

UNESCO-IHE INSTITUTE FOR WATER EDUCATION



Flood Modeling and Forecasting for Awash River Basin in Ethiopia

ABRAHA ADUGNA ASHENAFI

MSc Thesis WSE-HI.06-02
March 2006

UNESCO-IHE
Institute for Water Education





Flood Modeling and Forecasting for Awash River Basin in Ethiopia

Master of Science thesis
by

Abraha Adugna Ashenafi

Supervisors

Prof. R.K. Price, PhD, MSc
S.J.Van Andel, MSc

Examination committee

Prof. R.K. Price, PhD, MSc (UNESCO-IHE chairman)
Ir.S.Velickov, PhD, MSc (SKF)
Ir.A.Jonoski, PhD, MSc (UNESCO-IHE)
S.J.van Andel, MSc (UNESCO-IHE)

This research is done for the partial fulfillment of requirements for the Master of Science degree
Hydroinformatics specialization
At UNESCO-IHE Institute for Water Education, Delft, the Netherlands

Delft
March 2006

The findings, interpretations and conclusions expressed in this study do neither necessarily reflect the views of the UNESCO-IHE Institute for Water Education, nor of the individual members of the examination committee, nor of their respective employers.

Abstract

Nowadays, due to the climate change, urban development and other human interventions, flood is still a major challenge to humanity. Awash River basin is the most developed and exploited of all Ethiopian basins and it has massive economic significance for the development of the country. Most part of the basin is situated in flood prone areas and would be subjected to inundation unless some means of flood control strategies protect it. After the construction of the Koka dam, the features of the basin have changed in large extent. Heavy rainfall in the upstream catchments causes large inflow to Koka reservoir. At times, where the water level is already close to the maximum water level, the excess water must be discharged through the spillways in order to prevent overtopping of the dam. The discharge through the spillway and the tributaries discharge downstream of the reservoir have alternately been reported to cause flooding in the downstream areas of Koka reservoir in which irrigation development and settlements are found.

In recent years, some infrastructure plans to prevent the downstream areas from flooding and to increase the flood control are proposed and implemented by the government like constructing long dikes along the river. Fully implementation of the plan, constructing these measures, is delayed due to doubt on real cause of the problem of flooding in addition to a need of huge investments. The concern of the present research is to identify causes of flooding and to develop flood management strategies to reduce disaster.

For this reason, a MIKE 11 1D hydrodynamic model is built downstream of the reservoir to identify the causes of flooding by providing different boundary conditions and compare and identify critical sections on the reach using five years of measured stream flow data. The results showed that both the discharge from Koka and tributaries cause the flooding. Therefore, operation rule scenarios of the reservoir and downstream flood forecasting models were developed.

Result shows that without endangering the power supply, peak discharge reductions up to $300 \text{ m}^3/\text{s}$ can be achieved by advance release

Flood forecasting on the main tributary with a lead-time of 3 days for warning purpose using ANN and M5 model tree was developed. It shows that discharge peaks can be forecasted several days in advance. However, but with the present models, some peaks are underestimated or lagged. Further research can improve the accuracy of the models.

The overall experimental result showed that non-structural measures like reservoir management and flood forecasting could reduce flooding damage of farmlands and industries in Awash basin. A joint application of structural measures and non-structural measures could solve the flooding problem.

Key words: Awash basin, ANN, Koka reservoir, M5, MIKE 11, Flood control, flood forecasting and flooding

Acknowledgment

First and for most, I would like to acknowledge the tireless efforts of my Supervisors Prof.R.K.price and Schalk Van Andel without whose professional guidance, invaluable comments, discussion, constant inspiration and encouragement that all inspired me to work continuously and enthusiastically throughout this study. For them I hold respect and admiration.

I am very grateful to Dr.Ir.Andreja Jonoski for his help and guidance during the research proposal and data collection.

Also, I would like to thank the staff members of the core hydroinformatics and knowledge management, for having confidence in me while giving opportunity for doing the MSc. research through UNESCO-IHE- water mill project.

I would like to express my sincere thanks to my Employers, National regional state of Tigray, and most particularly the Tigray water resource and development commission. I hope that the knowledge and experience I gained during this period will serve my country in the best possible way.

Many thanks go to Berhanu Bekele, Fasil Abebe and Befekadu Getachew for providing the data and for keeping in touch during the research.

I want to say thanks to UNESCO- IHE staff and friends in Hydroinformatics (hydroinformaticians) who offered me technical and moral support during the study.

Last but certainly not the least; I would especially like to thank my family whose support, for remembering me in all their prayers and affection encouragement me to complete this research. Finally, my last words of appreciation and respect are reserved to my wife Helen Debebe, my son Yeshak Abraha and my daughter Solyana Abraha for their patience, encouragement, love and understanding they gave me all this period.

This piece of work is dedicated to the
most important persons in my life:
HD, YA and Solyana Abraha

Above all

I owe everything to the Almighty **God** who gave me strength and good health throughout entire period of my study. There is no loss in believing in him

Abraha Adugna

UNESCO-IHE, Delft, The Netherlands

March 2006

Table of contents

Abstract.....	iv
Acknowledgment	v
List of Figures	2
List of Tables	4
List of Abbreviations	5
1. Introduction.....	6
1.1 Problem description	6
1.2 The study area	6
1.3 Poverty and water resource strategy in Ethiopia and its impact in achieving millennium development goals (MDG)	9
1.4 Awash basin water management authority	12
1.5 Objective of the study	13
1.6 Methodology	14
1.6 Outline of the thesis	15
2. Flood Modeling and Management	16
2.1 Types of flooding	16
2.2 Flood Management	16
2.3 Flood Modeling.....	20
2.4 Parameters for data analysis and model performance analysis.....	28
3. Data inventory and analysis	33
3.1 Data inventory.....	33
3.2 Data preparation and analysis	37
3.3 Basic characteristics of the Awash basin	42
4. Flood analysis	44
4.1 Purpose and analysis set up.....	44
4.2 MIKE 11 hydro-dynamic model.....	46
4.3 Results analysis	52
4.4 Conclusion and Discussion	55
5. Reservoir operation.....	56
5.1 Analysis set-up.....	56
5.2 Reservoir inflow time series	56
5.3 Reservoir model	59
5.4 Reservoir release scenario's.....	61
5.5 Conclusion	68
6. Flood Forecasting.....	69
6.1 Set up analysis.....	69
6.2 Data preparation.....	69
6.3 Inputs selection	71
6.5 Flood forecasting using ANN	78
6.6 Error analysis	80
7. Conclusions and Recommendations	82
7.1 Conclusions.....	82
7.2 Recommendations.....	82
8. References.....	84

List of Figures

Figure 1.1 Awash basin and its classification in zones.....	7
Figure 1.2 Awash basin with irrigation schemes.....	13
Figure 2.1 Procedure of a Flood forecasting system.....	19
Figure 2.2 Machine learning (Solomatine.2005).....	23
Figure 2.3 Model tree (Solomatine, 2003).....	24
Figure 2.4 M5 model tree sample at MelkaSedi station.....	25
Figure 2.5 ANN's classification (Solomatine, 2000).....	27
Figure 2.6 Multi-layer perceptron network sample.....	27
Figure 3.1 the study area with rainfall stations.....	33
Figure 3.2 Study area with its discharge stations.....	34
Figure 3.3 maximum, mean and minimum temperature.....	35
Figure 3.4 soil map of upper basin.....	36
Figure 3.5 proportion of soil in the upper basin (data taken from FEWS-NET).....	36
Figure 3.6 Double mass curve of station MelkaHombole and MelkaKunture in upper basin.....	38
Figure 3.7 Flow between station Nurahera and upstream flows (Q1).....	39
Figure 3.8 Flow between station Methara and upstream flows (Q2).....	39
Figure 3.9 Flow between station Awash station and upstream flows (Q3).....	40
Figure 3.10 Example of missing values, 1991.....	42
Figure 4.1 Awash study area with potential cause of flooding: Koka reservoir and tributaries.....	44
Figure 4.2 The Awash River basin network (vertical and horizontal grids are in UTM). 49	49
Figure 4.3 Confluence between Awash main, welencheti, and arba (vertical and horizontal grids are in UTM).....	49
Figure 4.4 Example of a cross-section used to describe the awash main river (horizontal and vertical units are in meter).....	50
Figure 4.5 Model versus measured result at MelkaSedi station 1990-1993.....	51
Figure 4.6 Model versus measured result at MelkaSedi station in 1994.....	52
Figure 4.7 Water level result only using release discharge from the reservoir around MelkaSedi station.....	53
Figure 4.8 Water level profile only tributaries discharge.....	54
Figure 4.9 Water level profile with combined discharges, 1990, around MelkaSedi stations.....	54
Figure 5.1 GR3J model result at MelkaHombole station 40 km from the reservoir.....	58
Figure 5.2 Inflow Time series from inverse modeling and measured outflow and losses from Koka.....	59
Figure 5.3 idealization of reservoir operation.....	59
Figure 5.4 Different scenarios (minimum of August as reference and then 1.1* minimum etc).....	62
Figure 5.5 release schedule scenarios checked with minimum volume of water level for power generation.....	62
Figure 5.6 Flow chart for advance release scenario.....	63
Figure 5.7 five-year advance released schedule.....	64
Figure 5.8 a) Mike 11 water level profile with advance release, 1990, around MelkaSedi.....	64
Figure 5.9 Discharge time series around MelkaSedi with advance release.....	67
Figure 6.1 Effective rainfall versus discharge at MelkaSedi station.....	69
Figure 6.2 the model tree generated using full year data.....	74

Figure 6.3observed versus predicted in training using four (1990-1993) year data	76
Figure 6.4 Observed versus predicted in verification data(1994).....	76
Figure 6.5 Observed versus predicted in training using wet season data (1990-1993)	77
Figure 6.6 Observed versus predicted in test data using wet season data (1994).....	77
Figure 6.7The model architecture	78
Figure 6.8 Training prediction out put.....	79
Figure 6.9 Verification prediction out put	79
Figure 6.10 Scatter plot of observed and predicted	80

List of Tables

Table 1.1 Division of Awash basin.....	7
Table 1.2 Flood events at the Awash Basin and the estimated damage (DPPC report)	9
Table 3.1 soil unit and its properties in the upper Awash.....	36
Table 3.2 Comparison of flow record at Melka Hombole and Melka Kunture.....	41
Table 3.3 The characteristic of discharges in m ³ /s and area average rainfall in mm upstream of koka reservoir and downstream	43
Table 4.1 The hydrodynamic variables involved in the model set up	50
Table 4.2 The manning coefficient fixed with the model calibration.....	51
Table 5.1 Parameter values for calibration	57
Table 6.1 The correlation coefficient of relevant variables and time lags with predicted discharge of full year data downstream of koka reservoir	70
Table 6.2 The correlation coefficient of relevance variables and time lags with predicted discharge of wet season data downstream of koka reservoir	70
Table 6.3 The hydrological characteristic of 5 year time series	71
Table 6.4 The selected training and testing data type	71
Table 6.5 The corresponding correlation coefficient of input variables and their corresponding time lags with the predicted discharge.....	73
Table 6.6 the predicted result test of model trees using full year data input	76
Table 6.7 The predicted result test of model trees using wet season data input.....	77

List of Abbreviations

ANN Artificial Neural Network

DPPC Disaster Prevention and Preparedness Commission

EFAP Ethiopian Forestry Action Program

FAO Food and Agricultural Organization

GDP Gross Domestic product

GNP Gross National Product

HDI Human Development Indexes

HIV/AIDS Human Immune Deficiency Disease

ITCZ Inter-Tropical convergence Zone

MDG's Millennium Development Goals

MoA Ministry of Agriculture

MoWR Ministry of Water Resource

SDPRP Sustainable Development and Poverty Reduction Programme

UNDP United Nation Development Programme

WRMP Water Resource Management Programme

1. Introduction

1.1 Problem description

Since the ancient time, flooding has remained a major challenge to humanity. Flooding is one of the most dramatic interactions between man and his environment, emphasizing both the sheer force of the natural events and man's inadequate efforts to control them. Although man has been responding to floods since long before the day of Noah, they are still not fully understood by the hydrologists, engineers, planners and politicians and especially by the common people for whom floods are an unexpected, often inexplicable and always traumatic experience. The Awash River basin, with approximately 110,000km² contains flood prone areas. In 1960 a dam was constructed as a multipurpose facility for generating power, irrigation and flood control in the basin. The construction of Koka dam provides flood protection for the farms and other economical infrastructure in the valley by absorbing floods from the upper part of the basin. Downstream dikes prevent flooding of the adjoining farms from high flow in the river.

Integrated water resources management in Ethiopia is not in an advanced stage. While the country's water and land resources endowments are abundant, very little has been accomplished in the way of proper exploitation of these resources for economic benefit of the people. In fact, it is a paradox that the country has been the victim of food insufficiency and famine in recent years. Flood management being, part of integrated water resource management, has not been treated as one of the means to alleviate poverty in development plans. The only flood control and management activity being carried out in the country is in the Awash River basin using structural measures i.e. constructing long dikes along the riverbank downstream of the Koka reservoir.

In the present day, with the advances in modern technologies for water management, flood control may not only rely on these structural measures, but also use different kinds of modern water management approaches to minimize occurrence of flood and the resulting damages.

The concern of the present study is to develop a flood management strategy and a reliable flood-forecasting model for early warning in the Awash River basin based on the result of flood analysis and analyzing real cause of flooding in the basin downstream of Koka dam.

1.2 The study area

Location

The Awash basin (see Figure 1) is part of the Great Rift Valley in Ethiopia located from 8.5⁰N to 12⁰N. It covers a total area of 110000km² of which 64000km² comprises the western catchment, which drains to the main river or its tributaries. The

remaining 46 000km², most of which comprises the so-called Eastern catchment, drains in to a desert area and does not contribute to the main river course(ADF2003 proposal report).

The Awash basin has been traditionally divided into four distinct zones. These are; Upper basin, Upper valley, Middle valley and Lower valley (see Figure1).

