land cover change detection (km ²)				
	1973	1986	2011	1973-1986	1986-2011
farmland	92.168	144.021	164.700	51.853	20.679
forest	386.802	206.379	284.910	-180.424	78.531
grassland	438.124	578.059	743.119	139.935	165.060
shrubland	519.934	491.372	203.522	-28.561	-287.850
water	1.319	0.609	0.635	-0.710	0.026
woodland	161.791	179.698	203.252	17.907	23.554
sum	1600.138	1600.138	1600.138	-	-

Table 10: The land cover table for year 1973, 1986, 2011 (%)

Year	Land cover change detection (%)				
	1973	1986	2011	1973–1986	1986–2011
farmland	5.760	9.001	10.293	3.241	1.292
forest	24.173	12.898	17.805	-11.275	4.908
grassland	27.380	36.126	46.441	8.745	10.315
shrubland	32.493	30.708	12.719	-1.785	-17.989
water	0.082	0.038	0.040	-0.044	0.002
woodland	10.111	11.230	12.702	1.119	1.472
sum	100.000	100.000	100.000	-	-

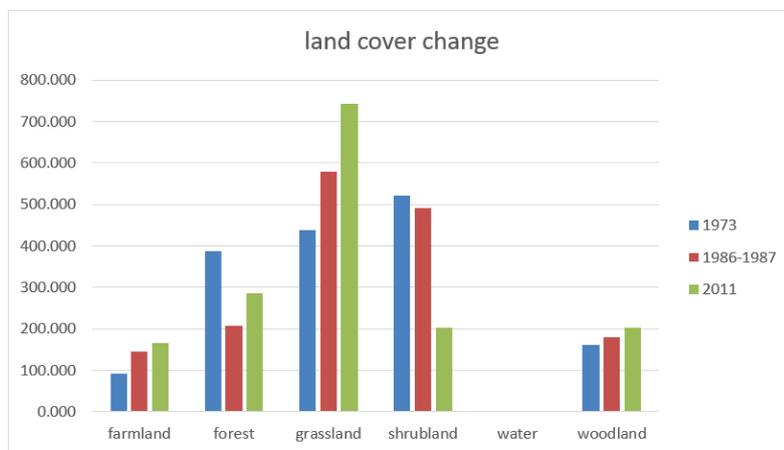


Figure 22: The figure of land cover change in year 1973, 1986, 2011

These two tables and one figure can tell the fact that during the year from 1973 to 2011, the area of farmland, woodland as well as grassland increased, especially for the grassland which changed from 438 km² to 743 km² (increase almost 20%). While the shrubland shrunk a little from 1973 to 1986 and decreased to 203 km² during 1986 the 2011 period. The change in the open water area is small and close to zero. Forest area is the only class that decreased from 1973 to 1986 and then increased from 1986 to 2011.

7.1.3. Land cover change for subbasins

Since in-situ data of this research was of very poor quality, hydrological modelling only was for subbasin 5 which has the best data available. (See 7.4.1.)

The land cover change detection for the subbasins is necessary especially for the subbasin 5. The Table 11 shows land cover in 1973, 1986 and 2011(km²).

Table 11: The land cover for subbasin5 in year 1973, 1986, 2011

subbasin5			
lulc	1973	1986	2011
farmland	82.04	48.62	92.18
forest	146.14	81.01	116.51
grassland	171.76	230.81	220.30
shrubland	105.76	153.69	68.76
water	1.10	0.11	0.34
woodland	82.78	75.07	91.01

The forest area in subbasin5 decreased from 146.14 km² in 1973 to 81.01km² in 1986 then increased to 116.51km² in 2011. A similar trend as shown for the farmland area, water, woodland. The grassland area increased from 171 km² to 230 km² and decreased a little to 220 km². The shrubland increased to 153 km² in 1986 and decreased to 68 km² in 2011.

7.2. PET DATA ANALYSIS

The PET data extracted from the MODIS 16 product was calculated 8-day sum value in each 1km × 1km raster.

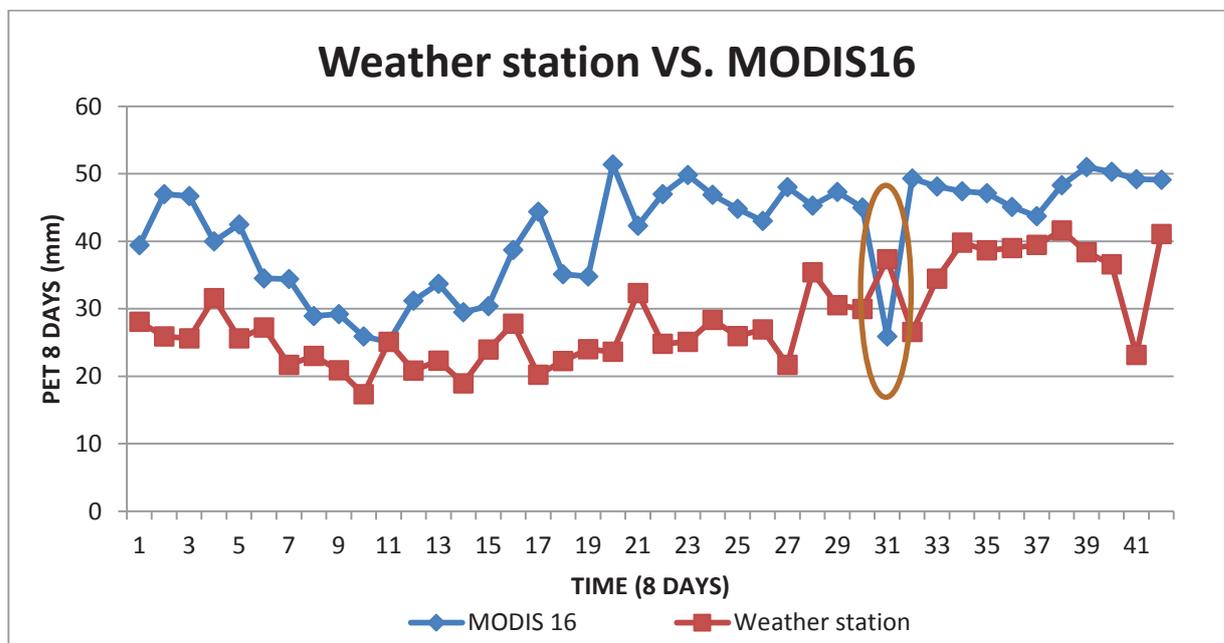


Figure 23: The comparison for the MODIS data and weather station data

The value of the PET data is compared with the reference ET_0 data that is calculated in the previous study. The previous study was using the input data from one weather station in the Malewa catchment. The result for this comparison is shown in Figure 23.

The result of the comparison shows a poor relation between the two data in the same location. Overall, the data extracted from MODIS16 product is higher than the data calculated by the weather station data, except the data in the orange circle. The differences may be caused by the atmospheric effects in the image (explained in Figure 25). The average bias of the PET data between MODIS product and weather station is 13.55 mm/8day which means approximately 1.70mm/day excluding the error point. The Figure 24 also shows the inconsistent about the data. The coefficient of determination (R^2) was only 0.21.

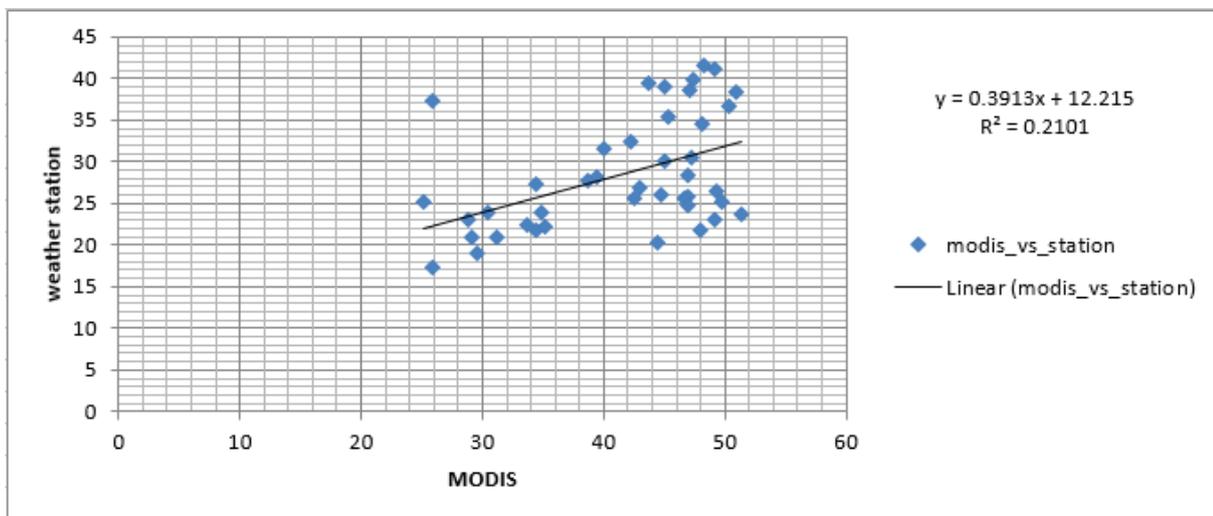


Figure 24: The consistent of MODIS16 and weather station

The Figure 25 shows the PET data extracted in different land cover classes. From this figure, all the data curves appeared to have similar trend. In the first circle, it obviously shows that the data in the same point is an error data which may be caused by the atmospheric effects in the remote sensing image. The second circle shows that the PET in the product has the range: Forest > woodland > Shrubland > Grassland > Farmland. This difference PET between the different land cover classes might be caused by the detail calculation method that is related to the difference land cover type. Since evaluation the MODIS16 product algorithm is outside the scope of this research, the exact cause is somewhat uncertain. The fraction of the algorithm is difference because of the different land cover classes.

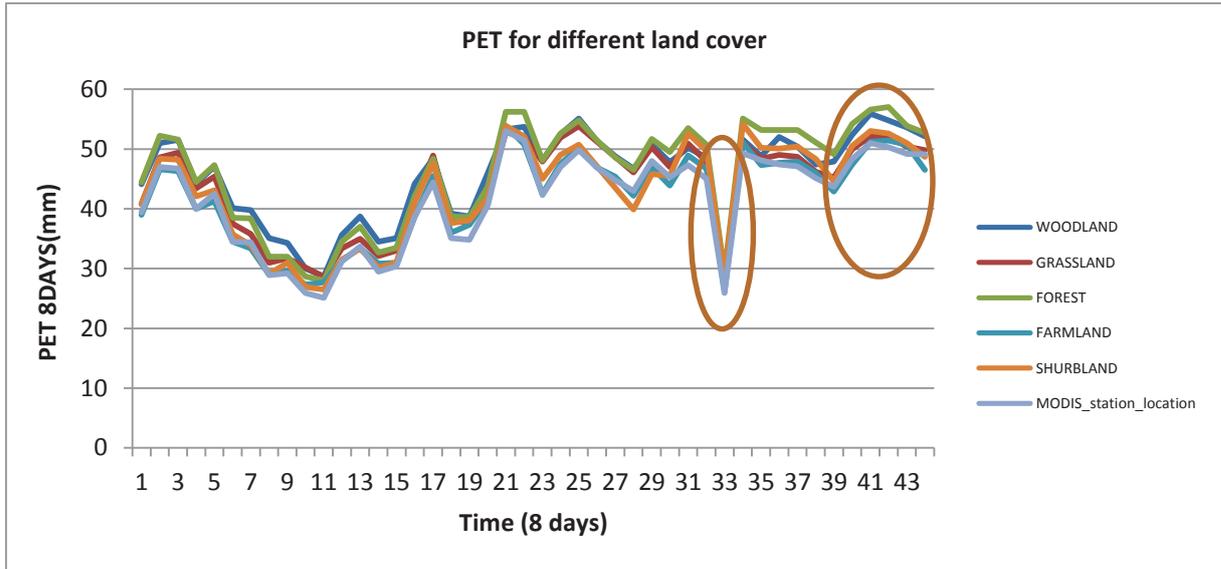


Figure 25: The PET from different land cover

Another reason that may cause the difference between the station result and MODIS16 product may be the way the PET data is calculated. Though the MODIS PET data and the weather station PET data were calculated based by the Penman-Monteith equation (Monteith, 1964), the equation used in these two PET data is a bit different depending on the different conditions. The previous study was based on the environment in the Naivasha basin while the MODIS16 product was calculated for the whole global.

The equation used in the previous study is described below:

$$ET_0 = \frac{0.408\Delta(R_n - G) + \gamma \frac{37}{T_{hr} + 273} u_2 (e^\circ(T_{hr}) - e_a)}{\Delta + \gamma(1 + 0.34u_2)} \quad (17)$$

Where

ET_0 means the potential evapotranspiration (mm/hour)

R_n means the net radiation at the grass surface (MJ/m²/hour)

G means soil heat flux density (MJ/m²/hour)

T_{hr} means hourly air temperature (°C)

Δ means saturation slope vapour pressure curve at T_{hr} (kPa/°C)

γ means psychrometric constant (kPa/°C)

$e^\circ(T_{hr})$ means saturation vapour pressure at air temperature T_{hr} (kPa)

e_a means average hourly actual vapour pressure (kPa)

u_2 means average hourly wind seed (m/s)

While the equation for the MODIS16 product is shown as Equation18:

$$\lambda E = \frac{\Delta * A + \rho * C_p * (e_{sat} - e) / r_a}{\Delta + \gamma * (1 + \frac{r_s}{r_a})} \quad (18)$$

Where

λE means latent heat flux and λ means latent heat of evaporation.

$\Delta = \Delta e_{sat} / \Delta T$ means the slope of curve relating saturated water vapour pressure to temperature

A means available energy partitioned between sensible heat, latent heat soil heat fluxes on the land surface

ρ means air density

C_p means the specific heat capacity of air

r_a means aerodynamic resistance

γ means psychometric constant (kPa/°C)

r_s means an effective resistance to evaporation from land surface and transpiration from the plant canopy

The remote sensing PET data has 10-30% range accuracy of ET observations in the global (Mu et al., 2011). Although the weather station data from previous study might be more suitable to the conditions in Naivasha basin. Because the data from the weather station cannot cover the whole time period (2007/02/01-2010/02/01) for this research. Also, only one weather station's PET cannot represent the PET for the whole area. Above all, the weather station input data of PET is not sufficient for this research. Then the PET data achieved from the remote sensing images have been assumed as the correct input PET in this research. And well prepared as the correct input format (.txt) for the SWAT model.

7.3. RAINFALL AND DISCHARGE DATA IN DIFFERENT YEARS

There is a coarse compare between the rainfall and discharge data in different years only for the subbasin 5. Because the land cover map was from 1973, 1986, 2011. And the data of year 2011 was not provided. The data of 2010 was chosen to instead the data of year 2011. The compare of rainfall and discharge data was between year 1973, 1986 and 2010 only.

The peak runoff discharge for the three years are 20.6 m³/s, 30.8 m³/s, 62.5 m³/s in year 1973, 1986, 2010 respectively. The increase in peak discharge could indicate that with the land cover change over the 37 years period, it has affected the catchment runoff behaviour. To conclude, however, rainfall, time series require investigation.

The total rainfall in the whole year has the value 837.5 mm, 1004.5 mm and 1039.5 mm, respectively. For the wet season from March to May, the total rainfall for the subbasin 5 is observed to be 252 mm, 488.2 mm and 469.7 mm respectively. The total runoff for the whole year was calculated and has the value 137.65 mm, 132.46 mm and 348.10 mm in year 1973, 1986 and 2010. And the total runoff in the wet season also

calculated with value 33.54 mm, 55.11mm, and 134.45 mm respectively in these three year. This result means that during the change in land cover the precipitation increased from 1973-2011, the discharge of the year decreased a little from 1973 to 1986 then increased in 2010. The result is shown in Table 12.