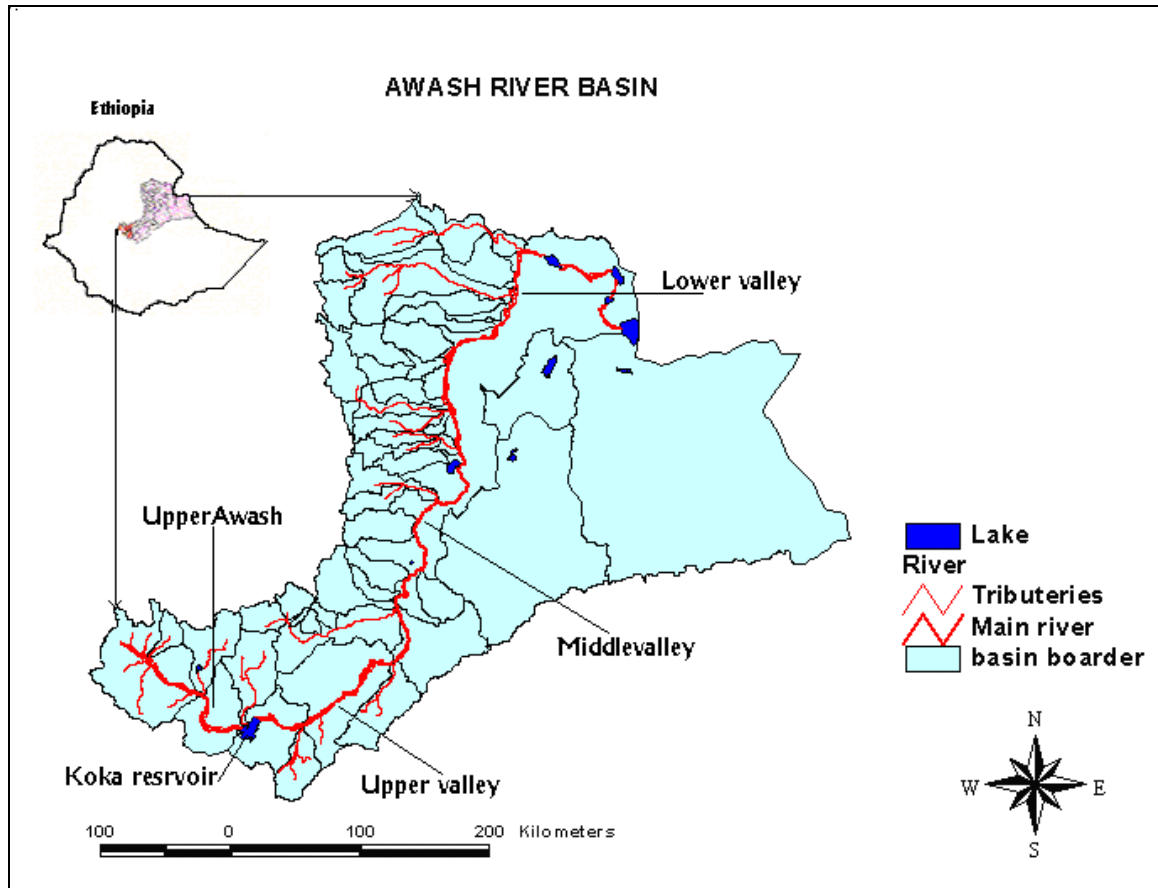


Figure 1.1 Awash basin and its classification in zones

Table 1.1 Division of Awash basin

Designation	From	To
Upper basin	Head waters	Koka dam (lake)
Upper valley	Koka dam	Awash station
Middle valley	Awash station	Gewane
Lower valley	Gewane	Lake Abe

The Awash basin covers the central and northern part of the rift valley and is bounded to the west, southeast and south by the Blue Nile, the rift Lakes and Wabi Shebele Basins respectively. The study is concentrated on the upper Awash, upper valley and middle valley area of the basin due to the data availability and vulnerability of the developed irrigation schemes for flooding.

Physiography and Geology

The topography of the Ethiopian Plateau is generally flat with elevations ranging from 2,500 m down to 2,000 m. A series of fault scarps leads from the plateau to the floor of the Rift Valley, which slopes northeast from the elevation of nearly 2000 m at Lake Ziway (seasonal flooded lake above koka) to less than 400 m where it becomes the Afar Triangle(lower valley). The flat floor of the Rift Valley is frequently broken by fault scarps and the effects of Pleistocene and Holocene volcanic activity. On the southeastern side of the valley fault steps lead to the Ogaden Plateau. This slopes southeast; draining towards the Indian Ocean and for that reason the Awash receives most of its tributaries from the western, Ethiopian Plateau, side.

Climate

The climate of the Awash Basin comes under the influence of the Inter-Tropical Convergence Zone (ITCZ) This zone of low-pressure marks the convergence of dry tropical easterlies and the moist equatorial westerlies. The seasonal rainfall distribution within the basin results from the annual migration of the ITCZ. In March, the ITCZ advances across the Basin from the south, bringing the small or spring rains. In June and July it reaches its most northerly location beyond the Basin that then experiences the heavy or summer rains. It then returns southwards during August to October, restoring the drier easterly air streams, that prevail until the cycle repeats itself in March.

The annual rainfall distribution resulting from this cycle is exhibited most clearly in the two distinct rainy periods, which are characteristic of the northern plains of the basin. Moving southwards the more prolonged exposure to the moist air-stream is evident in the tendency for the two dominant rainy periods to merge into a contiguous distribution. On the high plateau to the west of Addis Ababa, the rainfall distribution shows a continuous increase from the spring rains to the summer peak rainfall. The distribution of rainfall over the highland areas is modified by orographic effects and is significantly correlated with altitude (ADF 2003 proposal)

The mean annual rainfall varies from about 1600 mm at Ankober, in the highlands northeast of Addis Ababa (west limit of the basin) to 160 mm at Asayita on the northern limit of the Basin. Addis Ababa receives 90% of its annual rainfall between March and September. At Dubti, the same overall proportion is received during the two rainy periods, distributed 30% and 60% respectively. The mean annual rainfall over the entire western Catchment is 850 mm and over the headwaters of the Awash, as gauged at Melka Hombole it is 1216 mm. Over the Eastern Catchment, the mean annual rainfall is estimated to be 456mm. The annual and monthly rainfalls are characterized by high variability.

Mean annual wind speed at Koka average 1.2 m/s, the windiest months being June and

July with mean monthly values of 1.9 and 1.6m/s respectively. At Amibara (downstream of koka), the mean monthly wind speed values in June and July are over 2 m/s.

Recent Flooding Problem

The Awash basin and the flooding danger have been inseparable for the past decade. Huge destruction occurred in 1993 and 1996. According to the Ethiopian Disaster Prevention and Preparedness Commission (DPPC) estimation the number of people affected at different flooding events is shown in Table 1.2

Table 1.2 Flood events at the Awash Basin and the estimated damage (DPPC report)

Zone/wereda	Flood period	Remarks
Afambo(lower valley), Dupti(middle valley), Assaita(middle valley)	1993 1994 1995	144,400 154,900 145,700 people were affected
Bure-Mudiatu(middle valley), Gewanie(middle valley)	1993 1994 1995 1996	Not estimated Not estimated Not estimated Over 75,000 people were affected
Dullecha(middle valley), Ambira(Upper valley)	1996	Over 5million American dollar damaged

1.3 Poverty and water resource strategy in Ethiopia and its impact in achieving millennium development goals (MDG)

Large and landlocked, Ethiopia is the second most populous country in sub-Saharan Africa, and the most impoverished. Only seven of 177 other countries have lower human development indexes (HDI). 44% of Ethiopia's population lives below the national poverty line, and repeated droughts have contributed to regular food shortage and famine. Besides these problems natural disaster like flood, unemployment and HIV/AIDS are other challenges in Ethiopia (UNDP-Ethiopia report).

Based on 1999/2000 household incomes and expenditure survey report, the incidence of poverty has not improved in Ethiopia in the past five years. The proportion of the people in absolute poverty declined only marginally from 45.5% in 1995/1996 to 44.2% in 1999/2000. Furthermore, reports from subregional level consultation under the sustainable development and poverty reduction programme (SDPRP) suggest that poverty might have worsened since 1999/2000. The incidence of poverty in Ethiopia is considered to be one of the highest in the world. GNP per capita income is estimated at US\$110 with 55% of the population below the poverty line. In some regions, the population below the poverty line is as high as 85%. This group consists of small farmers in rural areas, landless, and unemployed workers.

Agricultural development, in the Ethiopian context, aims at the production of sustainable sufficient food for consumption as well as an increased supply of agricultural commodities for export. Ethiopia is well endowed with ample untapped natural resources but till now only 6% of potential irrigable land is developed and only 3% of its water resources in use. Agriculture also plays an important role to the economy. It accounts for about 50% of GDP, generating about 90% of export earnings, providing employment for 85% of the population and supplying 70% of raw materials for the countries Agro-industries (MoA, 2000). However, despite its

potential and efforts made in developing agriculture, the sector has remained at subsistence level for decades.

The idea of producing food security was a total failure. The national development plan is based on the strategy called ‘Agriculture led industrialization’ and aims at boosting agricultural productivity and improve the rural standard of living, which in turn will increase the demand for goods and services and lead to further industrial development.

The food deficit coupled with the alarming population increase and limited financial capacity improvement would make the problem of raising the food supply more critical in the country. The problem could even be more severe especially in the arid and semi arid regions of the country, as the marginal and erratic rainfall in these areas renders rain fed agriculture generally unreliable. Under such circumstances, irrigation development may be taken as a primary means for realizing a sustainable agricultural production. Irrigated agriculture may provide a degree of self-sufficiency in food, or at least contribute to ensuring national food security, raising rural population living standard, creating employment opportunities, and reducing urbanization pressure.

Applying irrigation practice is an important means of achieving food self-sufficiency in the country. Besides, it could also be used as a means of income if they produce surplus food. Hence, water resource development has been the focus of the government to address the twin problems of food insecurity and water scarcity. In dry land areas, irrigation intervention is believed to be improving the productive potential from very low to better. Yield increases by 100 to 400% have been recorded through irrigation (FAO, 1997).

Food insecurity and health problems are the main features of poverty. However, the government of Ethiopia has undertaken actions to address the poverty situation in the country and has thus adopted a twin-approach of promoting human development and broad based economic development, in line with achieving the Millennium Development Goals. The government of Ethiopia is also tackling low productivity and is making efforts to the aid to farmers. The government introduces irrigation development practices recognizing that it is a primary means of poverty alleviation in the country with in the specified period to 2015.

Irrigation development can meet its objective if it is managed properly. In line with the agriculture led industrialization policy, the water sector strategy of Ethiopia gives emphasis on the specific objective to combat and regulate floods in the Awash River basin in order to manage and practice irrigation development. It recognizes the potential of the basin to become a huge food basket for Ethiopia through sustainable flood mitigation, prevention and other practical measures. With persistent floods, pre-requisites to sustainable land use for agriculture in Ethiopia will include effective flood control programmes that incorporate engineering structures such as dikes, dams and terraces (structural measures) and non-structural measures such as flood forecasting and warning to decrease the disasters in terms of lives and property.

Watershed management programmes such as soil conservation, reforestation and the protection of fragile ecosystems are needed. To address environmental problems in

general and poverty alleviation in particular, Ethiopia has prepared environmental policies, conservation strategy and draft legislation. It has also put in place relevant institutional arrangements to plan and implement programmes, both at the federal and regional level.

Environmental Status

The deteriorating vegetation cover in Ethiopia and the resultant negative impact on soil erosion have long been causes for concern due to their impact on water flow in the river systems of the country. National forest cover declined from an estimated 35% in the 1950s to less than 3% in 2000. Presently, flooding is being accentuated by land degradation, soil erosion and silting up of river courses and reservoirs. In spite of the decreasing national forest resources, the demand for wood remains high due to the reliance on the resources for energy, building construction and on-farm food processing.

At present, wood removal from national forests is more than twice the rate of replacement through plantation establishment. According to the Ethiopian Forestry Action Program (EFAP), the national wood biomass demand for 1992 was estimated to some 47 million m³ of which 45-million m³ was in the form of fuel wood. The supply in the same year was only 14.4 million m³ indicating a deficit of some 32.6 million m³. Expanded afforestation programmes are thus needed to meet wood security and protect Ethiopia' fragile hillsides which have a direct influence or aggravate flooding downstream of Koka reservoir.

Water Resources Management Policy

In 1998, the MoWR issued its Water Resources Management Policy (WRMP) that sets the guidelines for water resources planning, development and management. The overall goal of the policy is to enhance and promote all national efforts towards efficient, equitable and optimum utilization of the available water resources of Ethiopia for significant socio-economic development on sustainable basis. The specific objectives of the policy are to:

- (i) enhance water resources and use the overall aquatic environment on sustainable basis;
- (ii) develop the water resources of the country for economic and social benefits of the people on equitable and sustainable basis;
- (iii) allocate and apportion water resources, based on comprehensive and integrated plans;
- (iv) manage and combat drought through efficient allocation, redistribution, transfer, storage and efficient use of water resources;
- (v) combat and regulate floods through sustainable mitigation, prevention, rehabilitation and other practical measures; and
- (vi) conserve, protect and enhance water resources and overall aquatic environment. (MoWR, 1998: report)

1.4 Awash basin water management authority

In the Awash River basin there is a semi autonomous organization called the Awash basin water management board, which is responsible for managing, coordinating and planning of integrated water resource in the basin particularly in the upper and middle valley. The board, which consists of representatives from the ministry of water resource, ministry of agriculture and the Ethiopian power and light authority, has manage the overall management in the basin and prepare disaster Prevention and Preparedness plan. As mentioned in the introduction the Koka reservoir plays an important role in the basin and the country as well; therefore the committee tries to balance power generation, flood control and irrigation development taking the primary optimized objective as power generation. This is one of the major tasks of the authority. The authority together with the people living around flood plain to control flooding of the farms and industries by constructing dikes along the river Awash, and it is felt that this needs huge investment.

The Millennium Development Goal (MDGs) are the world time bound and quantified targets for addressing extreme poverty in its many dimensions income poverty, hunger, disease, lack of adequate shelter, and exclusion-while promoting gender equality, education, shelter and security (Reader,2005).

The MDGs are the most broadly supported, comprehensive, and specific poverty reduction targets the world has ever established, so their importance is manifold. Here one important thing, which is not included or even not considered, in any of the MDG is natural disasters, which has a direct and indirect influence in the almost all Goals and targets.

The country under the millennium development campaign will work on poverty alleviation and sustainable development by formulating and establishing policy and strategies considering that the flooding problem in the Awash basin has a direct and indirect impact in achieving the millennium development goals. First, flooding directly relates to health problems caused by water born diseases through pollution of the water wells. Besides this, reducing the flooding of the farms has a direct impact in hunger reduction and improving the income level of the country in general and rural people in particular, which is the first goal of the MDGs (poverty alleviation).