Table 12: The total rainfall, runoff and runoff coefficient

Year	Whole year			Wet season		
	1973	1986	2010	1973	1986	2010
total rainfall (mm)	837. 47	1004. 50	1039. 55	252. 00	488. 19	469. 68
total runoff (mm)	137. 65	132. 47	348. 10	33. 54	55. 11	134. 45
Run off coeff.	0. 16	0. 13	0. 33	0. 13	0. 11	0. 29

A complicating factor, however, is that the runoff time series are unreliable. The result as such cannot be trusted because the consistency for the rainfall-runoff time series for these three years data is poor. Figure 26 shows the result of rainfall and runoff in different years. It is obviously that the data is inconsistent in the year 1986 and 2010. For the 1986, the high and dense rainfall in the time 40-60 did not generate a higher peak runoff discharge while for the year 2010, the sparse rainfall in 60-70 causes the peak runoff in the wet season. These patterns are unrealistic in the real world thus the data for these years are not correct and the results are not reliable.

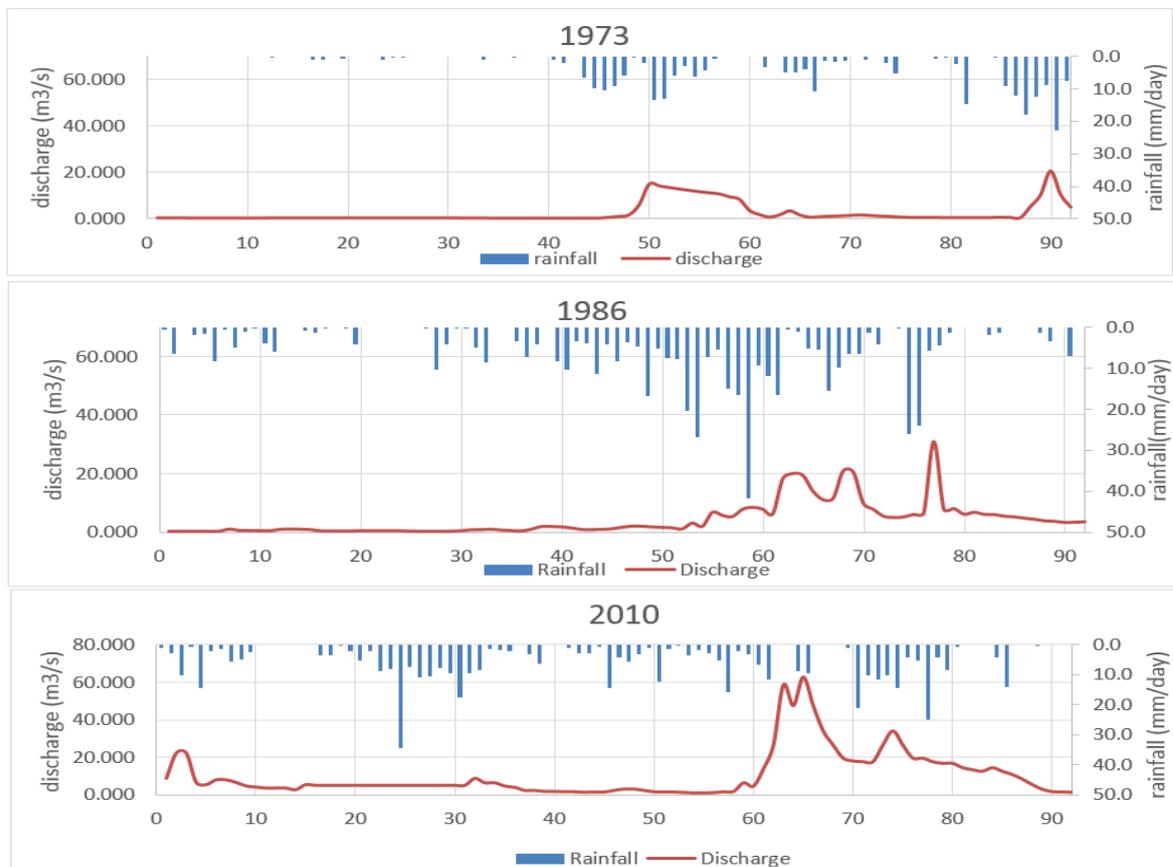


Figure 26: The rainfall and runoff comparison for the wet season in different years

7.4. THE MODELLING RESULT

7.4.1. THE FIRST RESULT OF ARCSWAT MODEL

For model simulations, the daily time step was applied by the objective of this study to also evaluate changes in peak discharge as a result of land cover change. The data used for running the model was from 2007/02/01 to 2010/02/01 (including the precipitation and PET data). For warming up the model, the input data for 2006/02/01 to 2007/01/31 needed to be used. However, the input data for PET was not sufficient from 2006/02/01 to 2007/01/31. Thus, this research assumed that the time period from 2006/02/01 to 2007/01/31 have the same input data as 2007/02/01 to 2008/01/31. And set this time period as the warming up year.

The ARCSWAT model provides the result for the whole catchment in the output.std file which includes the precipitation, surface runoff, ground water, lateral flow, actual evapotranspiration, PET, water yield (all variables mentioned above have the unit mm) and other results for the chemical elements in the soil which are ignored in this study. Besides, the model also gives the results for each subbasin (*.rch file) and HRUs (*.hru) in Dbase file (dbfs). Though a model interface are results converted from ASCII format to the Dbase format automatically. The discharge value for each subbasin in the file (*.rch file) named FLOW_OUT which has the unit m^3/s . The first simulation result is with default parameter values of the model (Figure 27).

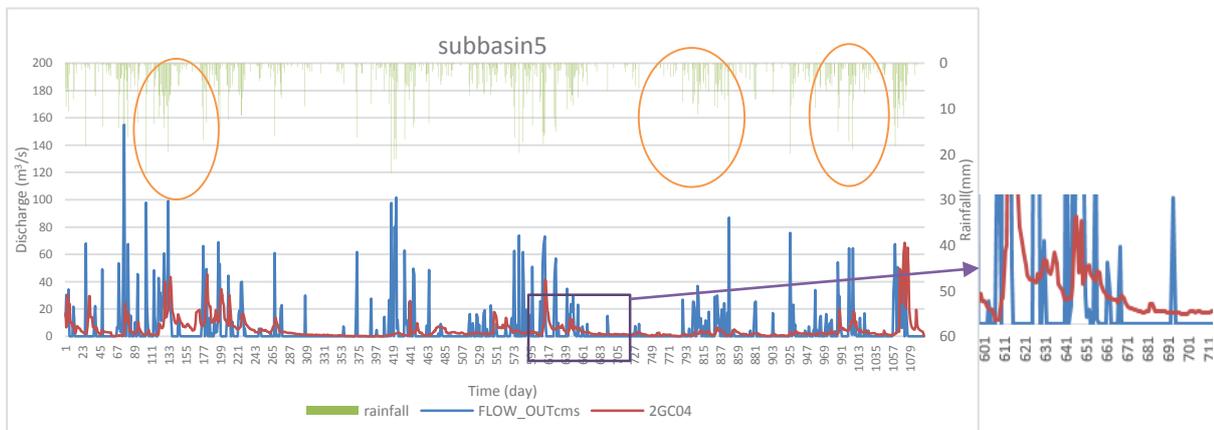


Figure 27: The first running result for ARCSWAT in subbasin 5

In Figure 27, the simulated result (blue line) with the observation one (red line) are compared. As indicated, results only are available for subbasin5. The figure illustrates that for the simulation result the discharge in the river comes from the rainfall directly, any rainfall directly causes runoff but also of there is no rainfall then there is no discharge. This phenomenon can be seen from the zoomed part in the purple rectangle. However, in the real world discharge values of zero normally only are in the dry season and applies to only ephemeral river system. The main reason for the poor simulation result is that the ground water part is not activated in the default conditions. Resulting in base flow of 0 for the first simulation. The orange circles

show the problems of the data, the rainfall data is higher and dense but the response of the observed discharge is not shown. As indicated, subbasin 5 has the best data and simulate discharge data from day 169 to 295 almost have the same response to rainfall as for the observation data. This is also for the discharge data from day 600 to 771 as well as from day 967 to 1037. So this subbasin and these periods are selected for calibration and validation of the model. All other time periods cannot be used and therefore are ignored from usage. First simulation results of the other subbasins are shown in the Appendix 3. However, it must be noted that results are very poor and thus rejected for further use. From the analysis above, the parameters which take control of the ground water would be taken into consideration of the parameters adjustment and the sensitive analysis.

7.4.2. THE MODEL CALIBRATION

The ARCSWAT model is a complex model which has a lot of parameters. These parameters would affect different parts of the hydrologic response. For these reasons, the SWAT-CUP is a helpful software to simulate the model and can adjust the parameters automatically and achieve the best parameters result. However, in order to have a better understanding of different parameters in the model, the calibration part of the model in this research was adjusted by manual operation only.

As mentioned before, the groundwater part for the first running of the model was not activated by use of the default parameter values in the research, in other words the base flow was 0. By the large number of model parameters with large value range, the manual operation and calibration is hard. Decision on parameter adjust were based on the effect of the simulated base flow. The parameters which may generate the groundwater contributions and thus change the volume of the groundwater were modified at first. Adjustments are systemic and consistent and part of a model sensitivity analysis. For selected model parameter values reference is made to Table 13.

Table 13: The parameters were chosen for this research

Parameter	Table	Description	Range	Unity
GW_DELAY	.gw	Ground water delay time	0-365	day
ALPHA_BF	.gw	Base flow alpha factor	0.1-1.0	-
RCHRG_DP	.gw	Deep aquifer percolation fraction	0-1	-
GW_SPYLD	.gw	Specific yield of the shallow aquifer	0-0.4	m ³ /m ³
SOL_AWC	.sol	Available water capacity of the soil layer	0-1	mm/mm
SOL_Z	.sol	Depth from soil surface to bottom layer	0-3500	mm
SOL_K	.sol	Saturated hydraulic conductivity	0-2000	mm/hour

7.4.3. THE EFFECT OF THE PARAMETERS

The four parameters affecting the groundwater table were tested first, the test values and the default values for those four parameters is shown in Table 14.

Table 14: The test value used for the GW parameters

GW_DELAY (31)	ALPHA_BF (0.048)	RCHRG_DP (0.05)	GW_SPYLD (0.003)
1	0.01	0.1	0.05
2	0.02	0.2	0.1
3	0.03	0.3	0.15
5	0.04	0.5	0.25
10	0.05	0.7	0.3
15	0.06	0.8	0.35
20	0.07	0.9	0.4

The effect of these parameters was analysis through the comparison between the simulate results of the subbasin 5 discharge data. However, the results for these four parameters rarely changed during the simulation in different scenarios. Above all, the results show that the problem is not the way that groundwater is simulated but much more the way runoff is generated by the combined and interacting subsurface model parts. As a consequence, parameter values for the soil part should be changed, although in previous study (Meins, 2013) these parameter values were set.

As stated, parameters for the soil property were identified and set in previous studies (Meins, 2013), and for the previous research also the ground water part shows the 0 results. So in this research, the adjustment for the soil depth (Sol_Z), available water capacity (Sol_AWC) and saturated hydraulic conductivity (Sol_K) was applied. Since these are the most important and sensitive soil parameters as shown in literature. However, the soil database has different soil type and also different soil depth for the three soil layers. If all the parameters for the different soil type were changed, then the research would have a lot of work to do. To solve the problem in a simple but feasible way, all the soil properties would be assumed to be homogeneous distributed over all HRUs. It means in the following steps the soil type for the whole area would have only one type and the three parameters of the three soil layers would be set with the same value. The value for the three parameters is shown in Table 15.

Table 15: The value for the three soil parameters in this research

SOL_Z100				SOL_Z700				SOL_Z1000			
K50	K100	K1000	K2000	K50	K100	K1000	K2000	K50	K100	K1000	K2000
SOL_AWC				SOL_AWC				SOL_AWC			
0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2
0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3
0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6

The methods used to analysis the effect of these parameters were through the visual comparison between the stream flow hydrographs with the observation data, but also through Nash-Sutcliffe (NS), coefficient of determination (R^2) and Root mean square error (RMSE) objective functions.

Results of sensitivity of the three parameters is shown in Table 16:

Table 16: The results for the chosen parameters: L means the depth of layer, K means saturated hydraulic conductivity

L100	K50				K100			
	0.1	0.2	0.3	0.6	0.1	0.2	0.3	0.6
NASH-SUTCLIFFE	-0.086	-0.013	-0.047	0.043	-0.301	-0.057	-0.079	-0.086
R2	0.123	0.180	0.181	0.179	0.091	0.123	0.104	0.110
RMSE	7.873	7.603	7.729	7.390	8.618	7.767	7.849	7.874
L700	K50				K100			
	0.1	0.2	0.3	0.6	0.1	0.2	0.3	0.6
NASH-SUTCLIFFE	-0.091	-0.134	-0.074	0.003	-0.129	-0.112	-0.097	-0.104
R2	0.111	0.115	0.150	0.149	0.104	0.077	0.087	0.092
RMSE	7.891	8.044	7.828	7.542	8.027	7.965	7.912	7.937
L1000	K50				K100			
	0.1	0.2	0.3	0.6	0.1	0.2	0.3	0.6
NASH-SUTCLIFFE	-0.206	-0.135	-0.071	0.006	-0.160	-0.110	-0.095	-0.103
R2	0.053	0.117	0.151	0.149	0.079	0.077	0.087	0.092
RMSE	8.295	8.048	7.817	7.534	8.135	7.960	7.906	7.933

L100	K1000				K2000			
	0.1	0.2	0.3	0.6	0.1	0.2	0.3	0.6
NASH-SUTCLIFFE	-2.561	-1.640	-1.501	-2.104	-3.710	-2.563	-2.521	-3.295
R2	0.035	0.037	0.032	0.034	0.028	0.028	0.025	0.026
RMSE	14.257	12.274	11.948	13.309	16.396	14.260	14.176	15.657
L700	K1000				K2000			
	0.1	0.2	0.3	0.6	0.1	0.2	0.3	0.6
NASH-SUTCLIFFE	-1.493	-1.302	-1.451	-1.985	-2.345	-2.211	-2.530	-3.188
R2	0.040	0.033	0.033	0.035	0.031	0.029	0.028	0.029
RMSE	11.927	11.462	11.827	13.053	13.818	13.536	14.193	15.461
L1000	K1000				K2000			
	0.1	0.2	0.3	0.6	0.1	0.2	0.3	0.6
NASH-SUTCLIFFE	-1.423	-1.302	-1.459	-1.996	-2.486	-2.220	-2.542	-3.203
R2	0.037	0.033	0.033	0.035	0.029	0.029	0.028	0.029
RMSE	11.760	11.462	11.847	13.076	14.105	13.556	14.218	15.488

All NS values in the table shows negative values. It is because the analysis data also including the data that have poor relation between rainfall and discharge. However, from this table, it also shows some useful

information. Compare the K50 and AWC 0.6 with different soil layer depth the NS in the column increase when the depth of the layer decrease. Thus the layer depth could be set at the range from 0-200mm. And for the K100, K1000 and K2000 the results seems even worse when the K value increase. It is illustrated that the value of parameter K might range from 0 to 50. The average values of three parameters in the original soil database was L700, K2000 and available water capacity 0.3, for this condition of soil data the value of $NS = -2.530$, $R^2=0.028$, $RMSE=14.193$. The result indicated that the original soil database has major errors with unrealistic the soil properties values.

7.4.3.1. THE EFFECT OF SOL_AWC

Through the comparison of simulation discharge data for different scenarios of parameter SOL_AWC, the Figure 28 shows the result of the effect of SOL_AWC. The change of parameter values is only for the parameter SOL_AWC. The values of parameter SOL_K and SOL_Z are not changed.

The figure shows that the parameter of SOL_AWC did effect the generation of the groundwater flow. Also it affected the peak runoff discharge. For the zoomed part in the left, this picture illustrated that the lower the fraction of available water capacity the higher the peak runoff. While, the right zoomed picture shows the lower fraction of available water capacity the higher volume of ground water flow occur, besides, the slope of the recession is also a bit higher when compare to the higher value of available water capacity. The facts also converse to be true. These results are reasonable because the fraction of available water capacity indicates the water that soil can store in the soil layer. The smaller holding capacity causes higher water percolation to the ground water store in the model. Also, it implies that the soil layers become saturated quickly even at low rainfall inputs. The saturated soil could not absorb the exceed part of rainfall thus causing the increase volume of surface runoff. It is the same with the higher runoff part in the Figure 28.

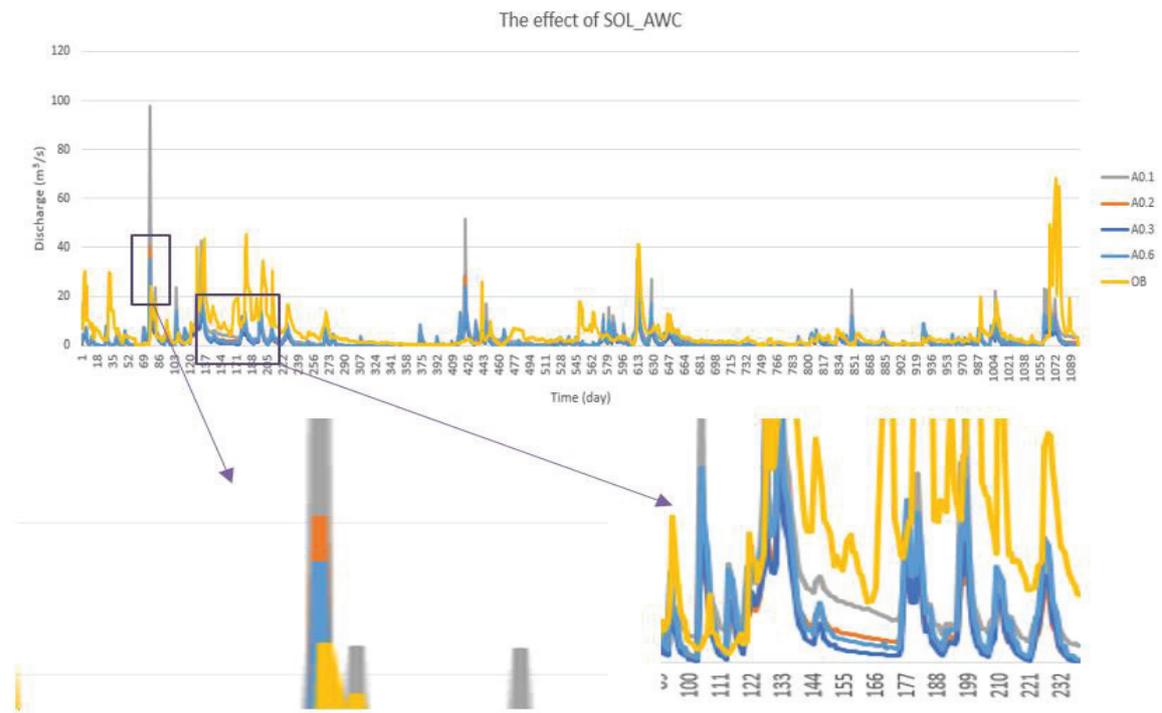


Figure 28: The effect of SOL_AWC

7.4.3.2. THE EFFECT OF SOL_K

Through the comparison of simulation discharge data for different scenarios of parameter SOL_K, the Figure 29 shows the result of the effect of SOL_K. The change of parameter values is only for the parameter SOL_K. The values of parameter SOL_AWC and SOL_Z are not changed.

The figure shows that the saturated conductivity K value in the soil also affected the volume of the ground water flow and the peak runoff discharge in the response of the rainfall. However, the effect of the saturated conductivity is not the same as the parameter SOL_AWC. When the value K was increased, the peak runoff discharge also increased while for the volume ground water shows decrease at the same time. This response is also reasonable because the saturated conductivity controls the water infiltration rate and triggers the lateral flow in the soil layer. The higher value of K, the higher water infiltration rate in the soil layer and the quicker the stream flow responses occur. Thus, the response of the discharge with the higher saturated conductivity is higher than with a lower value. Besides, the rainfall reaches the ground would leave less water for the ground water storage and increase the volume of lateral flow. This is the reason that the ground water flow shows a decrease trend when the value of saturated conductivity increase.

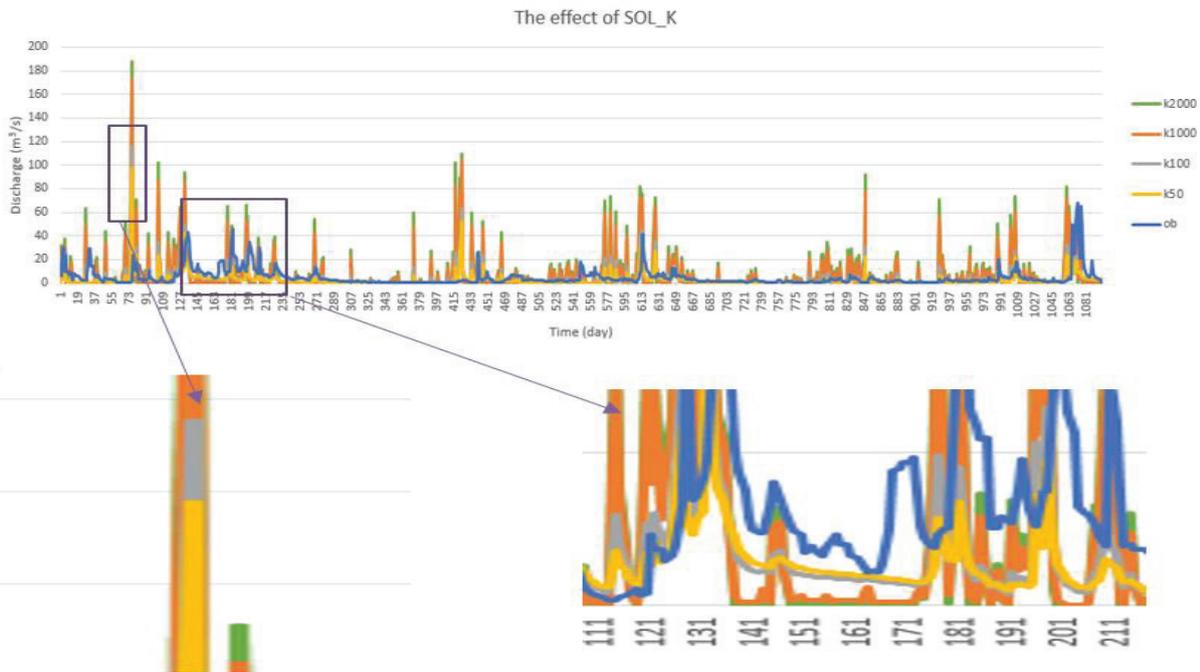


Figure 29: The effect of SOL_K

7.4.3.3. THE EFFECT OF SOL_Z

Through the comparison of simulation discharge data for different scenario of parameter SOL_Z, the Figure 30 shows the result of the effect of SOL_Z. The change of parameter values is only for the parameter SOL_Z. The values of parameter SOL_K and SOL_AWC are not changed.

This figure shows that depth of the soil layer also have the effect of peak runoff discharge and ground water flow. The zoomed part in left illustrates the lower value of the layer depth the lower volume of peak runoff discharge. However, the zoomed part in right shows the decreasing of layer depth causes the increasing of the groundwater flow. These results happen since the layer depth controls the volume of water stored in the soil layer. When other conditions are the same, the deeper the layer the stronger the capacity to absorb the water. And as explained in the manual document of SWAT, the soil layer part provide the lateral flow in the soil into the river. The more water in the soil layer becomes the lateral flow the less water percolation to the shallow aquifer to generate the groundwater flow, thus causing the decreasing of the groundwater flow.

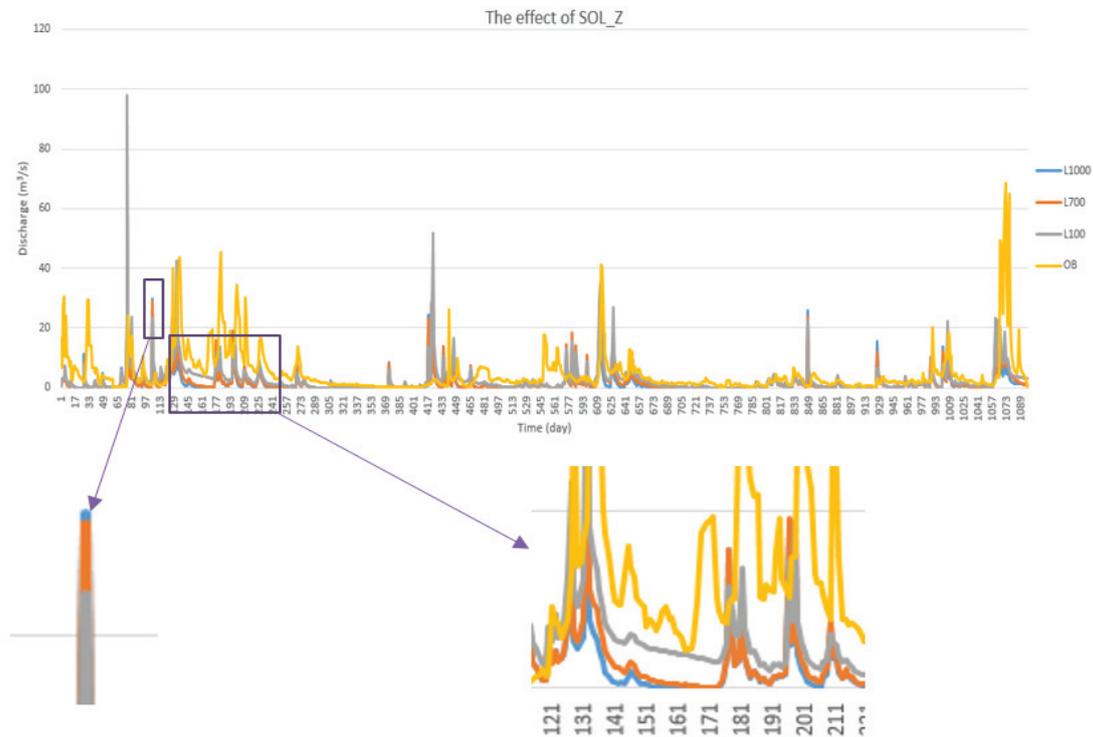


Figure 30: The effect of SOL_Z

7.4.4. THE MODEL CALIBRATION AND VALIDATION RESULTS

Because of the restrict time for the research, also because of the poor quality of the data, the manually adjusting of model could not change the parameters for 100 times. Depending on the analysis results for these parameters before, the final result of the model calibration and validation is shown in Figure 31.

The calibration data were chosen from day 63 to 413. While the validation data were chosen from day 609 to 1060. This decision was making because the poor relation between the rainfall data and discharge data effect the results a lot, it was necessary to choose the time period data which have a better relation from the long time period, thus making a reliable result for the model. For the calibration periods, the results of objective functions are $NS=0.60$, $R^2=0.62$, $RMSE=5.89m^3/s$, and $RVE=-16.90\%$, while for the validation periods, $NS=0.69$, $R^2=0.71$, $RMSE=2.71m^3/s$, and $RVE=8.9\%$ is calculated. Although the values for the objective functions are acceptable, from the figure, it still shows that some peak flow in the calibration periods does not fit well and it also happened in the validation part. However, the simulation for the ground water were almost fit with the observation value.

The Optimized parameter values for the model after calibration were: $SOL_AWC=0.04$, $SOL_K=40$, $SOL_Z=100$.

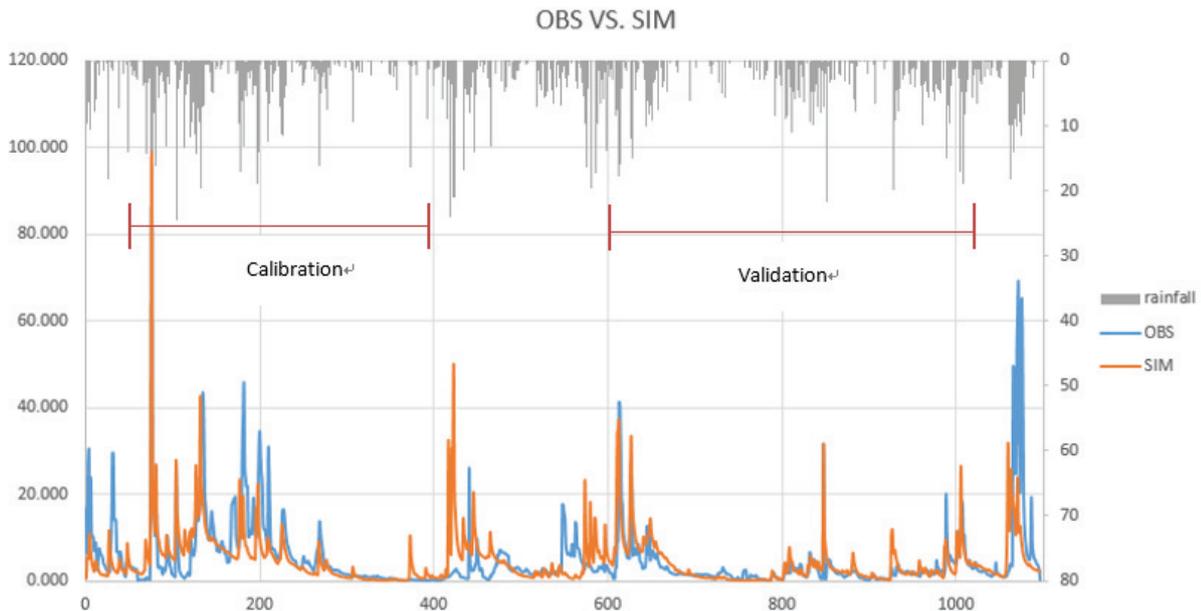


Figure 31: The result of model calibration and validation

7.4.5. THE MODEL RESULTS FOR DIFFERENT LAND COVER MAPS

7.4.5.1. The comparison of discharge in different land cover maps

The Figure 32 shows that in the subbasin 5 different land cover map affected the discharge results. For the land cover map in 1973 and the 1986, the results almost similar to each other which causes almost the overlay of the blue line (1986 land cover map) and grey line (1973 land cover map). Only a little difference between these two is that the result of 1986 map has a bit higher discharge when the rainfall is high, and the result of map 2011 always has lower discharge in dry seasons like the zoomed part in the right. However, in the wet season, comparing the biggest peak runoff discharge in different map, the result from 2011 map almost ranking the highest value. Only in a short period in the 1973 has the highest value like the zoomed picture in the left.