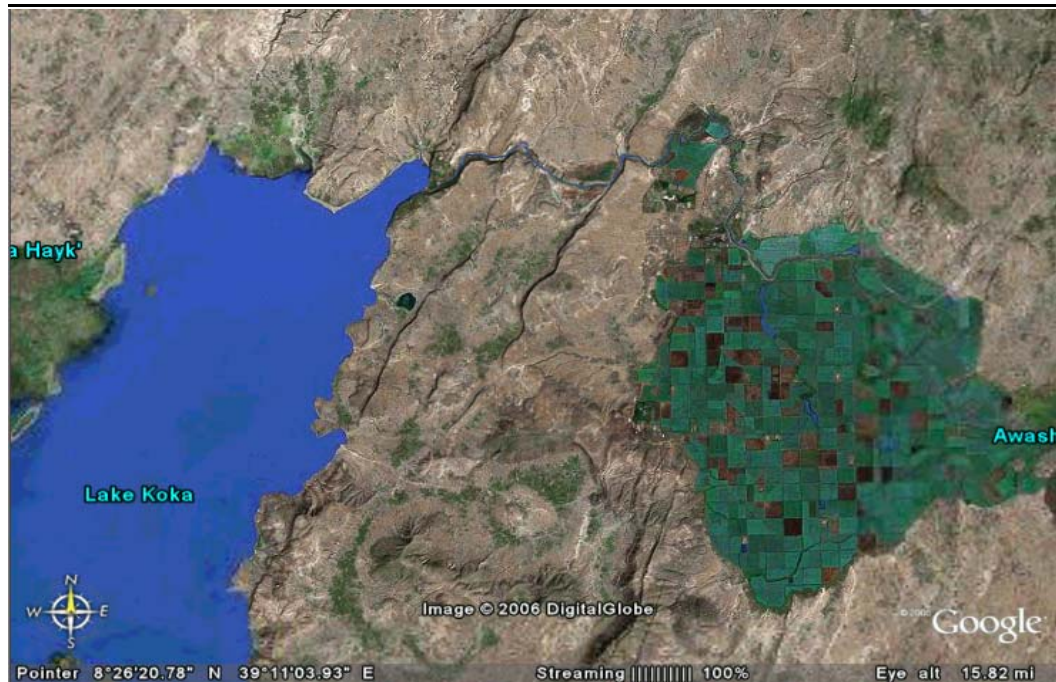


Figure 1.2 Awash basin with irrigation schemes immediately below Koka reservoir

1.5 Objective of the study

The main objective is to develop and provide flood management strategies for the Awash basin, in order to alleviate poverty in the region taking into account sustainable development.

The study carried out the following specific objectives:

- To identify causes of flooding downstream of Koka reservoir by doing flood analysis using MIKE 11 1D hydrodynamic modeling
- To develop advance release scenarios for Koka reservoir to reduce flooding risk downstream, using a reservoir routing model and MIKE
- To apply the M5 model trees algorithm and Artificial Neural Network for flood forecasting in Middle basin at MelkaSedi station to give warning to downstream property and save lives
- To give recommendations to practioners on how to combine non structural measures and structural measures to improve flood management

1.6 Methodology

The daily data of Awash River basin and reports about flooding in the basin is used as a case study, and the following study methods and techniques were applied in this study:

1. Literature study on flood analysis and flood modeling and forecasting techniques
2. Data preparation
Data preparation includes the following
 - Study of the modeled system
 - Collection of the relevant data
 - Dealing with missing data
 - Dealing with inaccurate data
 - Dealing with lack of data
3. MIKE 11 1D model to identify critical cross-sections and the possible cause by doing flood analysis
4. Based on the flooding analysis results reservoir operation models and flood forecasting models are build
5. Data analysis to prepare the input and output data
 - Choice of data sets
 - Number of data set
 - Method of data divisions
 - Data processing
 - Choice of model inputs
 - Choice variables
 - Choice of lag times
 - Choice model architecture
 - Connection type
 - Choice of stopping criteria
6. Flood forecasting using M5 model trees and ANN(artificial neural network)

1.6 Outline of the thesis

Chapter 2 Flood modeling and management: contains a brief introduction to the basic theory, characteristics and application of flood modeling and flood forecasting in flood management and the application of data driven modeling in flood management. Some statistical concepts associated with data analysis are introduced as well.

Chapter 3 Data preparation and analysis: gives a description of the study area and data and then introduces the data analysis results, checking data consistency and reliability.

Chapter 4 Flood analysis: discusses the different reports drawn about the cause of the flooding in the basin, following with identifying causes of the problem by building MIKE 11 1D hydrodynamic model. The analysis is focused on the middle valley of the basin, which has affected the irrigable land.

Chapter 5 Reservoir operation: first gives the catchment response of the basin above the reservoir and then discusses different scenarios of advance releases from the reservoir to reduce flooding downstream, without affecting power generation or the main objective function of the reservoir.

Chapter 6 Flood forecasting: first gives brief introduction of the purpose of flood forecasting in flood management and warning systems. Then forecasting models are developed with model tree and ANN. Different data sets are used i.e. first continuous full year data is used for training and testing and then with the flood season data only to improve the prediction of the model and the result is improved marginally. This is done with model tree and artificial neural network with same data to see the prediction capacity of the two systems and their understandability with decision makers.

Chapter 7 Conclusion and Recommendations: the overall result of the study is summarized, and some recommendations for further work to improve the forecasting and to give a lasting solution to the problems are given

2. Flood Modeling and Management

2.1 Types of flooding

Definition: Flood is relatively high water levels caused by excessive rainfall, storm surge, dam break or tsunami that overtop the natural or artificial banks of a stream, creek, river, estuary, lake or dam. Flooding in the river can be defined as the inundation of a normally dry area caused by an increased water level in a watercourse or in a body of water. This type of flooding can be further distinguished as flash flooding and simply river flooding.

1) Flash flooding: A flash flood can be defined as a flood that threatens damage at a critical location in the catchment, where the time for the development of the flood from the upstream catchment is less than the time needed to activate warning, flood defense or mitigation measures downstream of critical location.

A flash flood is characterized by a sudden and massive increase of water quantity (e.g. caused by short intense bursts of rainfall commonly from thunderstorms) often combined with mud and debris flows, leaving endangered communities only limited time to respond.

2) River flooding is marked by relatively slowly rising water levels of main rivers and a gradual inundation of floodplains due to continuous long duration rainfall. River flooding is essentially a natural process that helps shape the landscape.

Flooding along a river or stream occurs when sufficiently heavy or prolonged rainfall produces runoff, which overflows the banks of the watercourse and overwhelms the ability of the land to drain the water effectively. This is aggravated when the ground is already saturated and when river channels become blocked by debris. Flooding may also be exacerbated by the changes in land use such as urbanization and changes in agricultural practices. Throughout history, flooding has threatened human life and property.

Flash flooding imposes a different and indeed more daunting challenge to the alertness of flood prone communities whereas the larger scope and longer duration of river flooding constitute a major challenge to the scale and endurance of disaster management arrangements (Maskey, 2004).

2.2 Flood Management

The problem of floods

Floods are one of the most dramatic interactions between man and his environment, emphasizing both the shear forces of natural events and man's inadequate effort to control them. According to Roy Ward (1978), in most of the cases floods is not natural disaster but natural phenomena and form part of the normally occurring range of stream-flow conditions (just as does drought, at the other end of scale). Since normally stream channels can carry only a fraction of the peaks of extreme floods flow through and over, or be stored on, the flood plains.

In flood conditions, therefore channels and their adjacent flood plains are complementary and inseparable and together form proper conveyance of floodwaters. However, flood disasters are man made because man has put himself at risk by developing flood plains for settlement, agriculture and industry and building roads, bridges and railways lines in floodable position. Because of this increasing intrusion in to flood plain resulting from ignorance and economic reasons, flood problems are getting worst all over the world in terms of damages and loss of lives.

Following a line of present day thoughts (e.g. Miller, 1997) the reduction of flood damages requires a combined application of three strategies to the problem of flooding.

A first approach ,”keep the flood away from people”, has as its objective to contain or reduce floods through measures such as the construction of dams and reservoir, dikes and levees, demolishing of structures, which increase the inundation risk. In addition to structural measures, soil and water conservation programmes are also part of this strategy.

A second approach”keep people away from floods” focuses on the concept that flood plains are “dangerous places” where stable human activities should be avoided. Part of this strategy includes the definition of risk maps and implementation of non-structural measures such as flood forecasting and early warning systems.

The third approach “accept floods and clean up afterwards” is based on the acceptance of floods and concentrate on action to minimize damages and to promote a fast return to normality after the flood has occurred.

The conceptual inadequacy of any of these approaches considered in an isolated manner as well as the accumulated past experience demonstrated that sustainable and effective solution to flood problems have necessarily to incorporate a balanced view of strategy options and the use of an adequate integration between structural and non structural measures before, during and after the occurrence of floods.

Flood analysis

In the context of an integrated flood management point of view, deep conceptual understanding of physical nature of floods clearly plays a very important key role, being the necessary step for applying the different strategies both structural and non-structural. No successful solution at any stage or flood management system can be achieved without a good knowledge of hydrological processes that are behind the generation and propagation of the flood waves and use of instruments to describe them in a quantitative and as much as possible precise way. In such a condition, analyzing the flood problem and its cause is mandatory and is part of the solution for the problem, helping to apply specific measures and strategies to the problem and to use effective and efficient techniques.

Flood forecasting

Flood forecasting is one of the non-structural measures for flood management. Flood forecasting involves interpreting measured and forecasted rainfall, and river and tidal levels, often-using sophisticated mathematical models. This requires monitoring systems, which consist of a network of rain and river flow gauges and associated systems.

In last several decades, due to the improved understanding of the hydrological process and modern modeling technology, considerable improvements have been made in hydrological modeling, including many applications for flood forecasting. During this period, a great number of hydrological forecasting models were developed, including forecasting for flood problems, such as discharge or reservoir inflow volumes, flow in mainstream, urban and flash flood, storm surges in estuaries and coastal areas, operation of structure, etc.

The flood forecast, with the specified lead-time and probability, would be issued to the appropriate authorities or the public for initiating emergency contingency plans, operation of structures and flood protection reservoirs, or evacuation of affected area. Forecasting methods vary from simple extrapolation of upstream river levels to predict levels at given points downstream, through sophisticated predictive catchments flow forecasting modeling systems.

The key elements of a forecasting system operating in a real time environment are:

- Real time data acquisition for observed meteorological and hydrological conditions
- Hydrologic and hydraulic models for simulation
- Forecast of metrological conditions
- Updating and data assimilation
- Dissemination of forecasts

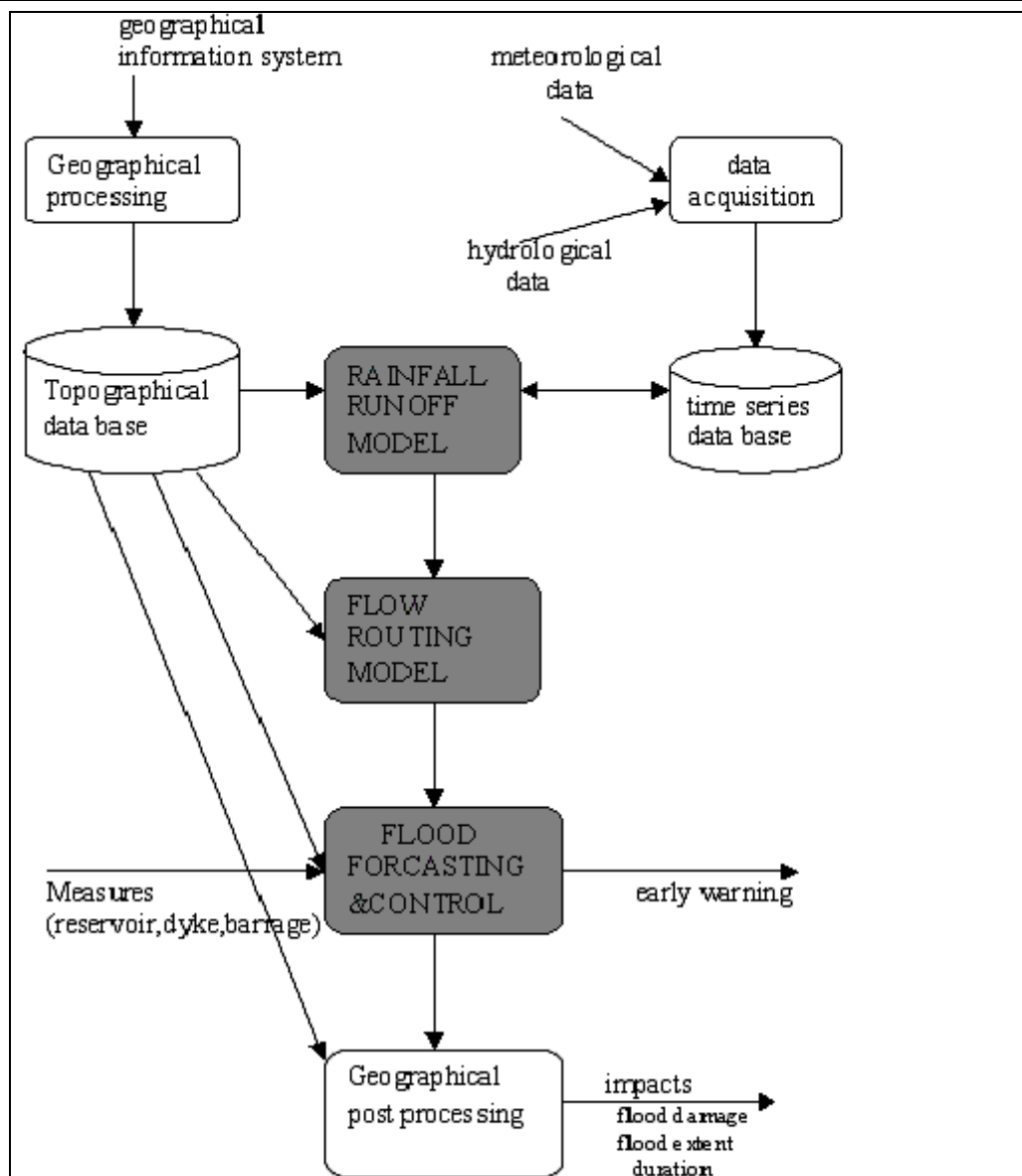


Figure 2.1 Procedure of a Flood forecasting system

Recent developments in weather forecasting, radar data and on-line meteorological and hydrological data collection require an increasing focus on data import and processing within a flood forecasting and warning system.

Flood forecasting is the most common measure in flood control practices; there are many available flood-forecasting methods under the hydrological model class

Hydrological forecasting

The World Meteorological organization (WMO-NO 168, 1994) distinguished five main types of hydrological forecasting depending on range and common features of forecasts:

- **Short-term forecasts:** forecast of hydrological variables up to two days a head

-
- **Medium term forecasts:** forecast of hydrological variables 2 to 10 days ahead
 - **Long-term forecasts:** forecast of hydrological variables beyond 10 days ahead
 - **Seasonal forecasts:** forecasts of hydrological variables several months ahead
 - **Hydrological warning:** emergency information on an expected hydrological event considered being dangerous.