7.4.5.2. The comparison of base flow in different land cover maps

The Figure 33 shows the results of base flow simulated by the SWAT model. It is illustrated that the base flow simulated with the map in 2011 always has the highest value when compare with other two maps, and the lowest value always happened on 1973. Comparing the starting time of the base flow increasing in the wet season, the result for 2011 also has the most quickly rate. The slope of the recession part in base flow illustrated the base flow decreasing rate. The base flow simulated from land cover 2011 has the fastest decreasing rate while the 1973's base flow is the slowest one.

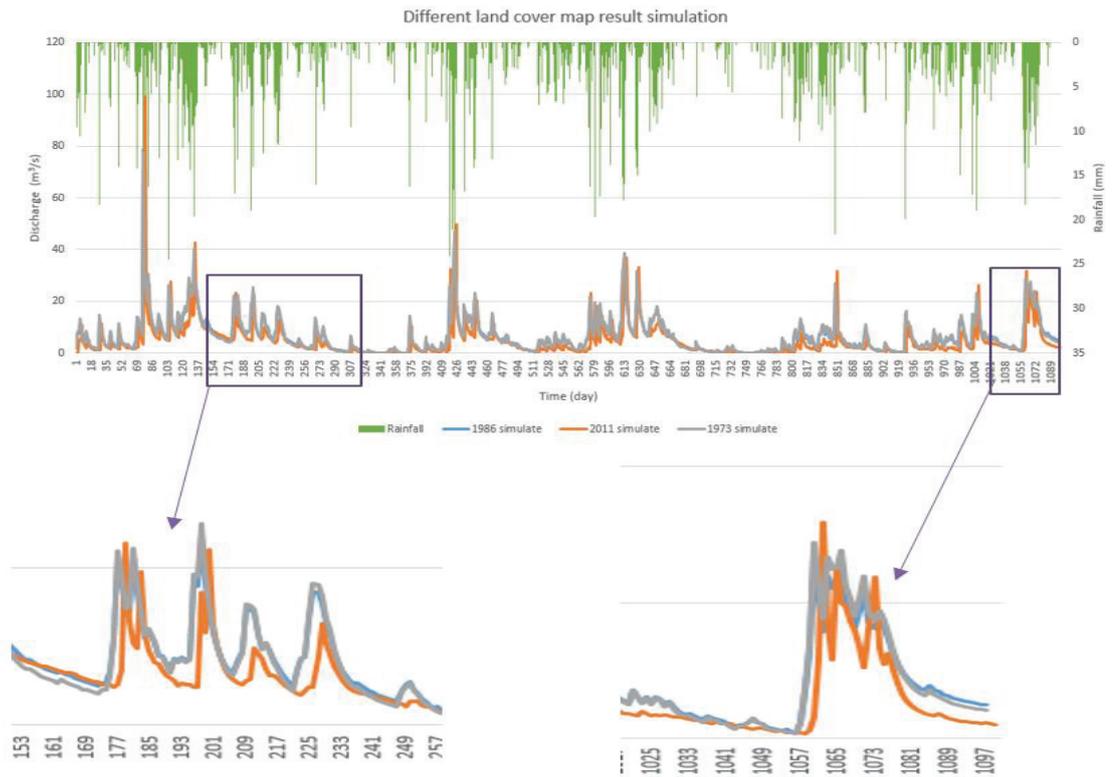


Figure 32: The comparison of discharge for different land cover simulation result

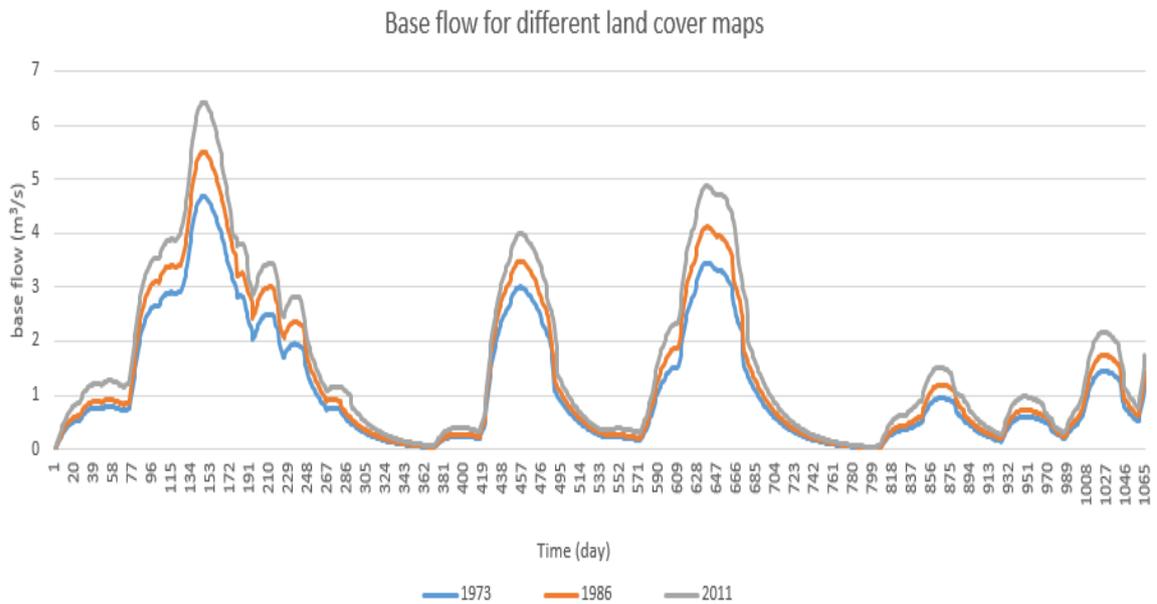


Figure 33: The comparison of base flow for different land cover simulation result

7.4.5.3. The comparison of surface runoff in different land cover maps

The Figure 34 shows the result for the surface runoff simulated by SWAT model with different land cover maps. The result illustrates that the surface runoff for the three land cover maps only have high value when

the rainfall is high. For the dry day, the values of surface runoff are approximately 0. The overlay lines also illustrates that the surface runoff for the 2011 land cover map has the highest value during this time period.

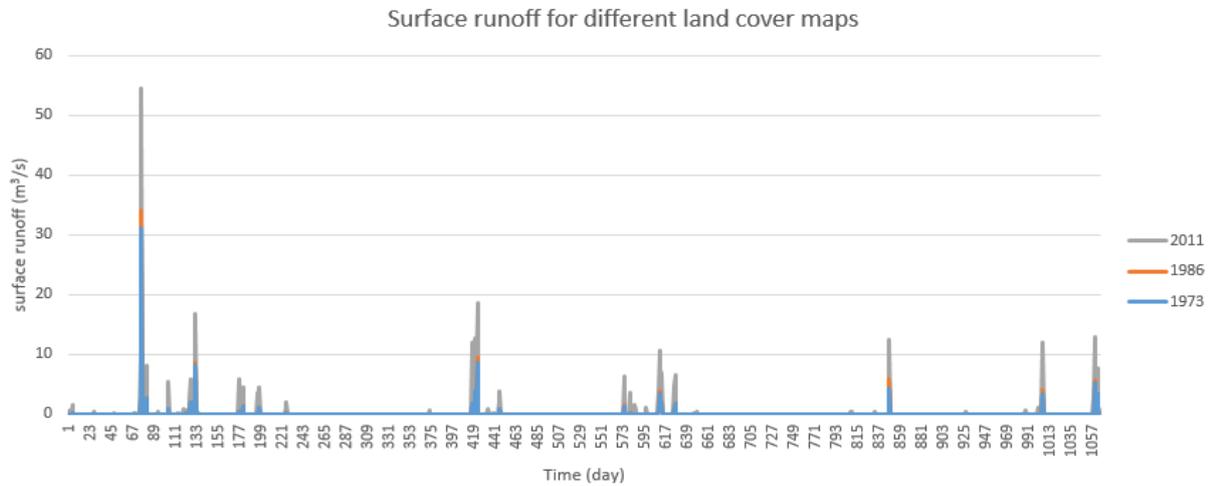


Figure 34: The comparison of surface runoff for different land cover simulation result

7.4.6. THE MODEL RESULTS FOR TWO SIMPLE SCENARIO

Because the land cover map in different years has a complex combination for the different land covers. Besides, since the results for the different land cover maps were analysed in the previous chapter, the actually effect of different land cover classes was still not clear. The results that achieved before is not clear about the effect of one land cover classes. Hence, there are two scenarios have been created to find out the effect of Forest land and Grass land on hydrological response.

The first scenario define the land cover in the subbasin 5 are full of forest land, while the second scenario define the land cover in the subbasin 5 are full of grass land. The Figure 35 shows the result of these two scenarios.

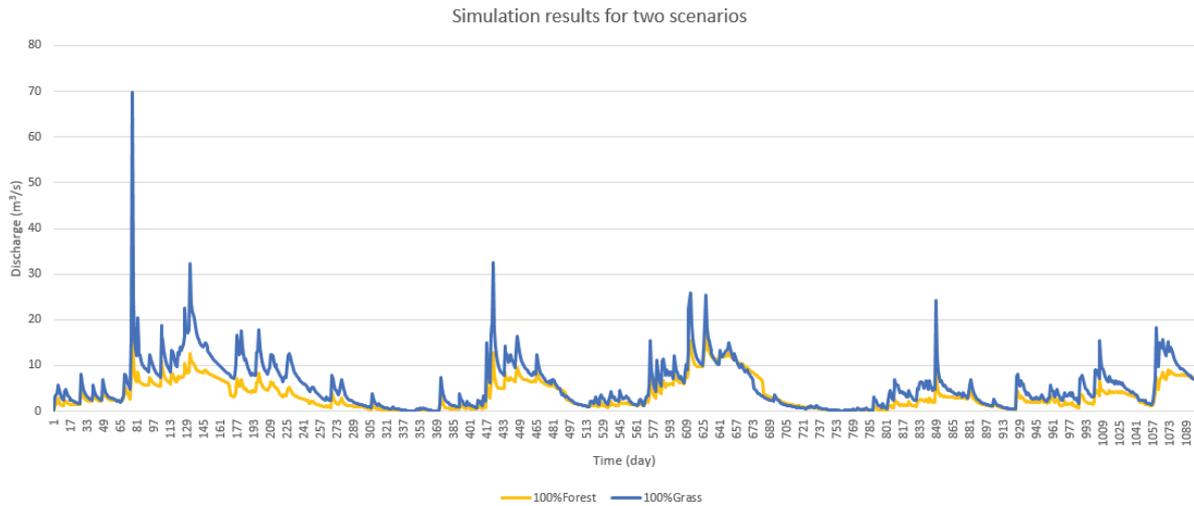


Figure 35: The comparison of two scenarios

In Figure 35, the area has forest (blue line) has higher peak flow than the area has grass (yellow line). This result is reasonable, because before the rainfall reach the ground, the water will be intercepted by the plants` leaves. Since the forest has the stronger capacity of interception of the rainfall, the discharge response of the peak flow is smaller than the grass which has a weaker capacity of interception.

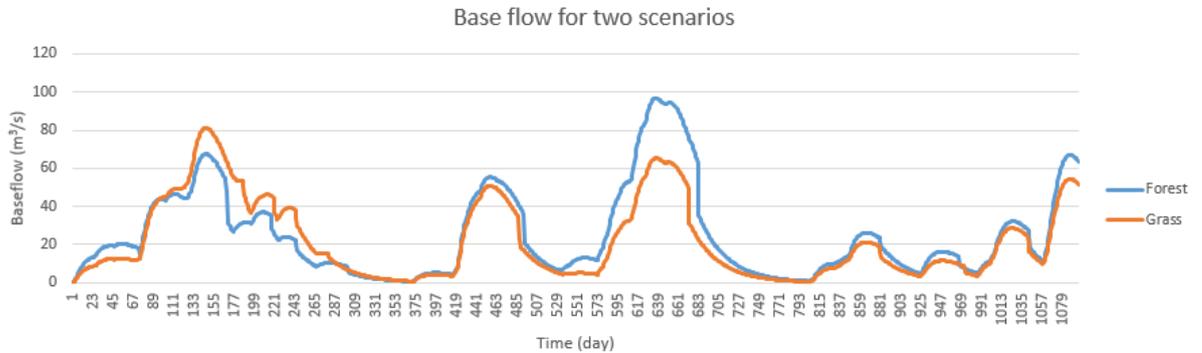


Figure 36: The comparison of base flow for two scenarios

Figure 36 shows the comparison results of base flow for the two scenarios. For the first year, the scenario for grass land has higher base flow than forest, while in the last two years, the forest land have higher base flow than grass land.

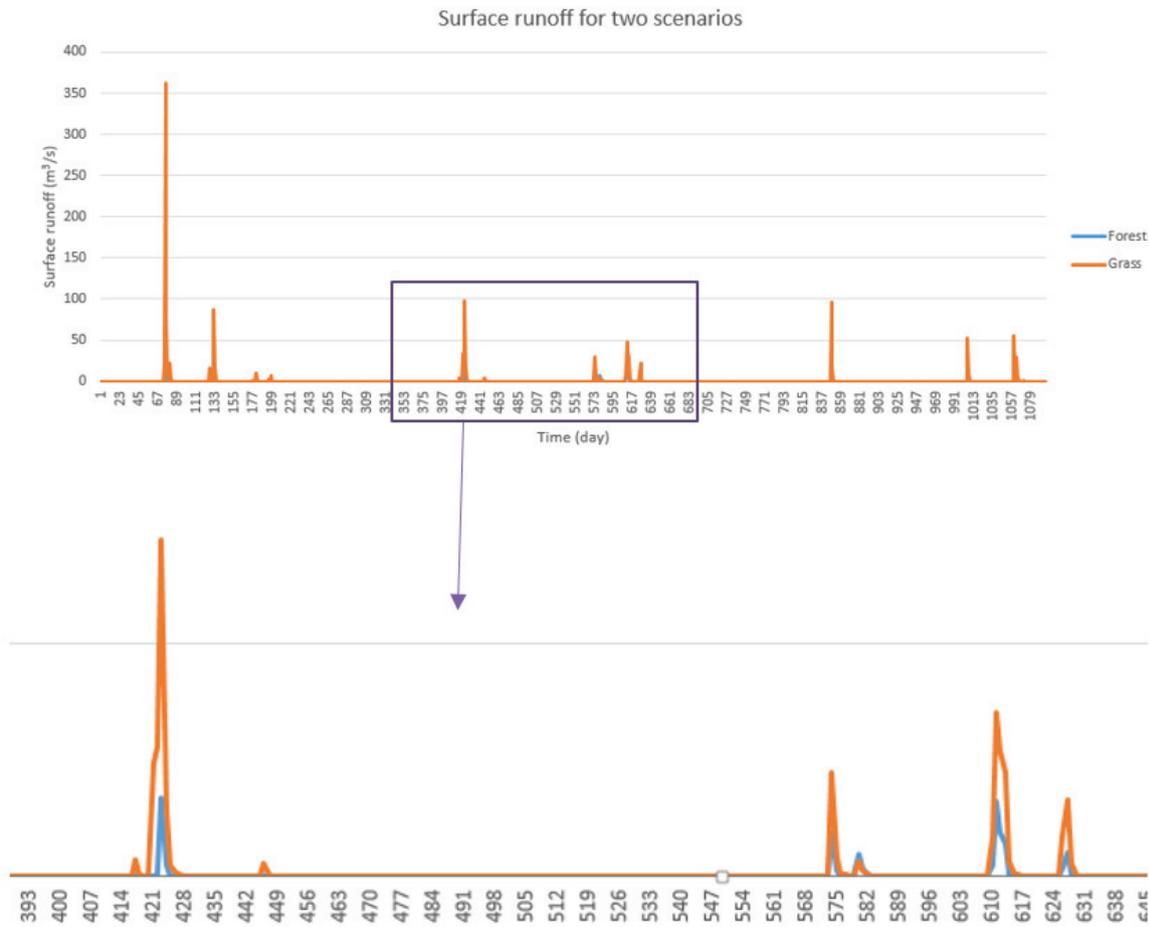


Figure 37: The comparison of surface runoff for two scenarios

The Figure 37 shows the results of surface runoff for the two scenarios in the subbasin 5. The grassland has more surface runoff than the forest land. This may also cause by the interception part of the land cover class.