Short to medium term forecasts are most often used for flood warning and for the real time operation of water resource system such as hydropower scheduling. Seasonal forecasting is usually applied to water resource and drought management for the management of reservoirs. The most wide spread use of hydrologic forecasting is converted into a flood warning on the extent, severity, and timing of the flood.

Because most hydrological processes are influenced by many factors and exhibit a high degree of non-linear and spatial temporary variability, hydrological models can also be classified according to the degree of describing the physical processes.

Flood forecasting modeling

In recent years flood forecasting modeling has become an attractive alternative to classical analytical methods. The benefit of flood forecasting can be issued as follows

- Flood forecasting provides an alternative to the more traditional structural methods of flood mitigation. On certain degree, it can reduce or avoid the vast investment on the construction of structures like dikes, embankments
- Forecasting also permits most effective operation of water resource system, water transfers and residents evacuation to avoid flood damages
- The environmental impact of implementing flood forecasting and warning systems is considerably less than many other flood control measures

Flood management by reservoir operation

Flood operation of reservoirs requires that information be made available in a form that can be readily assimilated by operators. In some cases, large reservoir releases have the potential to cause significant downstream damage and loss of life. In these cases, it is important to effectively communicate information about reservoir operation to those responsible for public safety and emergency response. In the past, it has been common practice to base reservoir operating rules on current information about reservoir storage without direct use of inflow forecasts. The justification has been that inflow forecasts are uncertain. However, there is a growing interest in taking account of forecast information, including forecast uncertainties, to improve levels of flood protection through more effective use of available reservoir capacity.

2.3 Flood Modeling

Need for modeling flood

Many of the phenomena that can be identified in water base system are so complex in their generation and prediction that we have to resort to models to help us understand

what is going on, and to make predictions or forecasts of what will happen (Price, 2001).

So according to the above statement some of the reasons we use models are;

- In trying to understand what is going on in complex water based systems
- To make forecasts of future values
- To predict the consequences of structural changes in the modeled system
- To predict the response of a system to extreme unseen events etc.

There are a number of different ways in which we model the real world. They are so distinct from each other that we can refer to different modeling paradigms. The key paradigms that concerns us in hydroinformatics are (Price, 2001)

- Physically based modeling
- Data driven modeling
- System modeling or Cybernetics

Physically-based modeling in flood forecasting

The essence of physically based modeling is to abstract and conceptualize the primary (physical) processes that are perceived to contribute to a given phenomenon, to represent these in a mathematical algorithm form, and to solve the resulting equations with appropriate boundary conditions and functional relationships.

i. Unit hydrograph

Perhaps the best-known and certainly most widely used tool in hydrological forecasting is the unit hydrograph method introduced by Sherman (1932). While even the unit hydrograph method is traditional regarding as black box model, it still has reasonable physical background: the mechanisms of runoff generation. One is the infiltration-excess runoff generation; following the hypothesis of Horton that runoff occurs once the intensity of rainfall exceeds the local infiltration capacity. The other one is saturation-excess runoff generation, whereby runoff results from rainfall on areas that are already saturated by a rising water table. Furthermore, different mechanisms may operate simultaneously in different parts of a catchment, with the contributing area expanding or contracting both during and between storm events.

The unit hydrograph is defined as the hydrograph of direct runoff resulting from a unit depth (usually 10 mm) of effective rainfall generated uniformly over the area at a constant rate during a specified duration. The derivation unit hydrograph includes base flow separation and effective rainfall calculation or losses (interception storage on vegetation, the replenishment of soil moisture deficits and storage in depression on ground surface) subtraction.

Today empirical hydrological method are often used in some components of more comprehensive models, e.g. the unit hydrograph is often used for stream flow routing and linear reservoir is often used to represent the ground water system in conceptual rainfall-runoff models (Refsgaard, 1996)

ii- Conceptual models

The black box becomes more of gray box in that its structure begins to resemble the systems representation of the land phase of hydrological cycle. In a *conceptual* type model the internal descriptions of the various sub processes are modeled attempting to represent, is partitioned into components that are routed through the sub processes either to the catchments outlet as stream flow or to the surface and deep storages or to the atmosphere as evapotranspiration. Conceptual approaches were recognized to be able to improve the description of the hydrological response of a basin in comparison with black-box modeling and this generally implies a better performance in discharge forecasting (e.g. Brathand Rosso, 1993). In fact, black-box models may obtain very good results in modeling events included in the calibration records but they often perform poorly in forecasting under out-of-sample data.

The lumped, conceptual model are especially well suited for the simulation of the rainfall-runoff process when hydrological time series exist that are sufficiently long for a model calibration (Refsgaard, 1996).

iii Distributed physical based model

Contrary to lumped conceptual models, a distributed physical-based model does not consider the water flows in an area to take place between a few storage units. Instead, the flows of water and energy are directly calculated from the governing continuum equation and partial differential equations, such as for instance the Saint Venant equations for overland and channel flow, Richards' equation for unsaturated zone flow and Boussinesq's equation for ground water. Today, several general-purpose catchments model of this type exist. The advantage and typical application of distributed physical based models has also been given by Refsgaard (1996). Such as effects of catchments changes, prediction of runoff from ungauged catchments or catchments with relatively short periods of data.

Data driven modeling

During the last decades, due to progress in vast digital data acquisition and improved processing ability a class of models often referred to as data driven model become quite popular. Physical based models can give us deep insight in to the natural processes and system behavior. The main advantage is that they can predict the events, which may have not occurred before, such as the effect of the building of a hydraulic structure, or extreme flood events. In order to describe the modeled system more accurately, the trends of physical based model are towards simulating a system in more detailed way. Data driven models are adapted to extract information from data and define the relation-ship between system state variables (input, internal, and out put variables) with limited knowledge of physical situations. The historical data are fed in to the model, which is called training. Once the training is performed, the verification is carried out to check the model universality on the other data, unseen by the model. Therefore, the adaptability, reliability and robustness of a data driven model depend on the source, range, quantity, and quality of the data sets.

Data mining, as a multidisciplinary field, is the source of data driven modeling. It relates to many knowledge fields as artificial intelligence, probability theory and statistics, computational complexity theory, information theory, data analysis and database etc. The key point of the data driven modeling is to minimize the difference (error) between the desired output and model output (see Figure 2.2)

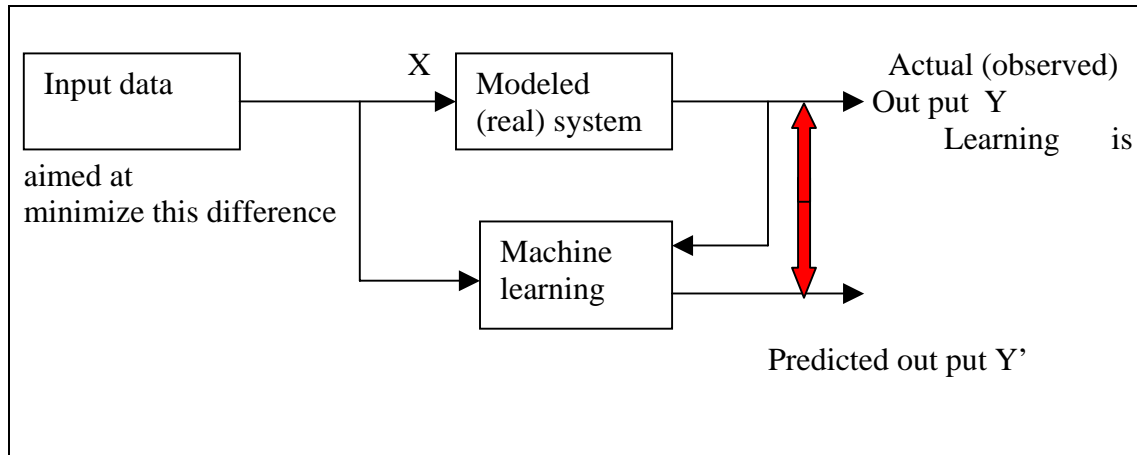


Figure 2.2 Machine learning (Solomatine.2005)

The advantage of data driven modeling can be generalized as follows:

- Easy to build because they don't involve too much information of the complex physical process and system
- Rapid to give the results in comparison to complicated physical based model
- Powerful to catch the multi linear or nonlinear relationship between input and out put variables
- Possible to deal with some missing data and noise in the data set

The limitations of data driven modeling are also generalized as follows:

- Difficult to extrapolate outside the value range of training data
- Difficult to explain the model results from the view of physical phenomena
- High requirement of the quality and quantity of data
- Lack of ability to accommodate to system changes. Normally new models are needed after the changes happen in the physical system

M5 model tree

This machine-learning technique uses the following idea: split the parameter space in to areas (subspaces) and build in each of them a linear regression model. In fact, the resulting model can be seen as a modular model, or a committee machine, with the linear models being specialized on the particular subsets of the input space. The M5 model tree approach, based on the principle of information theory, makes it possible to split the multi-dimensional parameter space and to generate the models automatically according to the overall quality criterion; it also allows for varying the number of models.

The splitting in MT follows the idea of a decision tree, but instead of the class labels, it has linear regression functions at the leaves, which can predict continuous numerical attributes. Model trees generalize the concepts of regression trees, which have constant values at their leaves (Witten & Frank, 2000). So, they are analogous to piece-wise linear functions (and hence nonlinear). Computational requirements for model trees grow rapidly with dimensionality. The major advantage of model trees over regression trees is that model trees are much smaller than regression trees, the decision strength is clear, and regression functions do not normally involve many variables.

The algorithm known as the M5 algorithm is used for inducing a model tree (Quinlan, 1992), which works as follows (Figure 2.3). Suppose that a collection T of training examples is available. Each example is characterized by the values of a fixed set of (input) attributes and has an associated target (output) value. The aim is to construct a model that relates a target value of the training cases to the values of their input attributes. The quality of the model will generally be measured by the accuracy with which it predicts the target values of the unseen cases.

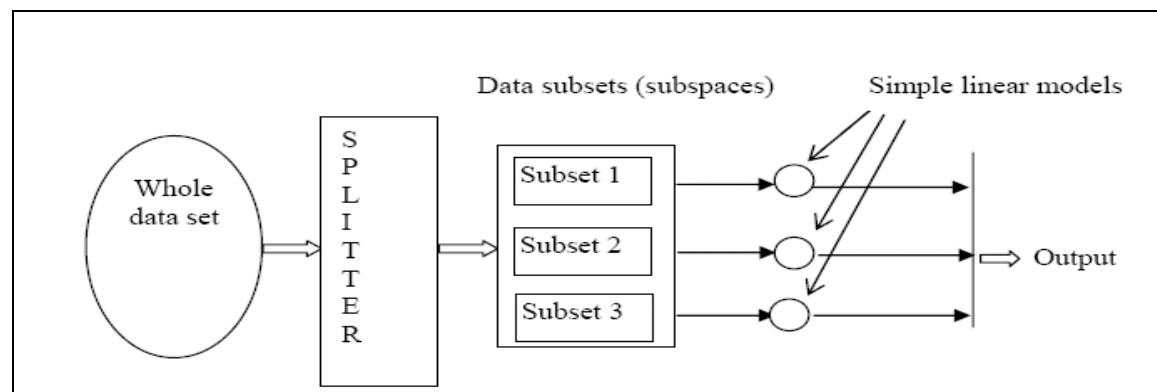


Figure 2.3 Model tree (Solomatine, 2003)

Splitting criterion for model trees

As introduced by Witten and Frank (2000), in Weka software the model trees splitting criterion is SDR (standard deviation reduction). It is used to determine which attribute is the best to split the portion T of the training data that reaches a particular node. It is based on treating the standard deviation of the class values in T as a measure of the error at that node, and on calculating the expected reduction in error as a result of testing each attribute at that node. The attribute that maximizes the expected error reduction is chosen for splitting at the node.

The expected error reduction, standard deviation reduction, is calculated by:

$$SDR = sd(T) - \sum_i \frac{|T_i|}{|T|} \times sd(T_i)$$

where T_1, T_2, \dots are the sets that result from splitting the node according to the chosen attribute.

The splitting process terminates when the class values of the instances that reach a node vary slightly, that is, when their standard deviation is just a small fraction (say, less than 5%) of the standard deviation of the original instance set. Splitting also terminates when just a few instances remain, say, four or fewer. Fig 2.4 shows graph of an M5 created for this research.

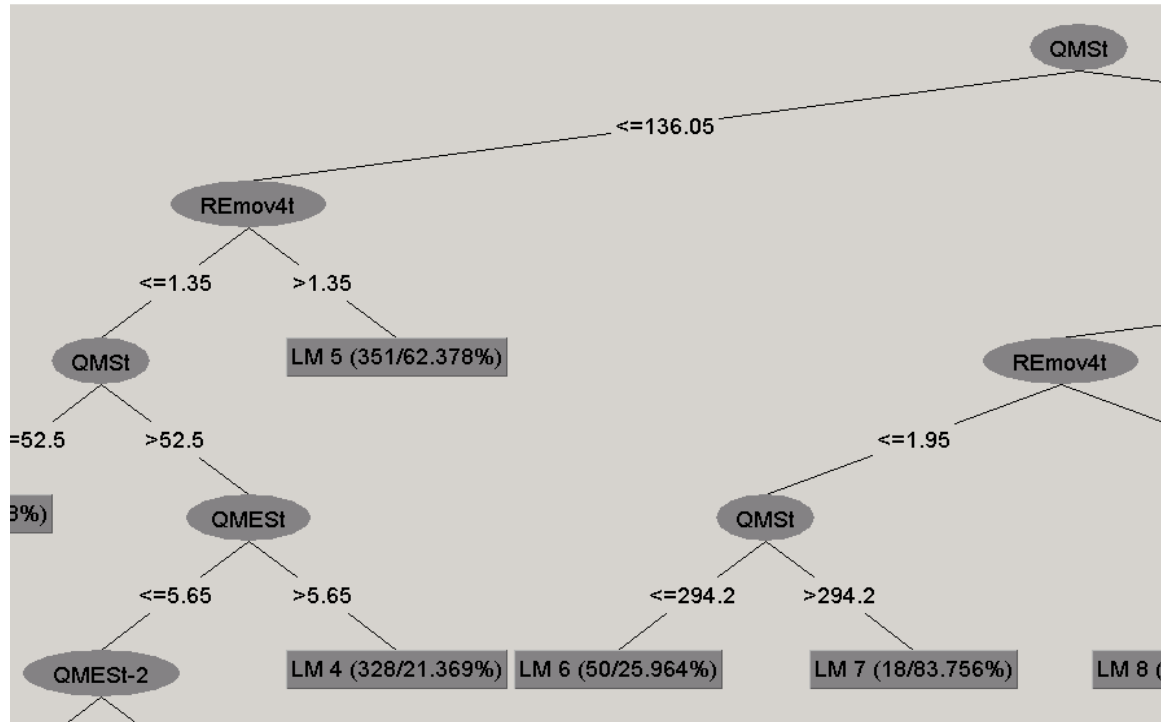


Figure 2.4 M5 model tree sample at MelkaSedi station

Pruning the trees

Over-fitting problems can happen during M5 model trees construction based on training data. Predictably, the accuracy of the trees over the training examples increases monotonically, as the trees are grown. However, the accuracy measured over the independent test examples first increases, then decreases. One quite successful method for finding high accuracy is post-pruning. A big advantage of post-pruning is that sometimes when two attributes individually seem to have nothing to contribute, they are powerful predictors when combined. This can be seen as a sort of combination-lock effect, when the correct combination of the attribute values is very informative, however the attributes taken individually are not (Witten and Frank, 2000).

Artificial neural networks

An artificial neural network (ANN) is an information-processing paradigm inspired by the way the densely interconnected, parallel structure of human brain processes information. Originally studied in the framework of AI, artificial neural network (ANN) has become now one of the primary technologies in machine learning, and a mainstream technology for data-driven modeling. Various types of networks are used in clustering, classification and prediction. ANNs loosely imitate functioning of

neurons in a human's brain, and it appeared it is possible to combine these neurons in such a way, that the network would reproduce any multi-variable multi-valued function, given enough points and values of this function. By analogy with the brain, the operation of the trained (learned) network is often called recall (Solomatin, 2000).

ANN has been developed as a generalization of mathematical models, human cognition or neural biology, based on the assumption that:

- Information processing occurs at many simple elements called neurons/units/cells/nodes
- Signals are passed between neurons through connection links
- Each neuron applies an activation function (usually non-linear) to its net input (sum of weighted signals) to determine its out put signals

A neural network is characterized by its particular:

- Architecture; which defines the pattern of connections between neurons
- Learning algorithms; which identify the method of determining the weights on the connections.
- Activation function; which determines its output

An ANN consists of a large number of processing elements, called neurons. Each neuron has an internal state called its activation or activity level, which is a function of the inputs it has received. Typically, a neuron sends its activation as a signal to several other neurons. A neuron can send only one signal at a time, although that signal may be broadcast to several other neurons.

ANNs exhibit three features, namely, distributed processing, adaptation and non-linearity, and it has been mathematically proven that adding up simple functions, as a ANN does, allows for universal approximation of functions (Kolmogorov 1957). This means that neural networks can approximate any function that best characterizes a time series. It is this property that has stimulated civil engineers to adapt, investigate, and improve the performance of neural networks associated with their applications.

ANN has the following limitations:

- The success of an ANN depends on both the quality and quantity of data
- Lack of physical concept and relations
- No standard way of selecting optimum network architecture and network parameters

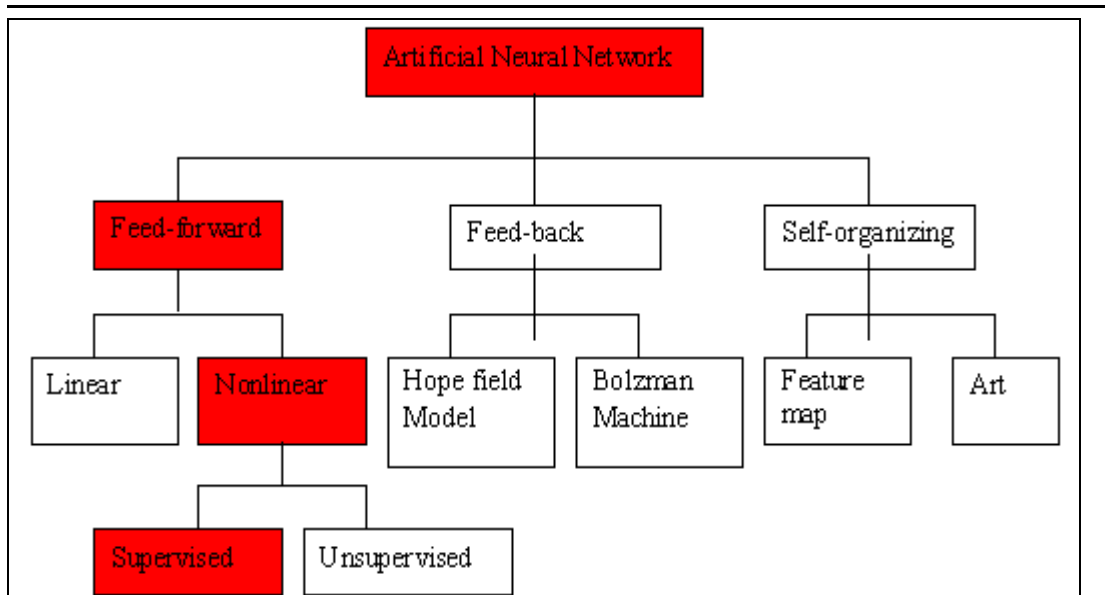


Figure 2.5 ANN's classification (Solomatine, 2000)

Multi-layer perceptron network

Multi-layer perceptron (MLP) is one of the most popular and successful neural network architecture (Figure 2.6) a MLP network can have several layers: input layers, output layers and layers in between input layer and output layers known as hidden layers, which do not have direct connection to the outer world. It is a feed forward network in which information passes from the input to the out put side. The nodes in one layer are connected to those in the next, but not to those in the same layer.

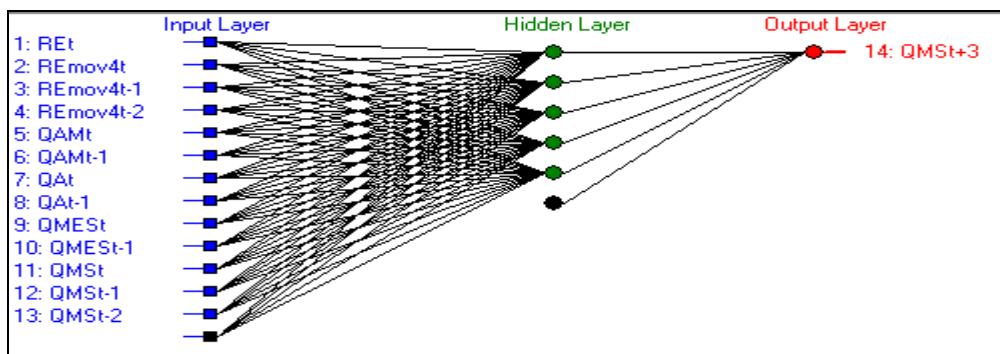


Figure 2.6 Multi-layer perceptron network sample

Choice of Training set

The size of training set is of fundamental importance to the practical usefulness of the network. If the training patterns do not convey all characteristics of the problem class, the mapping discovered during training only applies to the training set. Thus, the performance in the test set will be much worse than the training set performance. The

only general rules that can be formulated are to use a lot data and use representative data. If we do not have many data to train the ANN, then the ANN paradigms probably not the best solution to solve our problem. Understanding of the problem to be solved is of fundamental importance in this step.

Another aspect of proper training is related to the relation between training set size and number of weights in the ANN. If the number of training examples is smaller than the number of weights, one can expect that it may allocate one weight to each training example. This will obviously produce poor generalization (i.e. the ability to link unseen examples to the classes defined from training examples).

Network size

Establishing the size of a network is more efficiently done through experimentation (trial and error). The number of processing element in the hidden layer is associated with the mapping ability of the network. The larger the number, the more power full the network. However, if the network size is increased, there is a point where the generalization gets worse. This is due to the fact that we may be over-fitting the training set, so when the network is confronted with patterns that it has never seen before the response is unpredictable. The problem is to find the smallest number of degrees of freedom that achieves the required performance in the test set.

One idea is to start with small network and increase their size until the performance in the test is appropriate

Learning Parameters

The control of the learning parameters is an unsolved problem in ANN research. The point is that one wants to train as fast as possible and reach the best performance. Increasing the learning rate parameters will decrease the training time, but will also increase the possibility of divergence, and of rattling around the optimal value. Since the weight correction is dependent up-on the performance surface characteristics learning rate an adaptive learning parameter is necessary to obtain constant learning.

Stopping criteria

To prevent overtraining and to get good generalization, suitable stopping criteria should be applied during training. The most used criteria is to choose the number of iteration, or present the final error i.e. stop when mean square error(MSE) fall below certain predetermined level. Besides these, there is another criteria to prevent overstraining in which the data set is divided in to training, cross-validation and testing set and the training is stopped when the MSE on the cross-validation set starts to increase.

2.4 Parameters for data analysis and model performance analysis

Parameters for data analysis

The fundamental task in many data analysis is to identify the features of the data. here some basic concepts about the parameters for the data analysis and model performance analysis are presented

Mean

The mean is the sum of the data points divided by the number of data points. It can be expressed using the following formula:

$$\bar{Y} = \sum y_i / n$$

The mean is the term to measures of location and one of the estimators to check whether the underlying distribution of data set is normal. However, it lacks robustness of validity.

Median

The median is the value of the points, which has half data smaller than it and half data larger than it. If $y_1, y_2, y_3 \dots y_n$ is a random data set sorted from smallest value to largest value, then the median is defined as:

$$\tilde{Y} = y_{(n+1)/2} \text{ if } n \text{ is odd}$$

$$\tilde{Y} = (y_{n/2} + y_{n/2+1}) / 2 \text{ if } n \text{ is even}$$

The median an example of an estimator to check the robustness of validity but not the robustness of efficiency.

Mode

The mode is the value of the random sample that occurs with the greatest frequency. It is not necessary unique. The mode is typically used in a qualitative fashion. for example. There may be a single dominant hump in the data perhaps two or more smaller humps in the data. This is usually evident from histogram of the data. However for the samples from continuous populations, any specific value may only occur once. In this case, it will be more meaningful to use midpoint of the class interval of the histogram with the highest peak

Standard deviation

The standard deviation is the square root of the variance. That is,

$$Sd = \sqrt{\sum_{i=1}^n (y_i - \tilde{Y})^2 / (n - 1)}$$

The standard deviation is roughly the arithmetic average of the distance from the mean. It resorts the unit of the spread to the original data units. The standard deviation is an estimate that is the best we can do if the underlying distribution is normal. However, it lacks robustness of validity.

Coefficient of variation

The coefficient of variation is denoted as COV (X) or δx . It measures the degree of dispersion in the random variable. A smaller value of the COV indicates a smaller amount of uncertainty or randomness in the variable.

Range

The range is the largest value minus the smallest value in the data set. This measure is based only on the lowest and highest extreme values in the sample. And the spread near the center of the data is not captured at all.

Skewness

The skewness, known as the third central moment, is the measure of symmetry, or more precisely, the lack of symmetry. A distribution of one data set is symmetric if the left part of the center point is same as the right. For univariate data y_1, y_2, \dots, y_n , the formula for skewness is:

$$Skewness = \frac{\sum_{i=1}^n (y_i - \bar{y})^3}{(n-1)sd^3}$$

For a normal distribution, the skewness is equal to zero. And any symmetric data set should have a skewness near zero. Negative values for the skewness indicates that data set is skewed left and positive values indicates that data set is skewed right

Kurtosis

The kurtosis is the measure of whether the data are peaked or flat to a normal distribution. For univariate data y_1, y_2, \dots, y_n , the formula is:

$$Kurtosis = \frac{\sum_{i=1}^n (y_i - \bar{y})^4}{(n-1)sd^4}$$

Because the kurtosis is for a standard normal distribution is three, the excess kurtosis can be defined as:

$$Kurtosis = \frac{\sum_{i=1}^n (y_i - \bar{y})^4}{(n-1)sd^4} - 3$$

In this way, the standard normal distribution has a kurtosis of zero. Namely, data sets with high kurtosis tend to have a distinct peak near the mean, decline rather rapidly, and have tails. Data sets with low kurtosis tend to have a flat top near the mean rather than a sharp peak

Performance criteria for modeling

There are a large number of performance criteria used by the difference researchers to quantitatively measure the accuracy, efficiency and reliability of their models. It is difficult to select one criterion as a benchmark standard, and some criteria are only applied to certain specific problems. General rules are several criteria should be selected at the same time for model performance evaluation.

A variety of such criteria were proposed by the World Meteorology Organization (WMO, 1975) and other researchers. Generally, they are grouped in to graphical and numerical performance indicators (Wang, 2001).

Graphical criteria

The graphical performance indicators often provide a direct view of model over performance.

- The time series plot of the calculated and observed hydrograph for both the calibration and verification process. The zoomed in plot should further drawn offer the detail information.
- The scatter plot of the calculated versus the observed values for the verification period
- Box plots of the calculated and observed values. It is the indicator to check whether the statistical features of the original series is kept in the generated series

Evaluating numerical prediction

As shown in table below, several measures can be used to evaluate the performance of the models on simulation and prediction. Among which correlation coefficient, root mean squared error and mean absolute error are most common used. Specially, the relative error estimator (root relative square error and relative absolute error) is used in weka software as the indicator for the basic predictability or unpredictability of the out put variables. If the values of output variable are fairly close to their average value, then the expected prediction is to be good and the relative value will compensate for this. Otherwise, if the error value in one situation is far greater than in another situation, it may be because the quantity in the first situation is inherently more variable and therefore hard to predict, but not because the predictor is worse.