8. CONCLUSIONS AND RECOMMENDATIONS

8.1. CONCLUSIONS

In this research, both the measured data (rainfall and discharge data) and satellite data (PET, DEM and land cover map) were combined to create the input data for a hydrological model. The model SWAT was used to assess the effect of the land cover change to the hydrologic response in the discharge, surface runoff, base flow and runoff volume. The following conclusions are made by the results and analysis:

- (1) The data have a really poor relation between the rainfall and discharge. Analysis on the available data of the seven subbasins that make up the study catchment, showed that only the data for subbasin 5 has the best data for the research. However, the quality of input data for subbasin 5 is not good for the whole time period. Thus, only some period for the whole time series were selected for the model simulation.
- (2) The PET data achieved from the remote sensing image in MODIS 16 product have a poor relation with the data that calculated from the weather station data in the previous study. Although both PET were calculated use the equation based on the Penman method. Since the way to choose the coefficient in the equation is different as well as the detail of the calculation steps, the results for PET are different. For the precious study, it was calculated use the condition that depend on the Naivasha catchment, while the MODIS 16 data for the PET was calculated in order to adopt the demand of PET data for the whole world. Therefore, the calculated result of PET are different.
- (3) The land cover maps for year 1972, 1986-1987 and 2011 show the results of land cover changes by human influence, for the time period 1973-2011, the grassland increased from 438 km² to 743 km² while the farmland and woodland also increased from 92 km² to 164 km² and 161 km² to 203 km², respectively. The increasing area of the three land cover classes cause the decrease of forest land and shrubland from 386.8 km² to 285 km², 519 km² to 204 km².
- (4) The hydrological modelling results illustrate that the distributed model SWAT is suitable to simulate the observed discharge with fair match. Performance assessment by objective function also showed fair performance with NS=0.6, R²=0.62, RMSE=5.89 and RVE=-16.90 in the calibration time period as well as NS=0.69, R²=0.71, RMSE=2.71 and RVE=8.9 in the validation time period. For this research, the validation period data has a better result than calibration period data.

- (5) The result of model for different land cover map in year 1973, 1986 and 2011 indicate that most peak runoff discharge in 2011 have the highest value in these three maps, only a few peak runoff discharge in 1973 have the highest value. The base flow for these three year shows that because of the change in the land cover, the base flow value increased from 1973 to 2011 as well as the surface runoff. In the two scenarios which define the whole area covered by forest and grass, the forest result shows lower peak runoff discharge and surface runoff when comparing with the grassland. For the base flow, only the first year in the simulation shows that the grass land has a higher base flow than forest. The next two years of the base flow shows the deforestation might cause the decreasing of the base flow.

8.2. RECOMMENDATION

The following recommendations and suggestions can be made:

- (1) The decision of choosing the time series of input data should be more carefully. It will be better to analysis the data for the whole time series then choose the time period which has a good quality of data. The good quality input can get better result for this research. Or before the model set up, the poor quality data should be corrected first.
- (2) The data for PET was achieved from the satellite data. It may have coarse resolutions and affect the accurate of the result when compare with the PET data that calculated from the weather stations. It is better to set more weather stations in Malewa catchment in order to get continually data for the temperature, solar radiation or other data that used to calculate the PET.
- (3) The SWAT model does not have a satisfy result for the research. The time step of this research is daily while for the further study, the time steps for monthly is suggested to assess the land cover change effect to hydrologic responses. Of course, in order to have a good result of the SWAT model by monthly step, more data will be acquired to support the further study.
- (4) In this research, because of the limitation of time, only parameter values in the groundwater and soil table were analysed and adjusted for the model. However, for further study, the parameters for surface runoff and management could be selected to figure out the effect for the discharge.
- (5) The observation data was only for discharge. It made the simulated surface runoff and base flow results have no comparison objective to assess the accurate of model. For further study, it is better to get the observation data of base flow or surface runoff.
- (6) In this research, the effect of pumping and dam were ignored. In order to make the model result closer to the reality, the consideration of pumping water and dam can be involved in further study.

LIST OF REFERENCES

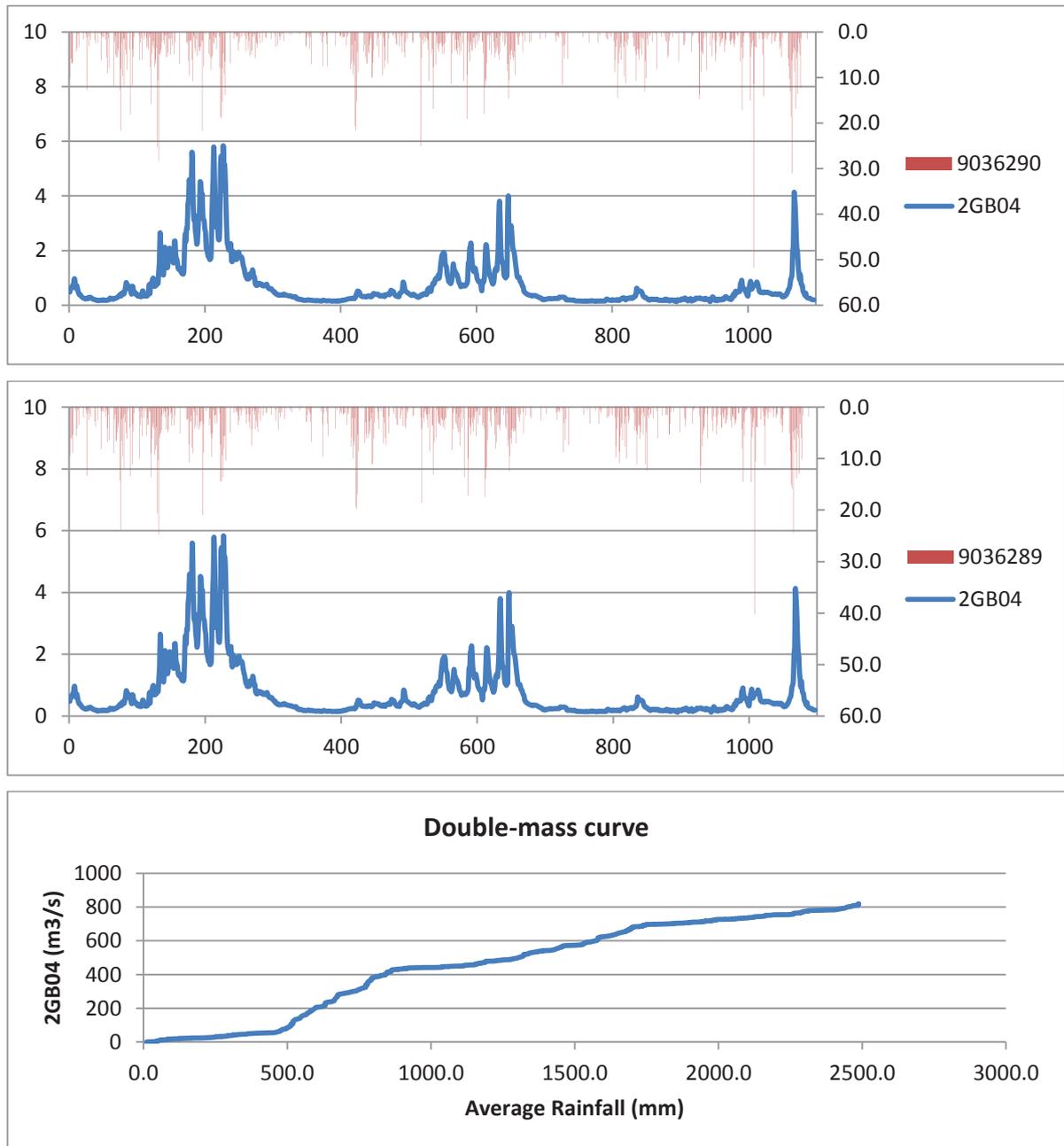
- Abiya, I. O. (1996). Towards sustainable utilization of Lake Naivasha, Kenya. *Lakes & Reservoirs: Research & Management*, 2(3-4), 231-242. doi: 10.1111/j.1440-1770.1996.tb00067.x
- Awange, J. L., Forootan, E., Kusche, J., Kiema, J. B. K., Omondi, P. A., Heck, B., Fleming, K., Ohanya, S. O., & Gonçalves, R. M. (2013). Understanding the decline of water storage across the Ramsar-Lake Naivasha using satellite-based methods. *Advances in Water Resources*, 60(0), 7-23. doi: <http://dx.doi.org/10.1016/j.advwatres.2013.07.002>
- Becht, & Harper, D. (2002). Towards an understanding of human impact upon the hydrology of Lake Naivasha, Kenya. In D. Harper, R. R. Boar, M. Everard & P. Hickley (Eds.), *Lake Naivasha, Kenya* (Vol. 168, pp. 1-11): Springer Netherlands.
- Becht, R., Odada, E., & Higgins, S. (2011). Lake Naivasha: experiences and lessons learned brief. International Lake Environment Committee, Lake Basin Management Initiative.
- Deelstra, J., Farkas, C., Engebretsen, A., Kværnø, S., Beldring, S., Olszewska, A., & Nesheim, L. (2010). Can we simulate runoff from agriculture-dominated watersheds? Comparison of the DrainMod, SWAT, HBV, COUP and INCA models applied for the Skuterud catchment. *Bioforsk Fokus*, 5, 119-128.
- Gaudet, J. J., & Melack, J. M. (1981). Major ion chemistry in a tropical African lake basin. *Freshwater Biology*, 11(4), 309-333. doi: 10.1111/j.1365-2427.1981.tb01264.x
- Hundecha, Y., & Bárdossy, A. (2004). Modeling of the effect of land use changes on the runoff generation of a river basin through parameter regionalization of a watershed model. *Journal of Hydrology*, 292(1-4), 281-295. doi: <http://dx.doi.org/10.1016/j.jhydrol.2004.01.002>
- Liu, X., Ren, L., Yuan, F., Singh, V. P., Fang, X., Yu, Z., & Zhang, W. (2008). Quantifying the effect of land use and land cover change on green water and blue water in northern part of China. *Hydrol. Earth Syst. Sci. Discuss.*, 5(4), 2425-2457. doi: 10.5194/hessd-5-2425-2008
- Lukman, A. P. (2003). *Regional impact of climate change and variability on water resources : case study Lake Naivasha basin, Kenya*. ITC, Enschede. Retrieved from http://www.itc.nl/library/papers_2003/msc/wrem/lukman.pdf
- Meins, F. M. (2013). Evaluation of spatial scale alternatives for hydrological modelling of the Lake Naivasha basin, Kenya.
- Monteith, J. (1964). Evaporation and environment, the state and movement of water in living organisms. Symposium of the society of experimental biology, Vol. 19 (pp. 205- 234): Cambridge: Cambridge University Press.
- Morán-Tejeda, E., Ceballos-Barbancho, A., Llorente-Pinto, J., & López-Moreno, J. (2012). Land-cover changes and recent hydrological evolution in the Duero Basin (Spain). *Regional Environmental Change*, 12(1), 17-33. doi: 10.1007/s10113-011-0236-7
- Mu, Q., Zhao, M., & Running, S. W. (2011). Improvements to a MODIS global terrestrial evapotranspiration algorithm. *Remote Sensing of Environment*, 115(8), 1781-1800. doi: <http://dx.doi.org/10.1016/j.rse.2011.02.019>
- Mulvaney, T. (1851). On the use of self-registering rain and flood gauges in making observations of the relations of rainfall and flood discharges in a given catchment. *Proceedings of the Institute of Civil Engineers of Ireland*, 4, 18-31.
- Neitsch, S. L. e. a. (2002a). Soil and Water Assessment Tool Theoretical Documentation. *Texas, Texas Water Resources Institute, College Station*.
- Neitsch, S. L. e. a. (2002b). Soil and Water Assessment Tool User`s Manual. *Texas, Texas Water Resources Institute, College Station*.
- Rientjes, T. H. M., Haile, A. T., Kebede, E., Mannaerts, C. M. M., Habib, E., & Steenhuis, T. S. (2011). Changes in land cover, rainfall and stream flow in Upper Gilgel Abbay catchment, Blue Nile basin – Ethiopia. *Hydrol. Earth Syst. Sci.*, 15(6), 1979-1989. doi: 10.5194/hess-15-1979-2011
- Sombroek, W. G., Braun, H. M. H., & van de Pauw, B. J. A. (Cartographer). (1982). Exploratory Soil Map and AGRO-Climatic Zone Map of Kenya, 1980.
- Sotomayor Maldonado, M. E. (2011). *Remote sensing based hydrologic modeling in the Bababoyo river sub - basin for water balance assessment*. University of Twente Faculty of Geo-Information and Earth Observation (ITC), Enschede. Retrieved from http://www.itc.nl/library/papers_2011/msc/wrem/sotomayor.pdf

- Stoof-Leichsenring, K., Junginger, A., Olaka, L., Tiedemann, R., & Trauth, M. (2011). Environmental variability in Lake Naivasha, Kenya, over the last two centuries. *Journal of Paleolimnology*, 45(3), 353-367. doi: 10.1007/s10933-011-9502-4
- Tang, L., Yang, D., Hu, H., & Gao, B. (2011). Detecting the effect of land-use change on streamflow, sediment and nutrient losses by distributed hydrological simulation. *Journal of Hydrology*, 409(1–2), 172-182. doi: <http://dx.doi.org/10.1016/j.jhydrol.2011.08.015>
- van Oel, P. R., Mulatu, D. W., Odongo, V. O., Meins, F. M., Hogeboom, R. J., Becht, R., Stein, A., Onyando, J. O., & van der Veen, A. (2013). effects of groundwater and surface water use on total water availability and implications for water management : the case of Lake Naivasha, Kenya. *JournalWater resources management*(IN PRESS), 16 p.
- Yan, B., Fang, N. F., Zhang, P. C., & Shi, Z. H. (2013). Impacts of land use change on watershed streamflow and sediment yield: An assessment using hydrologic modelling and partial least squares regression. *Journal of Hydrology*, 484(0), 26-37. doi: <http://dx.doi.org/10.1016/j.jhydrol.2013.01.008>
- Zhang, X., Liu, Y., Fang, Y., Liu, B., & Xia, D. (2012). Modeling and assessing hydrologic processes for historical and potential land-cover change in the Duoyingping watershed, southwest China. *Physics and Chemistry of the Earth, Parts A/B/C*, 53–54(0), 19-29. doi: <http://dx.doi.org/10.1016/j.pce.2011.08.021>