Table 2.1 Performance measures for numerical prediction (p are predicted values and a are actual values)

Correlation coefficient R	$\frac{\sum_{i=1}^n (p_i - \bar{p})(a_i - \bar{a})}{\sqrt{\sum_{i=1}^n (p_i - \bar{p})^2} \sqrt{\sum_{i=1}^n (a_i - \bar{a})^2}}$
Root mean-squared error RMSE (penalizes errors at high values)	$\sqrt{\frac{(p_1 - a_1)^2 + \dots + (p_n - a_n)^2}{n}}$
Mean absolute error MAE (shows the deviation from observed values regardless of sign; does not penalize high errors)	$\frac{ p_1 - a_1 + \dots + p_n - a_n }{n}$
Maximum absolute error	$\max(p_i - a_i)$
Root relative squared error (measure of fit especially at lower values)	$\sqrt{\frac{(p_1 - a_1)^2 + \dots + (p_n - a_n)^2}{(a_1 - \bar{a})^2 + \dots + (a_n - \bar{a})^2}}$
Relative absolute error	$\frac{ p_1 - a_1 + \dots + p_n - a_n }{ p_1 - \bar{a} + \dots + p_n - \bar{a} }$

Source (Solomatine, 2005)

3. Data inventory and analysis

After posing study objective, one important step in modeling is to prepare the data for the data analysis and meeting the requirements of modeling. Data preparation process involves collecting data, checking the data quality, correcting the errors; we can have a better understanding about the nature of data and getting insight the problem the data describes. In this chapter, the data collection and processing activities conducted for this study are presented

3.1 Data inventory

Rainfall and discharge

In Awash River basin, there are 512 rainfall stations both manual and automatic and 113 discharge stations are found. Out of this 20 discharge stations and 24 rainfall stations (Figure 3.1) are managed and coordinated by Awash basin authority who has a responsibility in upper and middle part of the basin including the Methara and Tendaho farms and Koka dam. The Awash basin authority consists of a committee from Ministry of Water Resource and Development, Ethiopian power and light authority and Ministry of Agriculture. The committee manages the release of water from Koka reservoir, power generation and irrigation development works. The three discharge stations on the main river upstream of Koka reservoir are MelkaHombole, Melka Kunture and Bello, and one station on the tributaries Akaki are taken for studying the area upstream of the reservoir. In addition, for downstream flood analysis and flood forecasting six discharge stations along the Main River and 10 discharges, stations along the tributaries are taken (Figure 3.2).

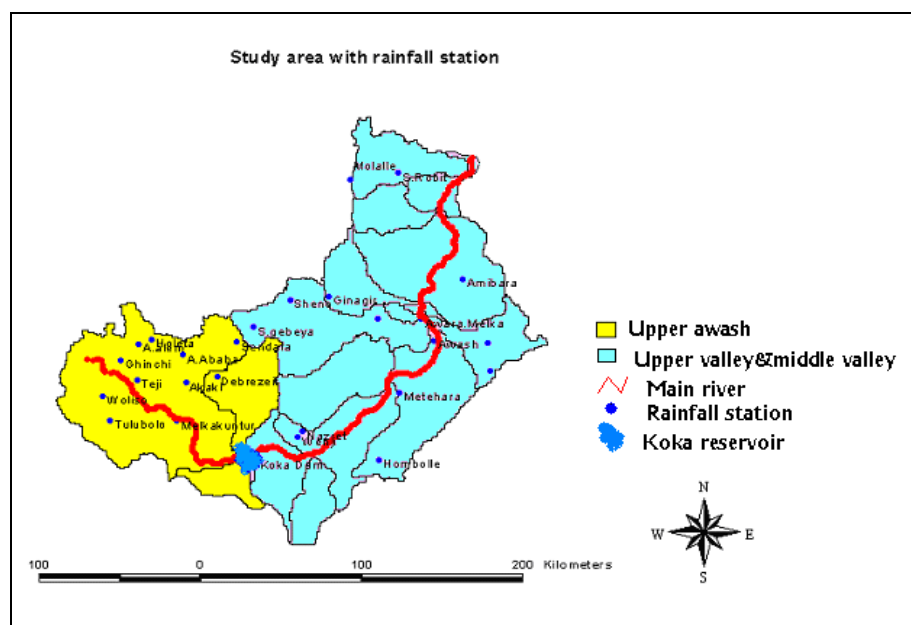


Figure 3.1 the study area with rainfall stations

Two types of data available upstream of the reservoir:

- 1) The daily discharge time series of 10 year (1990-2000) at three station on the main river Awash and on one station on tributary of Awash
- 2) Daily rainfall data time series of 10 year of 7stations. The weight of each station is calculated using thiessen polygon.

And downstream of the reservoir:

- 1) The daily discharge of more than 10 years discharge data in more than 10 stations
- 2) Daily rainfall data time series of 5 year of 8 stations

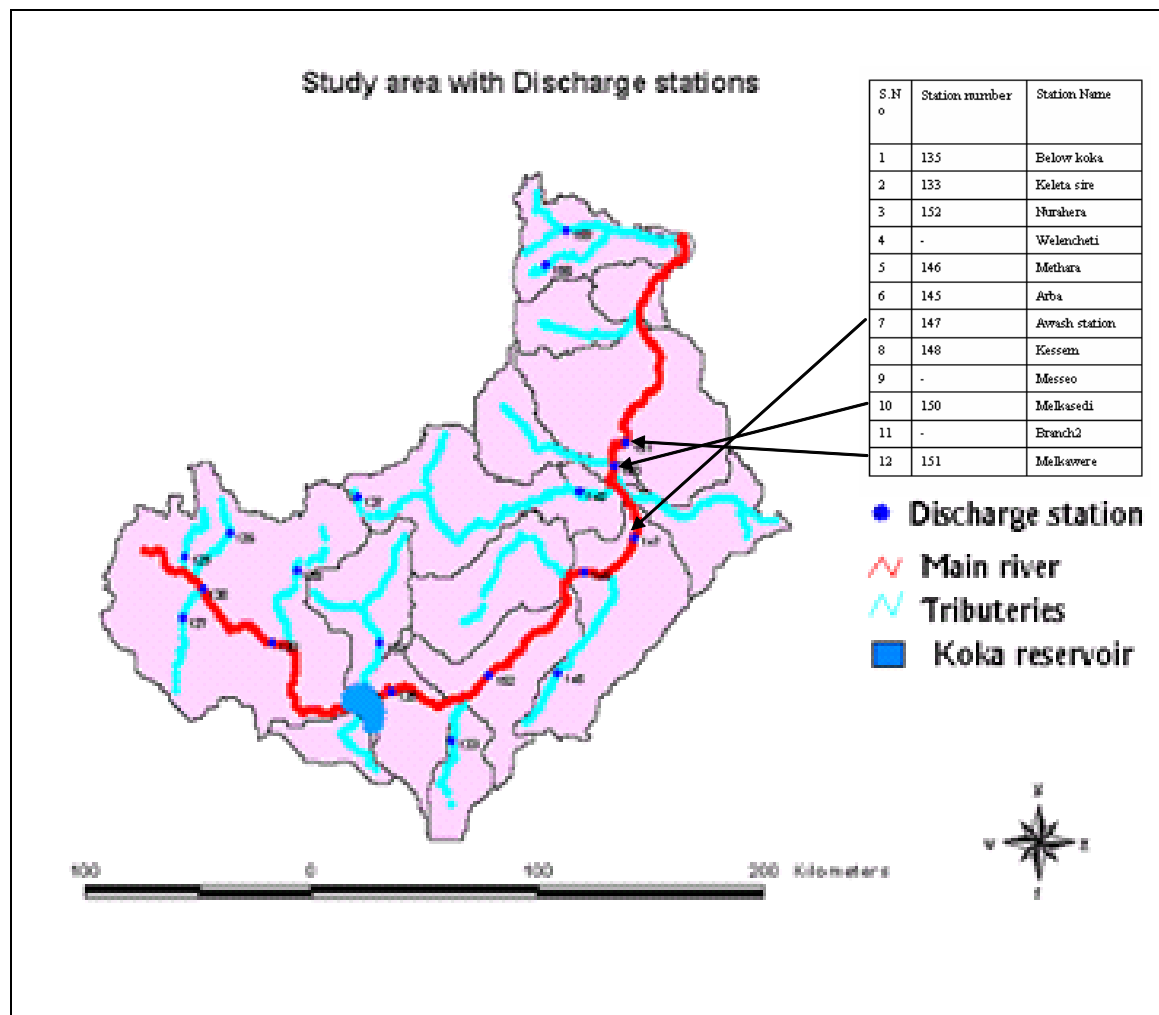


Figure 3.2 Study area with its discharge stations

Temperature and Evaporation

Mean Maximum and Minimum Temperature data of 8 stations including Koka station at the reservoir are available. Using thornthwaite formula the daily evaporation of the station is calculated and the area average is taken following same procedure as that of rainfall using thiessen polygon method. And for open water, evaporation for reservoir loss Penman open water Evaporation is used. Figure 3.3 shows temperature data for

Addis Ababa. Evaporation and evapotranspiration data is prepared to use for water balance of koka reservoir and rainfall-runoff modeling

Thornthwaite Method

Thornthwaite formula is commonly used and hence adopted in this work. This method uses air temperature as an index of the energy available for evapotranspiration and provides a figure for the consumptive use of short closed vegetation with adequate water supply. In this method the monthly gross potential evapotranspiration is calculated using the following formula:

$$ET_p = 0.16 \left(10 \left(\frac{T_n}{J} \right) \right)^a$$

where ET_p - Monthly gross potential evapotranspiration in centimeter per month.

T_n - Mean monthly temperature of the consecutive month of the year in °c

n - 1, 2, 3... 12 the number of the month

J - Annual heat index it is the summation of monthly heat index, j that is given by:

$$j = \left(\frac{T_n}{5} \right)^{1.514}$$

$$a = (0.49 + 0.0179 * J - 0.0000771 * J^2 + 0.000000675 * J^3)$$

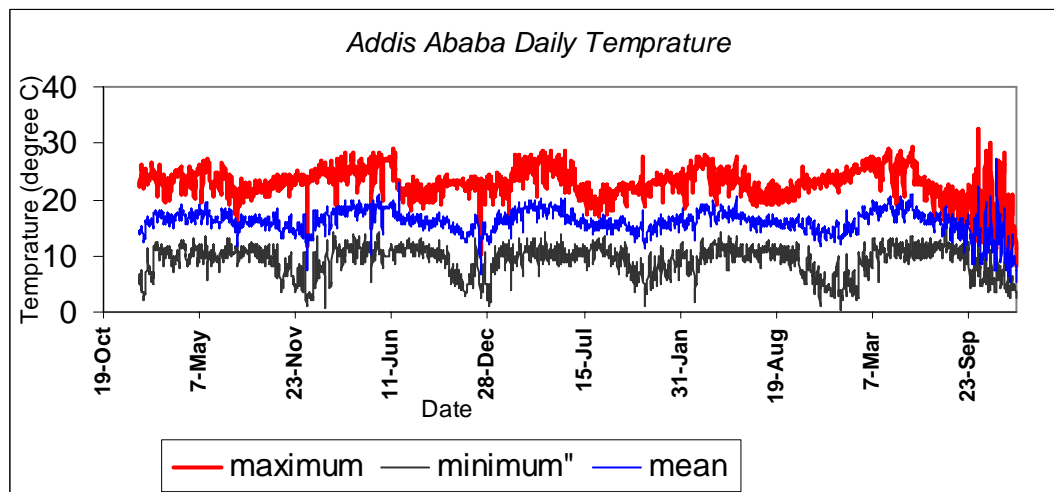


Figure 3.3 maximum, mean and minimum temperature

Soil Data

Soil data as a major medium in the rainfall runoff process, needs to be carefully prepared and its influence in the rainfall runoff process properly established. The infiltration, ground water recharge, and surface runoff are determined in the GR3J

model, which calculates inflow to the reservoir from the upstream catchment, based on the soil data input.

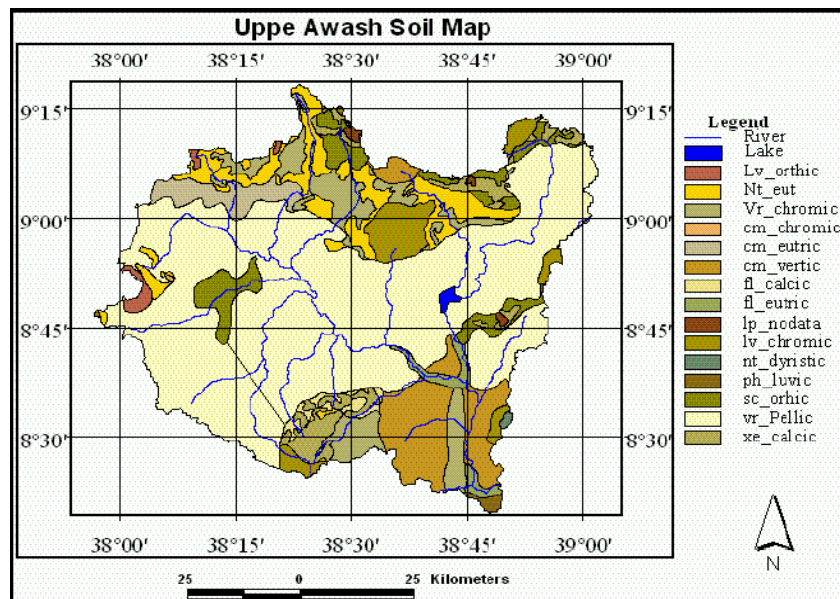


Figure 3.4 soil map of upper basin

The drawback of the map was that it has only the soil unit name (Figure 3.4). A research report of soil research stations in the basin and the FAO soil database from Global Soil Data Products are used to relate the soil unit and the corresponding soil property. Conversion formulas written in the manual of the FEWS-SFM African database were implemented to obtain the representative values of the various soil properties as per the model requirement

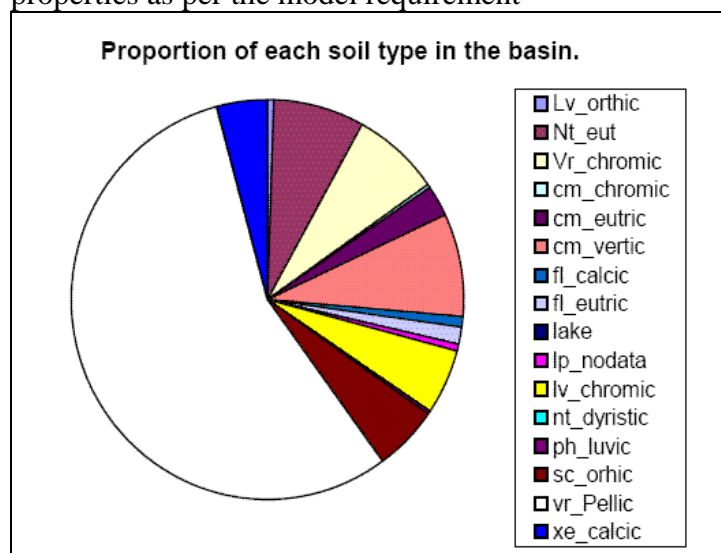


Figure 3.5 proportion of soil in the upper basin (data taken from FEWS-NET)

Table 3.1 soil unit and its properties in the upper Awash

Soil Unit	symbol	HSG	Rclfd	Texture	Ks	Depth	WHC
Orthic Acrisols	Ao	B	2	1	0.000772		0.1808
Cromic Cambisols	Bc	B	2	2	0.00012	86	0.1232
Eutric Cambisols	Be	B	2	2	0.000632	55	0.1727
Orthic Cambisol*	Bo	B	2	3	0.0001		0.107
Vertic Cambisols	Bv	C	3	2	0.00015		0.1211
Dyristic Gleysol	Gd	C	3	2	0.00033		0.139
Haplic Phaeozems	Hh	C	3	2	0.000537		0.1646
Luvic Phaeozems	Hl	C	3	2	0.00062		0.1685
Lithosols*	I	D	4	2	0.001		0.06
Cacarcic Fluvisols*	Jc	B	2	2	0.001	20	0.1
Eutic Fluvisols	Je	B	2	2	0.00045	150	0.1621
Calcic Fluvisol*	Jk	B	2	3	0.0001		0.95
Chromic Luvisols	Lc	B	2	2	0.000575		0.1621
Orthic Luvisols	Lo	B	2	1	0.000752		0.179
No data	N_d			2	0.001		0.1
Dystric Nitosols*	Nd	B	2	2	0.0008		0.1
Eutric Nitosols	Ne	B	2	3	0.00005		0.1028
Cambic Arenosols*	Qb	A	1	3	0.0001		0.107
Eutric Rgosols	Re	A	1	1	0.000818	78	0.1883
Mollic Andosols	Tm	B	2	3	0.000138		0.1131
Vitric Andosols	Tv	B	2	1	0.000369		0.1001
Chromic Vertsols	Vc	D	4	2	0.00038	125	0.1302
Pellic vertisols	Vp	D	4	3	0.000107	123	0.0375
water body	W			5	0.000001		0.01
Haplic Xerosols	Xh	B	2	1	0.000748		0.1816
Caloic Zerosols	Xk	B	2	2	0.000462		0.094
Gypsic Yermosols	Yy	B	2	2	0.000659		0.0745
Orthic Solonchaks	Zo	B	2	1	0.000727		0.1798

* Data taken from FEWS-NET Africa Data set.