9. APPENDIX

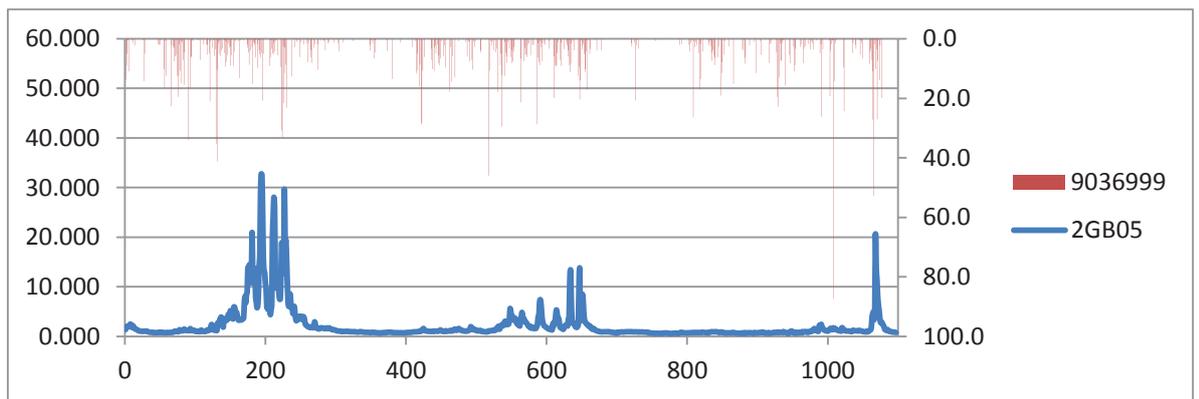
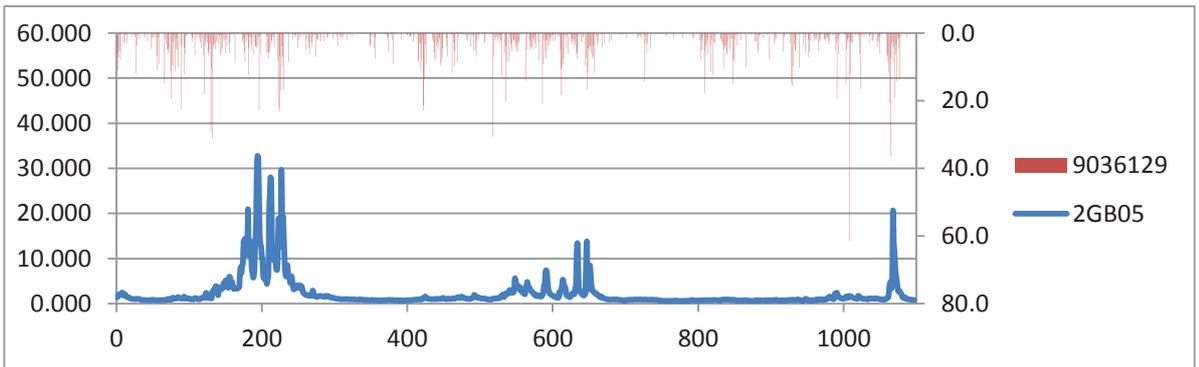
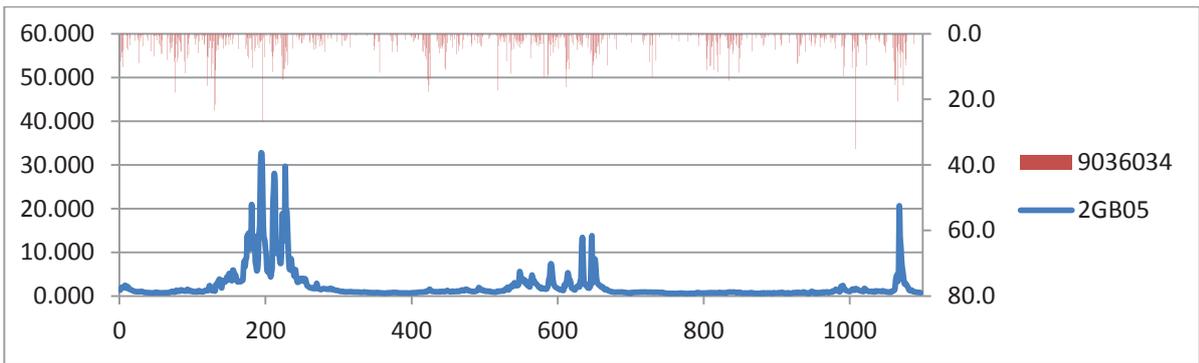
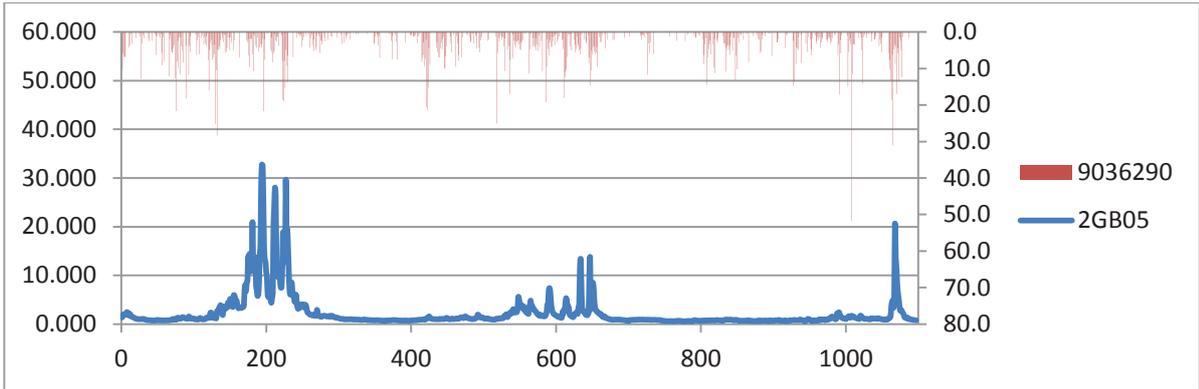
9.1. APPENDIX 1

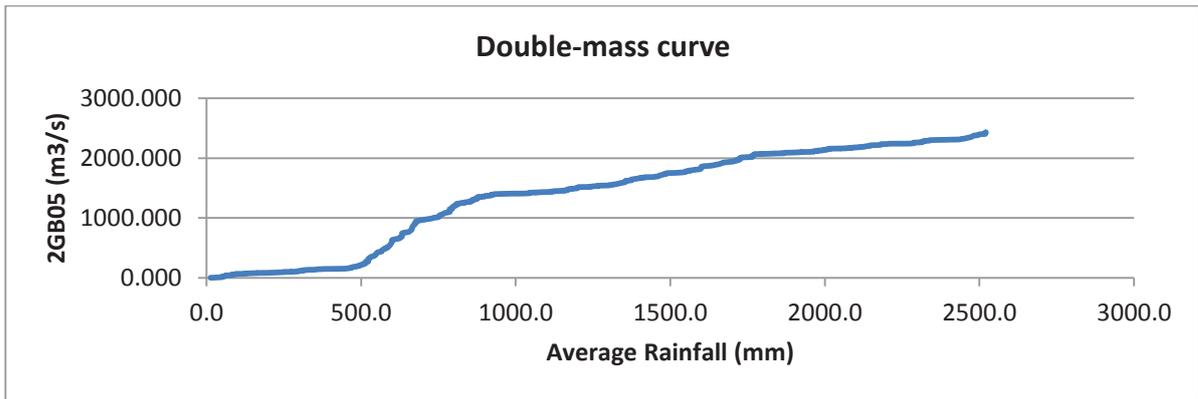
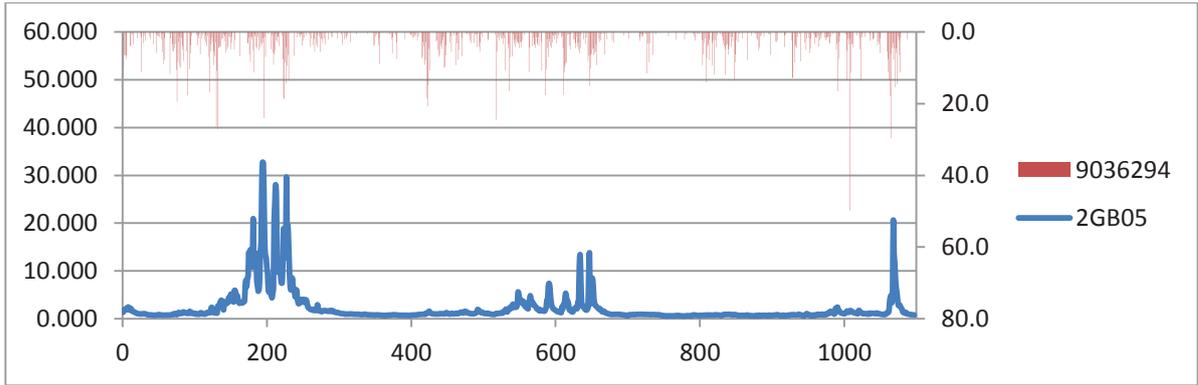
(1) River station 2GB04 (outlet of subbasin4):



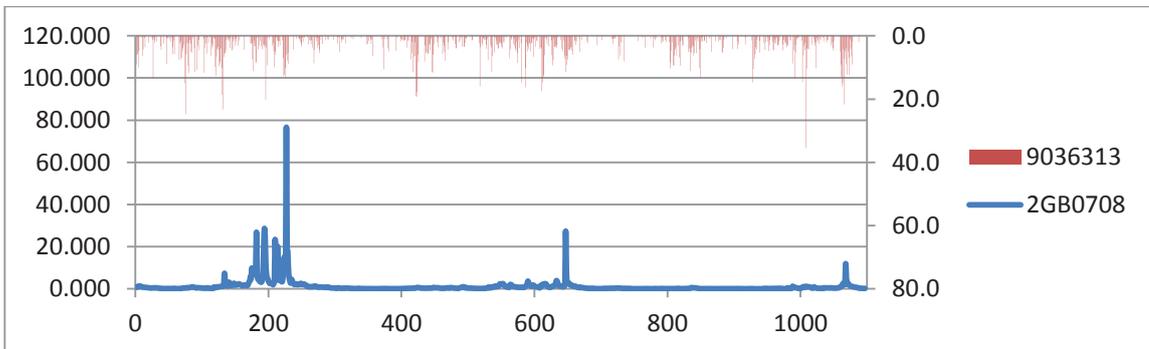
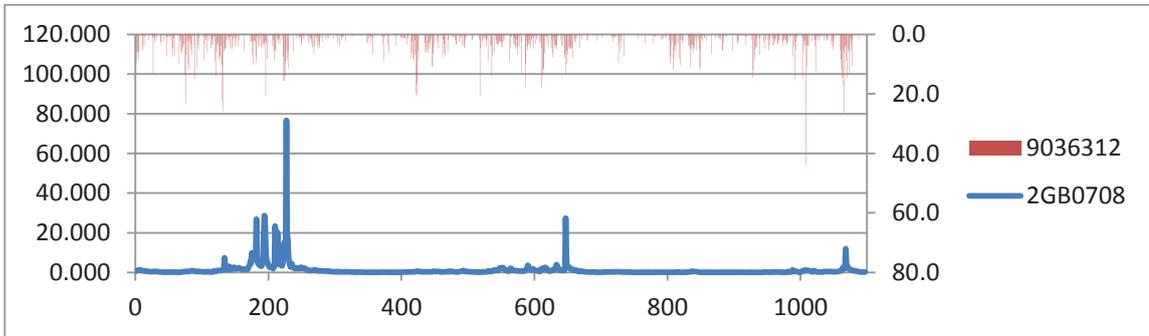
(2) River station 2GB01 (outlet of subbasin1 or the outlet of whole catchment) (See section 5.1.2)

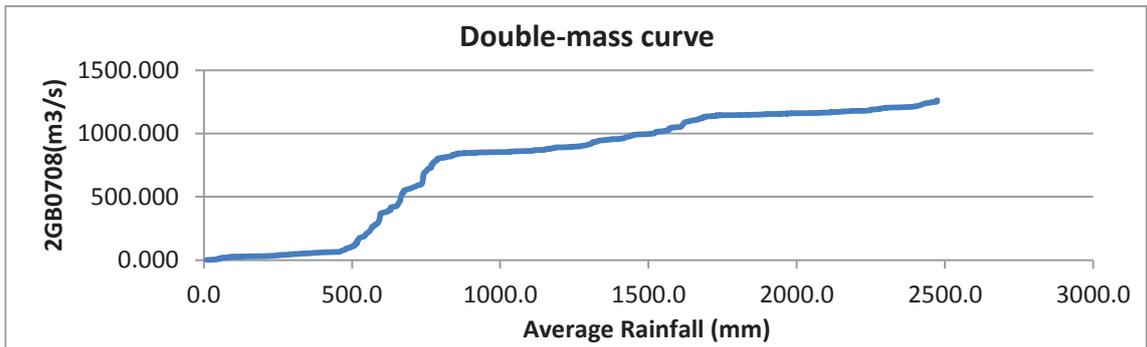
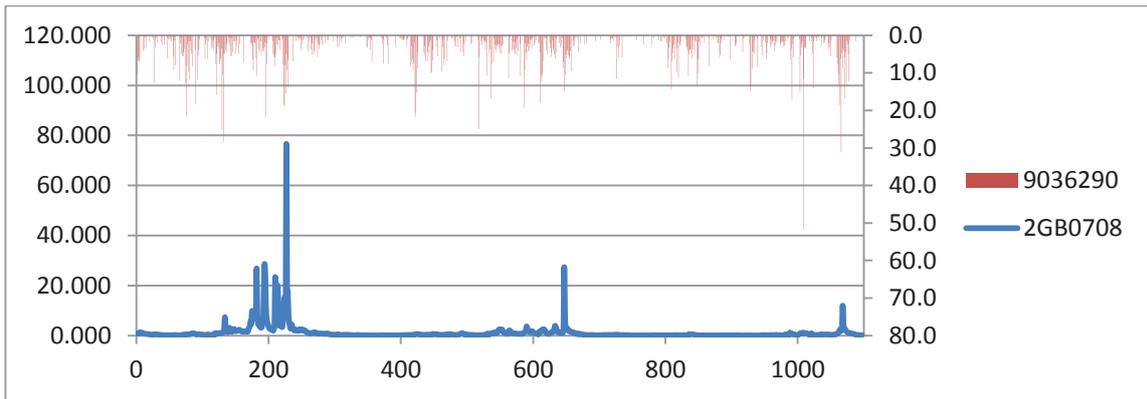
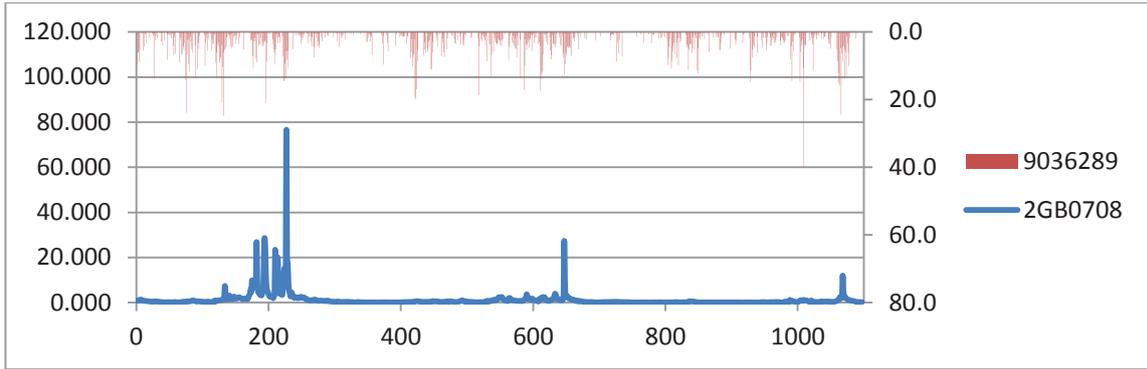
(3) River station 2GB05 (outlet of subbasin2):



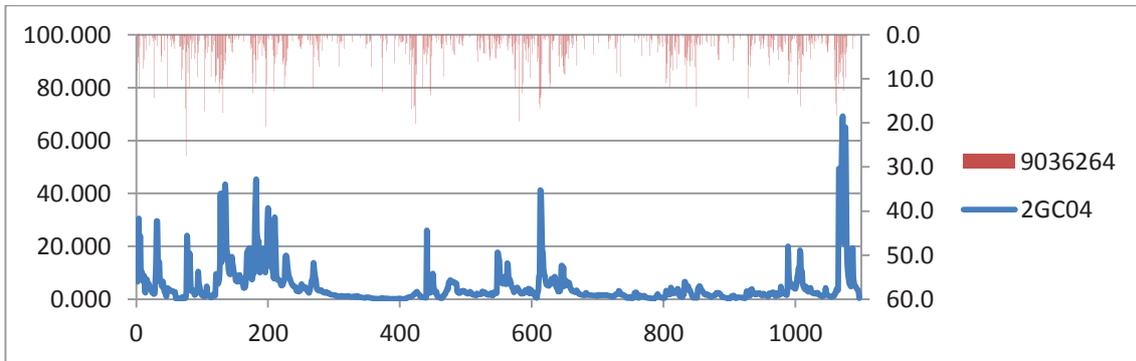


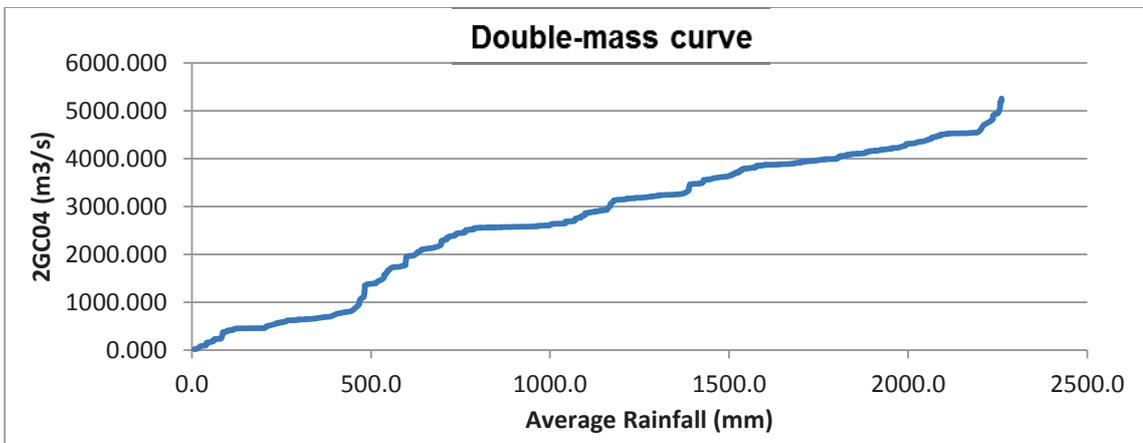
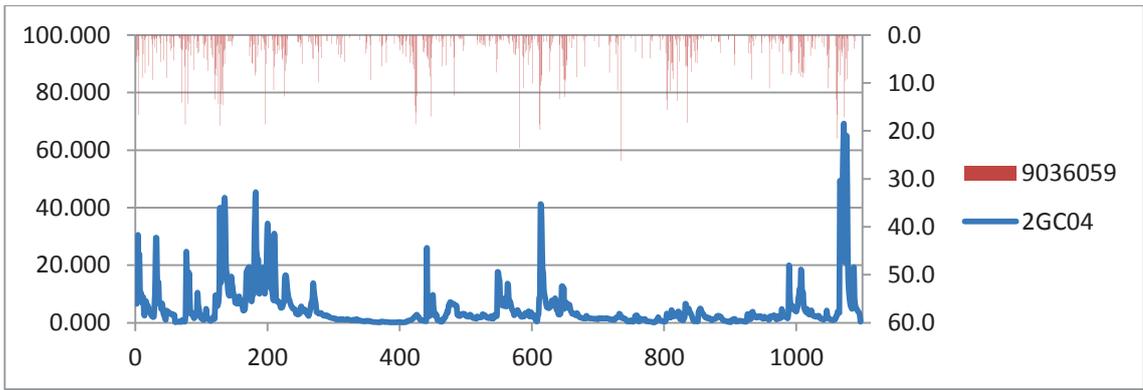
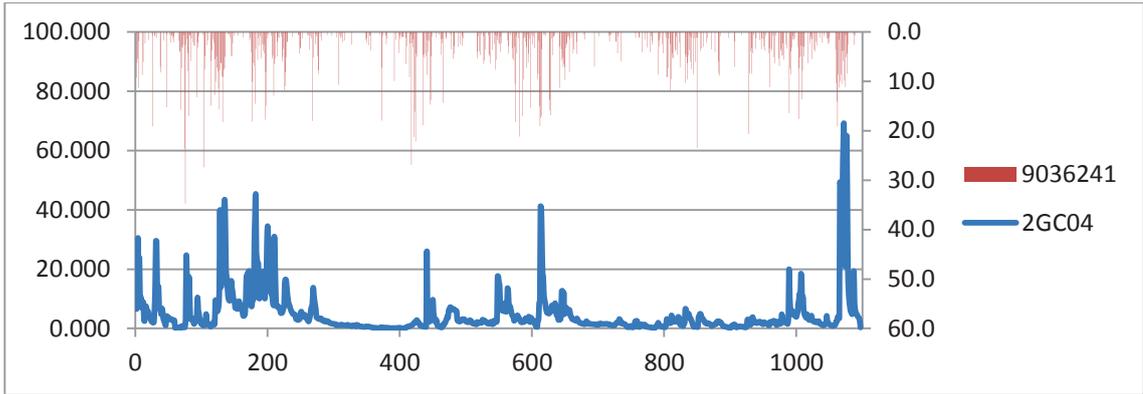
(4) River station 2GB0708 (outlet of subbasin 3):



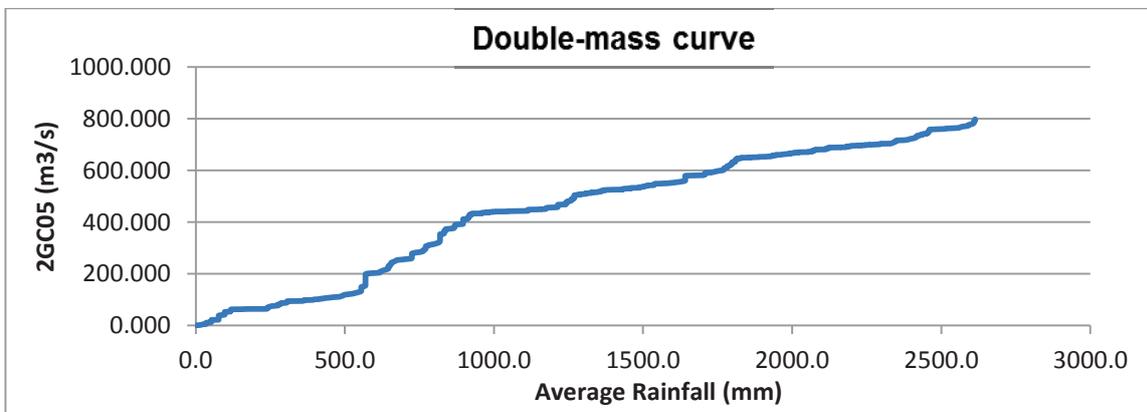
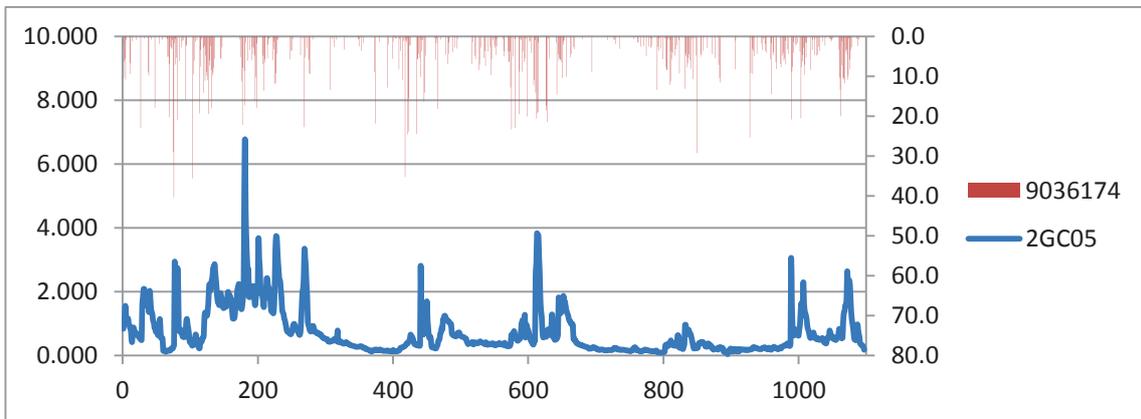
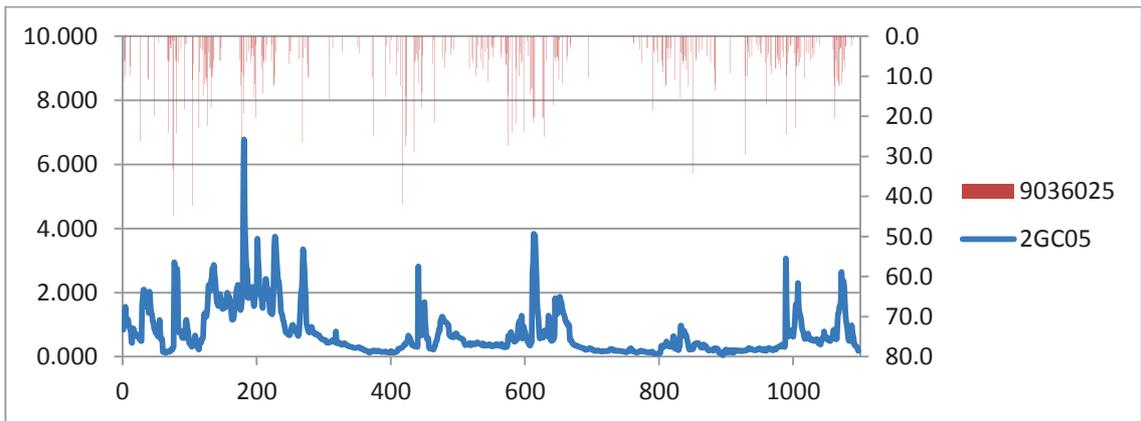
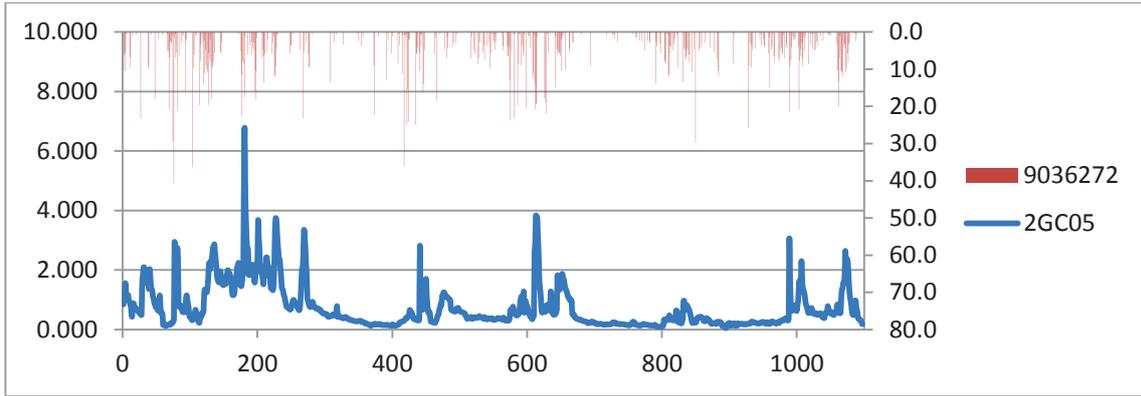


(5) River station 2GC04 (outlet of subbasin 5)

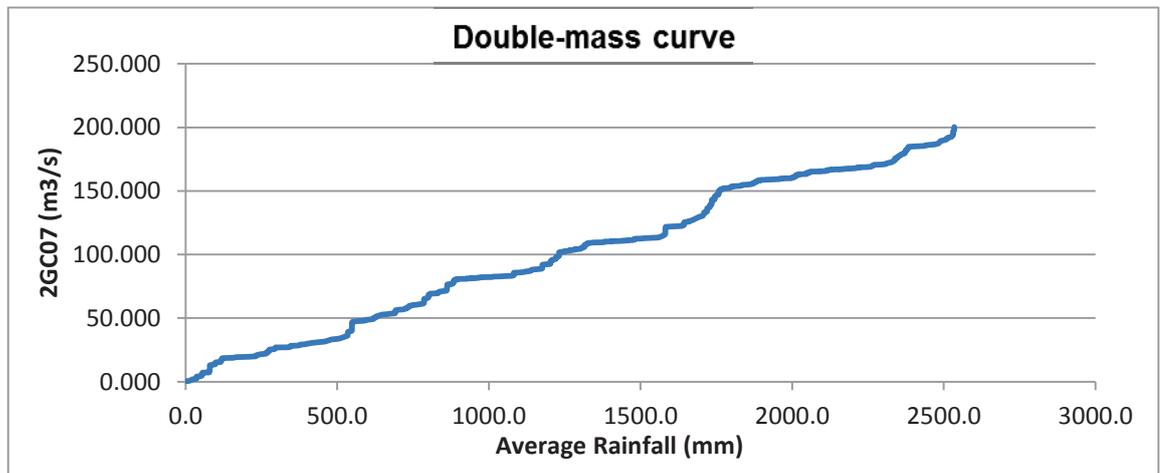
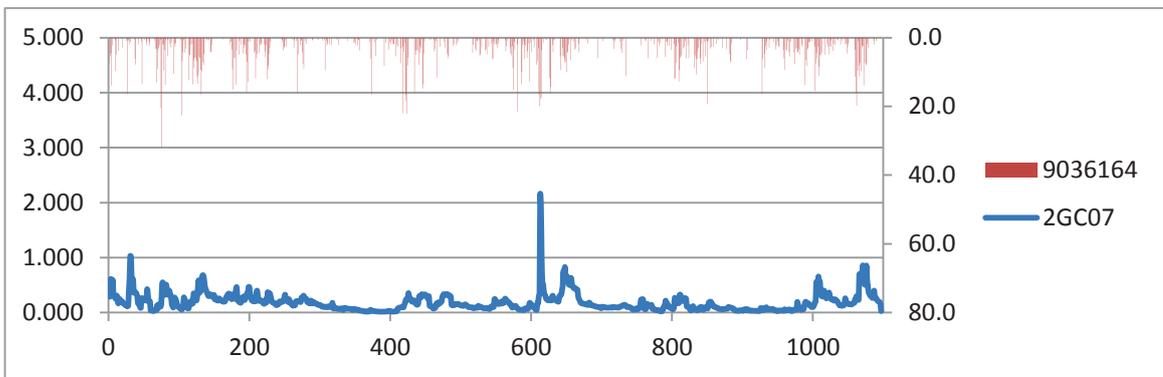
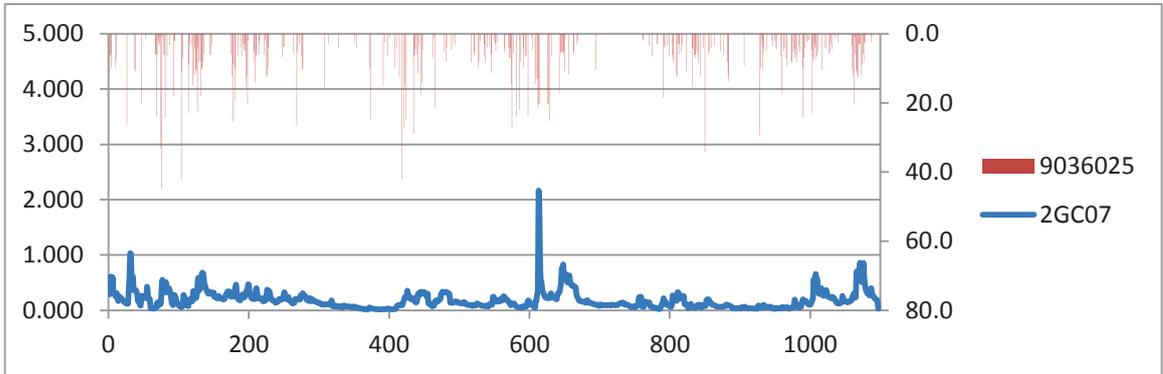