3.2 Data preparation and analysis

Data correction and completion

All the data source are from a special organization, Awash River basin authority, and have been verified and validated by specialists but some of the data are still need validation. For instance peak flow of 596m³/s has occurred in March15, 1992 but the respective rainfall value was 4.1mm in same day and 8mm for the previous day at Awash station. Similarly, comparing the daily flow records at MelkaHombole and Melka Kunture showed considerably higher flow records upstream than MelkaHombole for about two weeks, where it takes not more than a couple of days for a peak flow to reach MelkaHombole, as shown (Table3.3)

Among the upstream discharge stations, two were stations of particular relevance to the simulation of the Upper Awash basin. The first station in the middle of the basin is Melka kunture, which has a relatively consistent data, and the second MelkaHombole at the 40 km above the reservoir were taken for data validation (Figure3.5). Therefore, the MelkaHombole record was checked against the Melk Kunture station by using double mass curve analysis. The curve fitted is shown below and shows that the data are reliable.

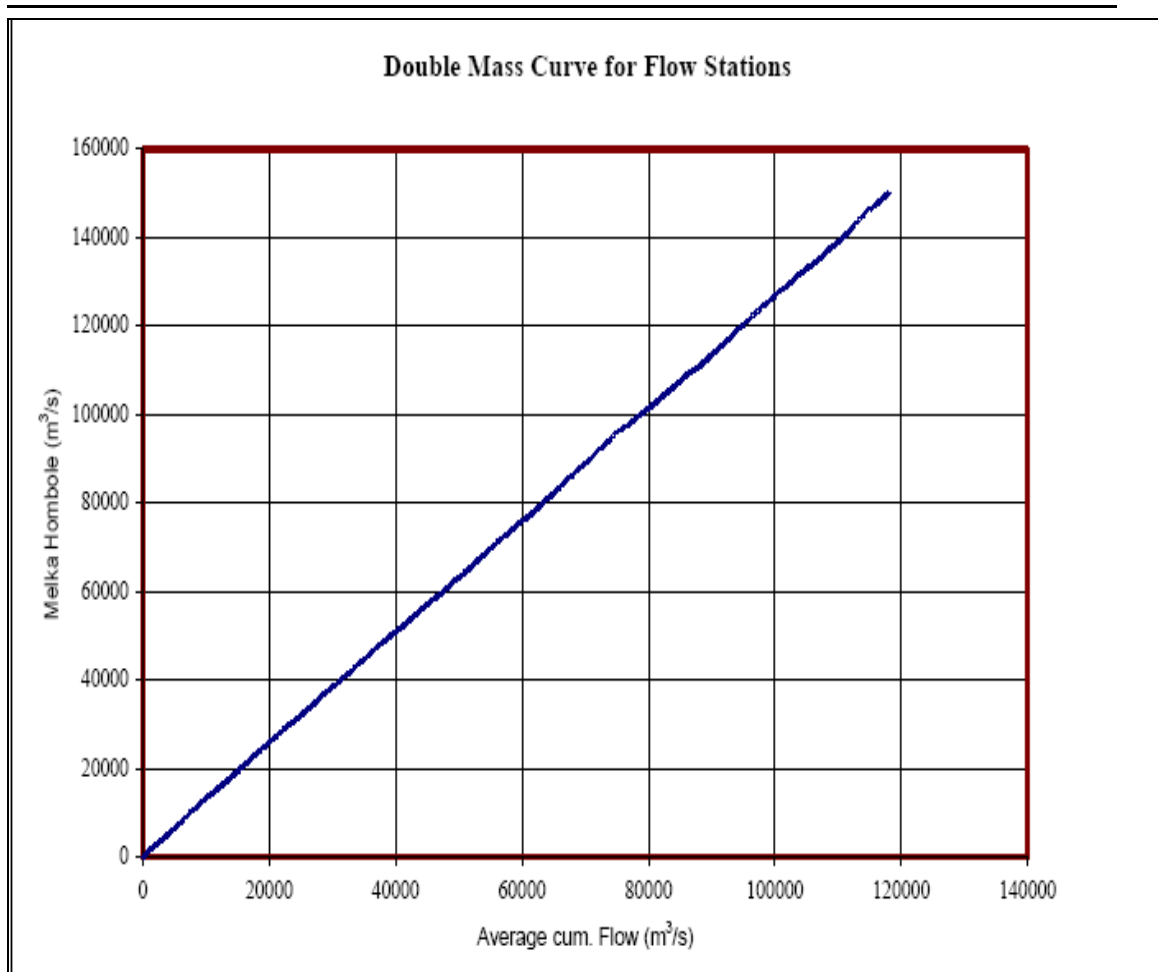


Figure 3.6 Double mass curve of station MelkaHombole and MelkaKunture in upper basin

Data analysis in the basin downstream of koka reservoir

From the above stations (Figure 3.2) Nurahera, Methara, Awashstation, MelkaSedi and MelkaWere are located in the main river Awash below the reservoir consecutively. It should be that the discharge at these stations is the summation of discharges upstream of it on the main river and its tributaries and rainfall-runoff in between, especially in wet seasons provided that the geological conditions are good or no significant loss of water due to cracks and faults occurs. In order to check this, system analysis is undertaken in the basin by calculating the discharge between stations upstream and downstream and compare the results and this is showed in the following graph below. From the result in MelkaSedi and MelkaWere stations, the downstream discharge is lower than the summation of upstream discharges and for the other station Nurahera, Methara and Awash station the downstream discharge is higher than the summation of upstream stations this may be due to rainfall in between stations. Because the difference for MelkaSedi limited, this station will be used for the flood analysis. Because much flow seems to be loss between MelkaSedi and MelkWerer. The result of calculations is depicted below.

The result is calculated based on:

Station 152(nurahera) =station135 (below koka) +station 133(keleta) +q1 (discharge due to rainfall in between the stations)

Station 146(Methara) =station 152(nurahera) + river welencheti+q2

Station 147(Awash station) = station 146(Methara) + station 145(arba) +q3

Station 150(Melka Sedi) =station 147(Awash station) +station 148(Kessem) +Messeo+q4

Station 151(Melka Werer) =station 150(Melka Sedi) +Branch 2+q5

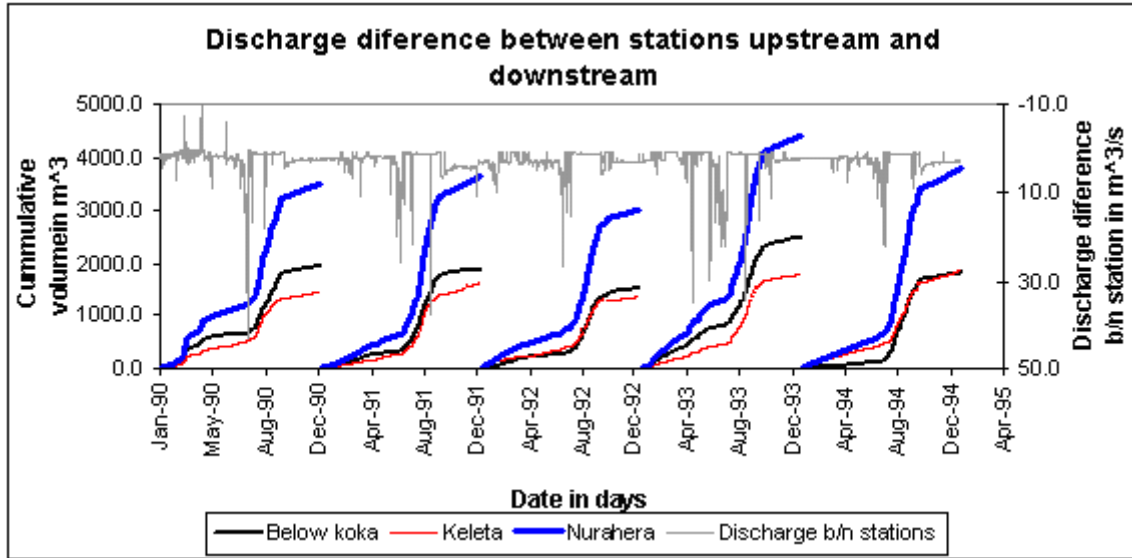


Figure 3.7 Flow between station Nurahera and upstream flows (Q1)

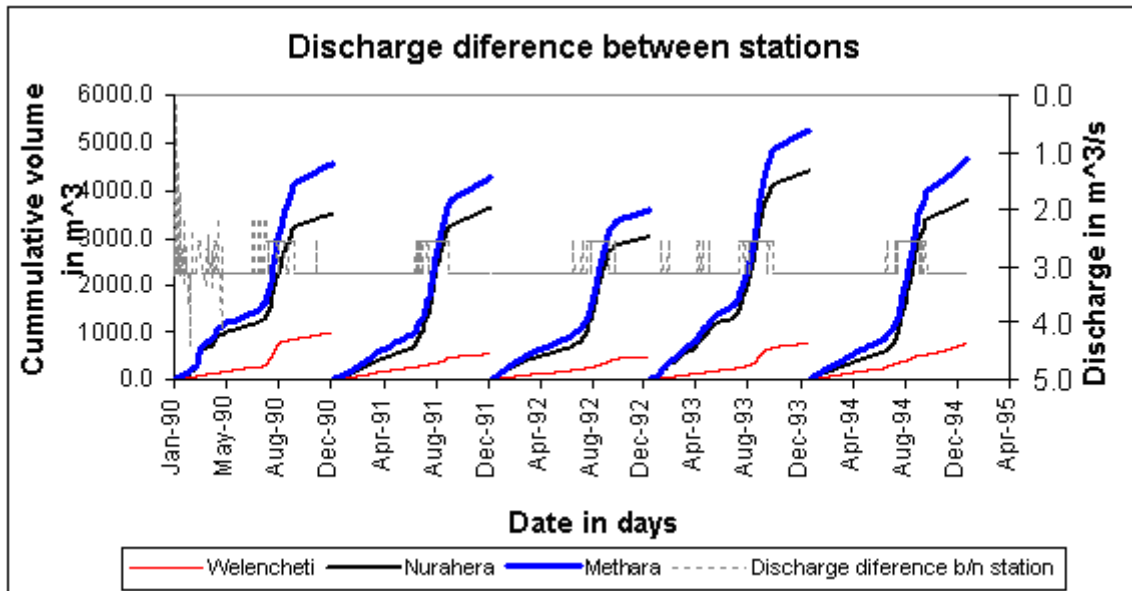


Figure 3.8 Flow between station Methara and upstream flows (Q2)

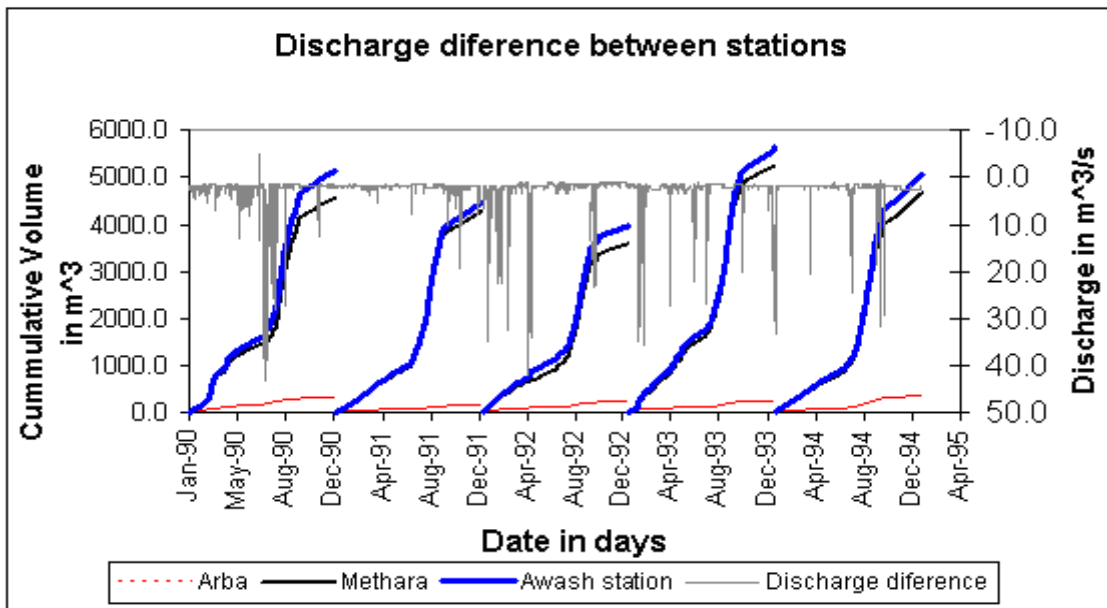


Figure 3.9 Flow between station Awash station and upstream flows (Q3)