(6) River station 2GC05 (outlet of subbasin6):

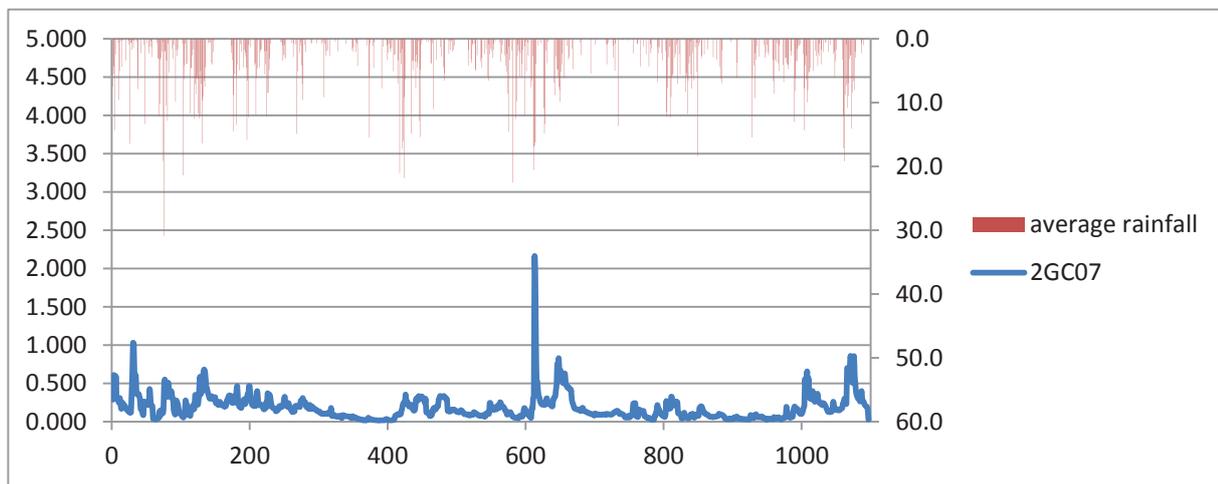
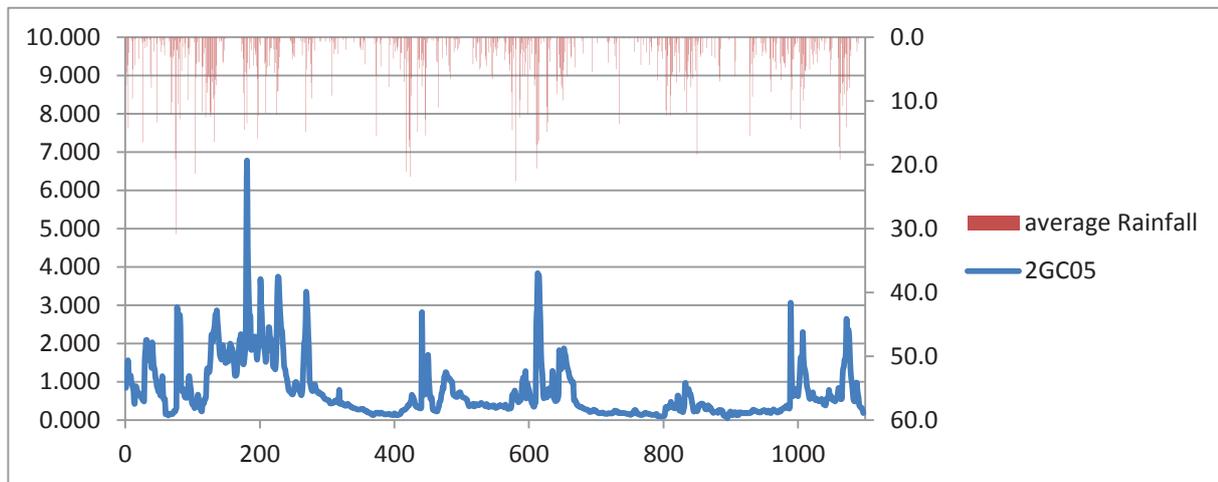


(7) River station 2GC07 (outlet of subbasin7):

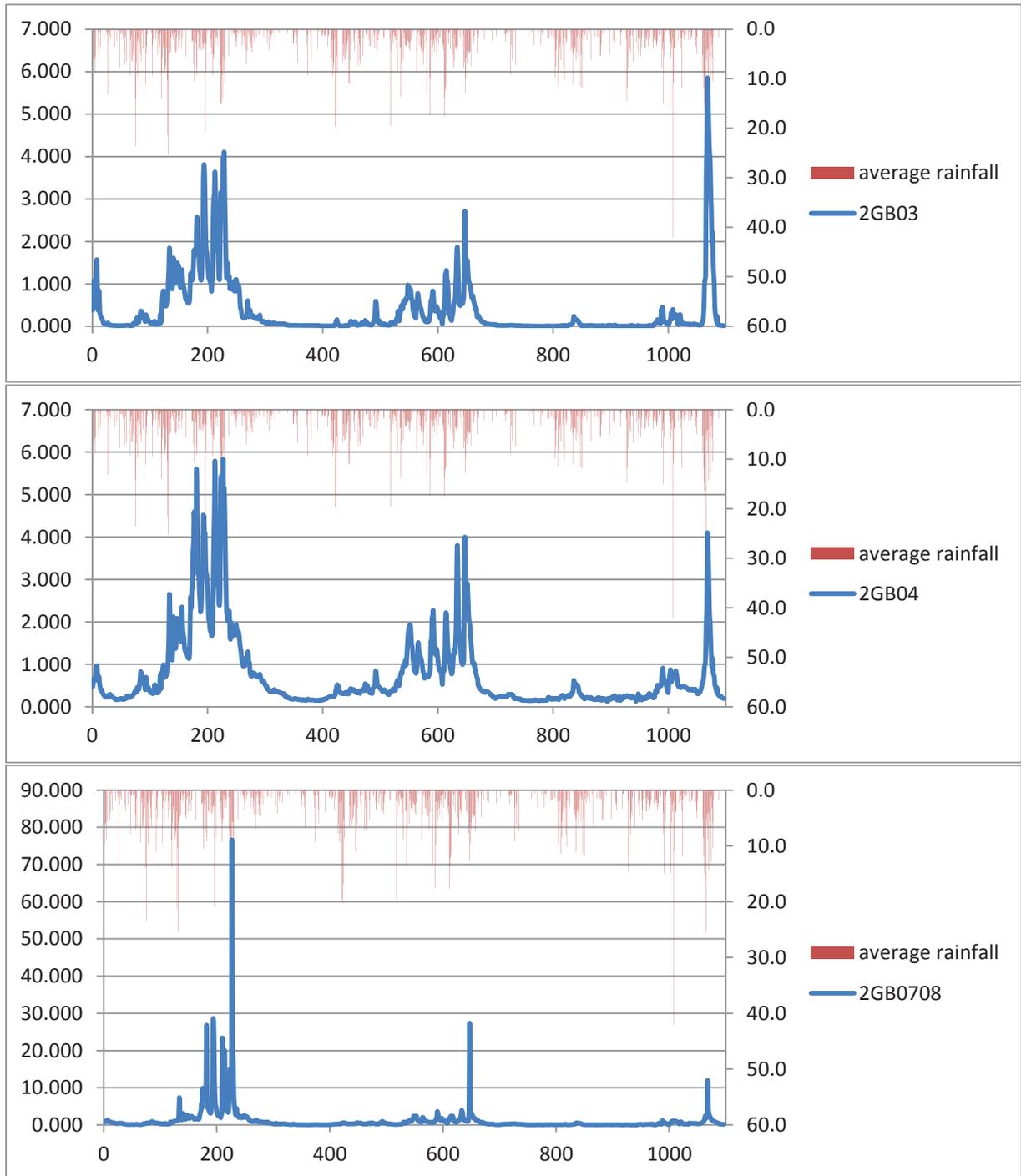


9.2. APPENDIX 2

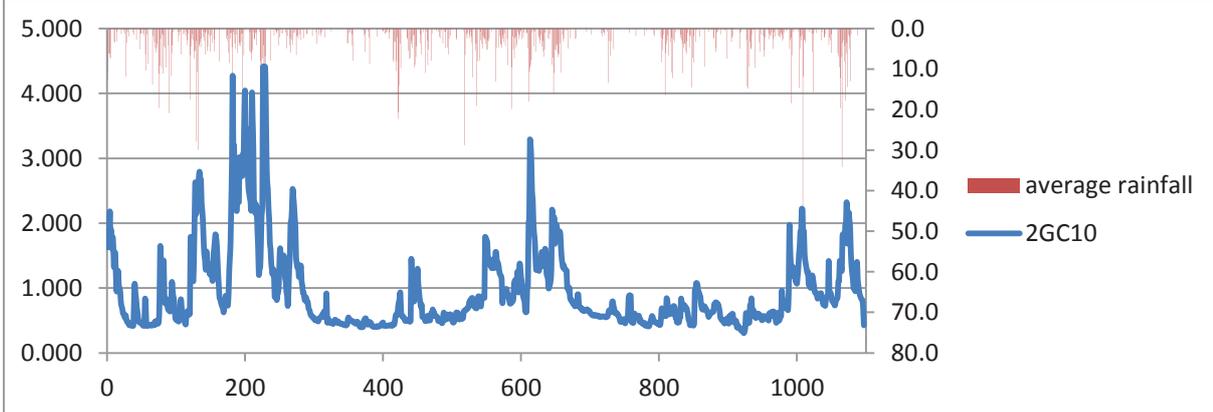
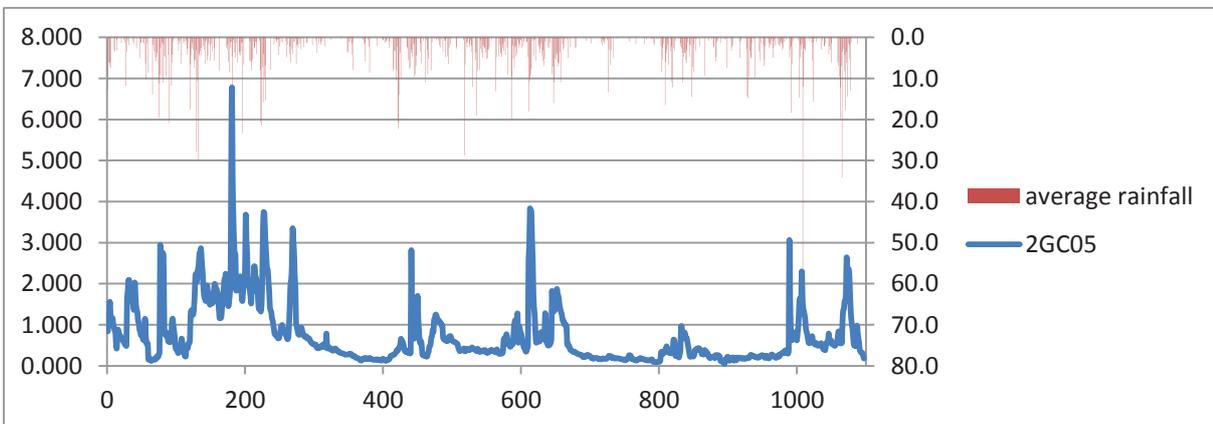
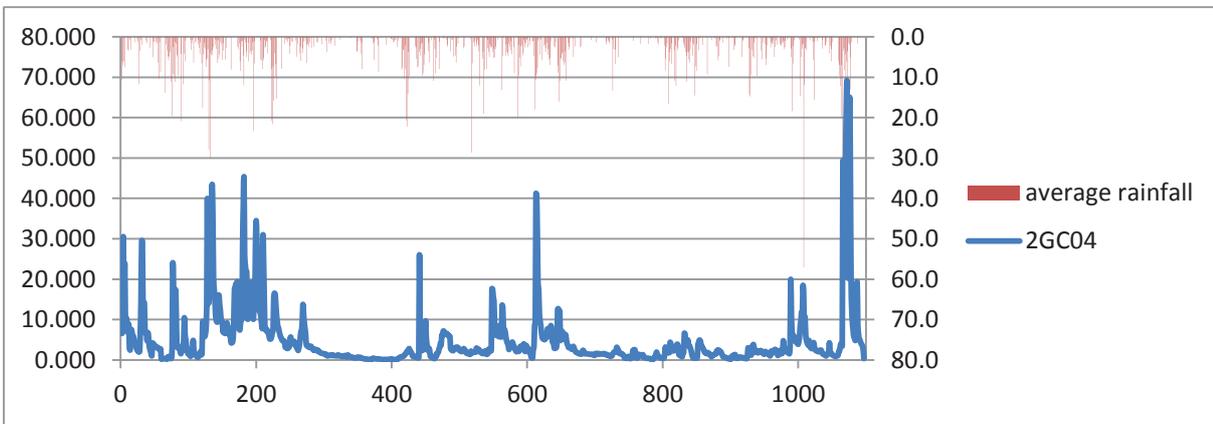
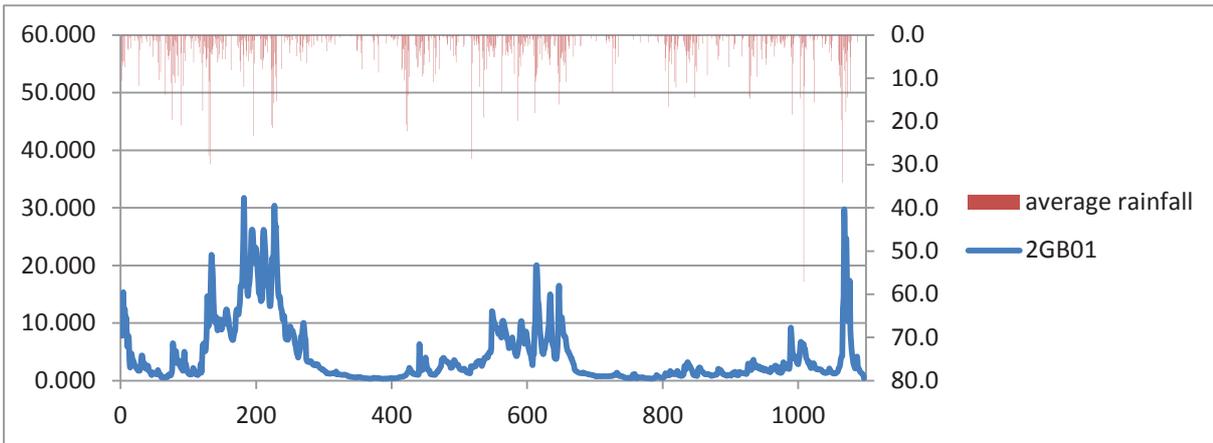
The comparison figures for average rainfall and discharge stations in part A.



The comparison figures for average rainfall and discharge stations in part B.



The comparison figures for average rainfall and discharge stations in part C.



9.3. APPENDIX 3