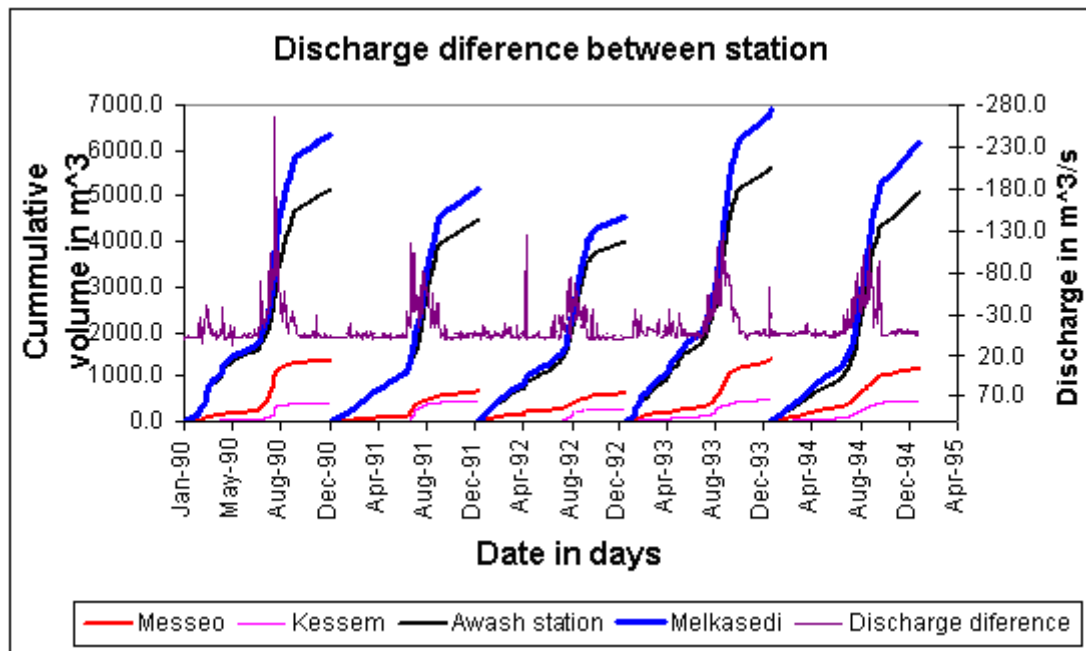


Figure 3.10 Flow between station MelkaSedi and upstream flows (Q4)

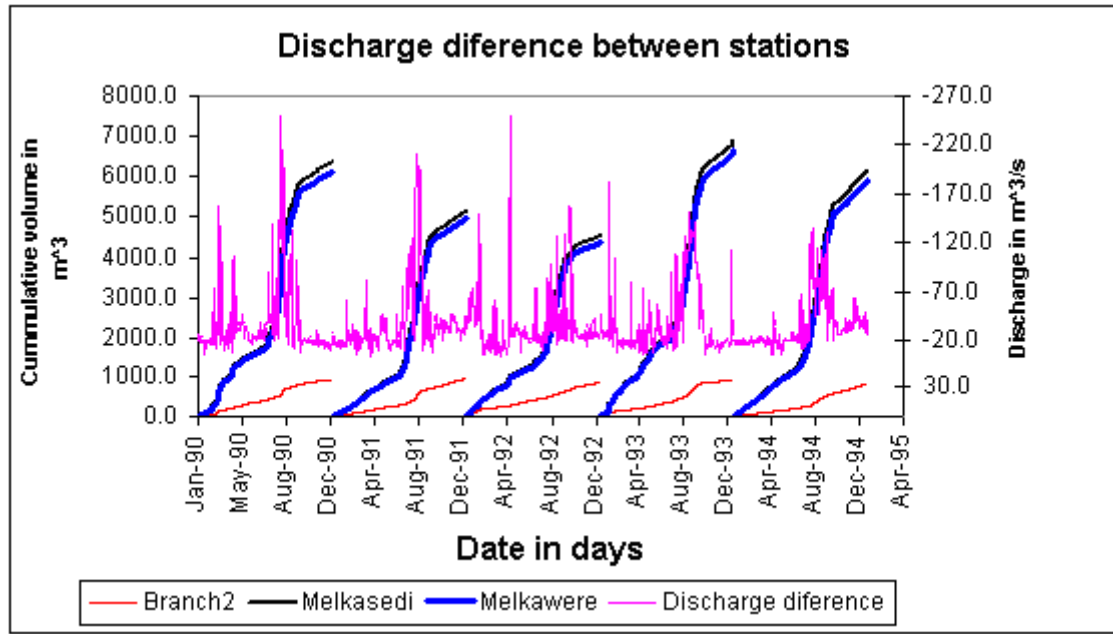


Figure 3.11 Flow between station MelkaWere and upstream flows (Q5)

Table 3.2 Comparison of flow record at Melka Hombole and Melka Kunture

Date	M.Hombole	M.Kunture
18-Aug-97	119	163
19-Aug-97	116	148
20-Aug-97	110	121
21-Aug-97	103	116
22-Aug-97	218	102
23-Aug-97	166	98
24-Aug-97	117	100
25-Aug-97	88	78
26-Aug-97	69	66
27-Aug-97	62	146
28-Aug-97	39	86
29-Aug-97	39	73
30-Aug-97	31	60

Based on the data overview a double mass curve is prepared to check the consistency of the data between stations. In a similar condition in Awash stations downstream of Koka reservoir there was some missing data so these data are filled by using the neighboring stations discharge using the following formula and multiple linear regressions.

$$Y = C + C_1X_1 + C_2X_2 + C_3X_3 + etc$$

Where Y a series of values of the base station (dependent variables)

X_i a series of values of neighboring stations i (independent variables)

C the equation's constant

C_i the equation's coefficient

$$Q_{est} = \frac{\sum \left(\frac{Q_i}{r^b} \right)}{\sum \left(\frac{1}{r^b} \right)}$$

Where Q_{est} estimated discharge at the missing station
 Q_i measured discharge of stations near by
 r distance to stations
 b power of distance (usually 2)

N.B. The formula is developed for rainfall estimation from the neighboring stations here I am using for discharge estimation. Adopted from workshop on hydrology lecture note (P.J.M.de Laat, 2003)

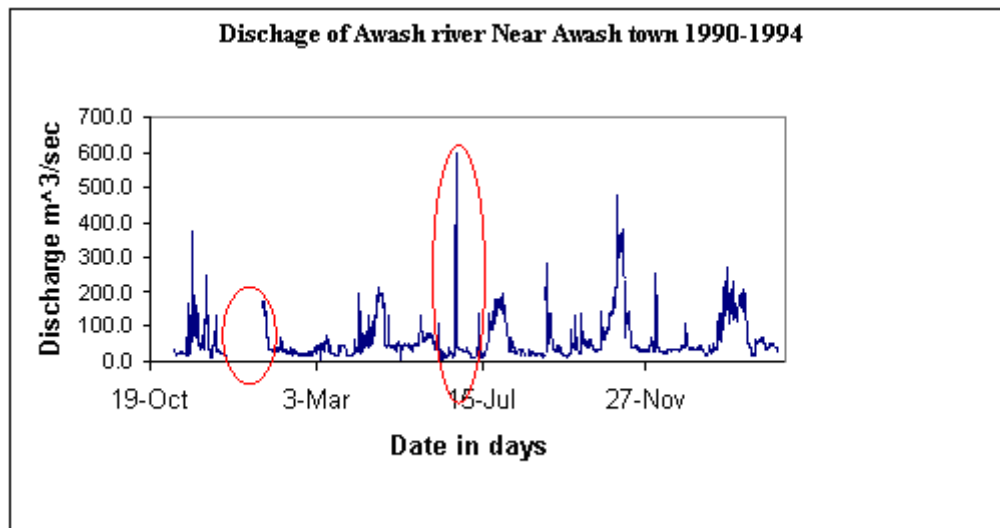


Figure 3.10 Example of missing values, 1991

3.3 Basic characteristics of the Awash basin

Rainfall and evapotranspiration

The seasonal rainfall distribution within the basin results from the annual migration of the ITCZ. In March, the ITCZ advances across the Basin from the south, bringing the small or spring rains. In June and July it reaches its most northerly location beyond the Basin that then experiences the heavy or summer rains. It then returns southwards during August to October, restoring the drier easterly air streams, which prevails until the cycle repeats itself in March.

The annual rainfall distribution resulting from this cycle is exhibited most clearly in the two distinct rainy periods, which are characteristic of the northern plains of the basin. Moving southwards the more prolonged exposure to the moist air-stream is

evident in the tendency for the two dominant rainy periods to merge into a contiguous distribution.

The mean annual rainfall over the entire Western Catchment is 850 mm and over the headwaters of the Awash, as gauged at Melka Hombole it is 1216 mm. Over the Eastern Catchment the mean annual rainfall is estimated to be 456 mm the annual and monthly rainfalls are characterized by high variability.

Potential evapotranspiration (PET) is also significantly correlated with altitude. At Wonji, in the Upper Valley, the mean annual PET is 1810 mm, over twice the mean annual rainfall, with average monthly rainfall exceeding average monthly Evapotranspiration only in July and August. At Dubti, in the Lower valley the mean annual PET is 2348 mm, which is over ten times the mean annual rainfall. Mean annual temperatures range from 20.8°C at Koka to 29°C at Dubti, with the highest mean monthly temperatures at these stations occurring in June, at 23.8°C and 33.6°C respectively

Runoff

As shown in table 3.3, the discharge of Awash River and its tributary are relatively small at normal time and even nearly dry out in dry season, but the flood peak is relatively large, which is nearly 8 to 10 times of its mean discharge.

Table 3.3 The characteristic of discharges in m³/s and area average rainfall in mm upstream of koka reservoir and downstream

Upstream of the reservoir

	Bello (QB)	Akaki (QA)	Melkakature (QM)	Melka Hombole (QMB)	RE (Rainfall)
Average	8.71	16.59	26.12	45.31	3.01
Standard dev	15.63	29.34	47.44	85.57	4.17
Maximum	541.25	363.60	398.63	803.14	26.58
Minimum	0.30	0.00	0.00	0.00	0.00
Time series	1991-2000	1991-2000	1991-2000	1991-2000	1991-2000

Downstream of the reservoir

	Nurehera	Methara	Awash station	AwraMelka	MelkaWere
Average	116.4	141.6	153.9	12.9	177.3
Standard dev	152.3	168.5	175.6	41	212.6
Minimum	14.6	18.4	19.5	0.0	23.9
Maximum	946.5	1021.1	1078.1	712.2	2343.9
Time series	1990-1994	1990-1994	1990-1994	1990-1994	1990-1994

4. Flood analysis

4.1 Purpose and analysis set up

The purpose of the present work is to identify the causes of flooding in the Middle valley part of the Awash River basin, which is a highly developed irrigable area of the basin (Figure 4.1).

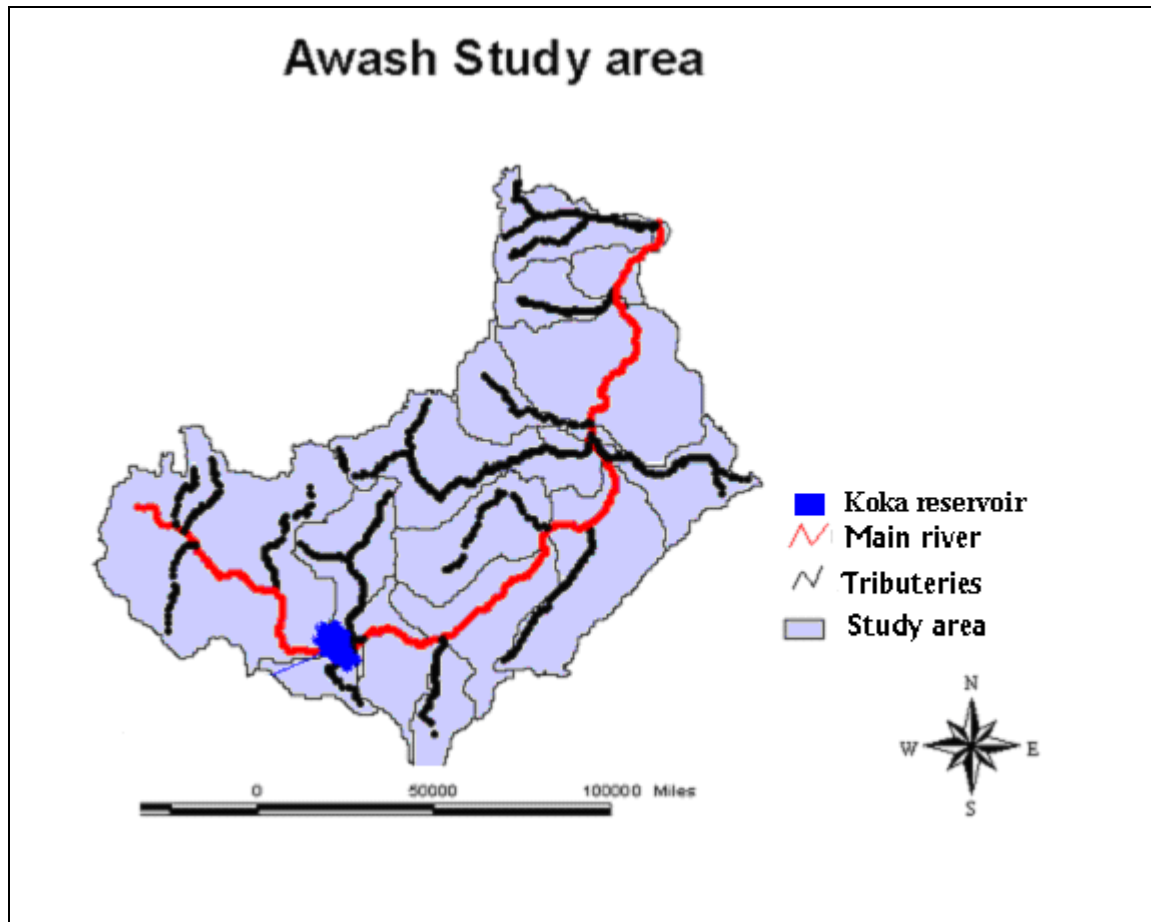


Figure 4.1 Awash study area with potential cause of flooding: Koka reservoir and tributaries

The direct cause of flooding is subjected to discussion following the reports of

1) UN-EUE (United Nation Emergency Unit of Ethiopia) and UNICEF reported in 1999 as follows:

In previous years, especially in 1993 and 1996, the lowlands around Wonji, about 10 kilometers south west of Nazareth town and further down the river around Metehara were flooded and state and private owned agricultural plantations were put at risk. In this upper section of Awash River, the flooding was usually caused by the release of water from the Koka dam's floodgates, which is necessary when the reservoir reaches its maximum capacity. Unfortunately, in previous years the reservoir authorities had trouble in releasing water in controlled intervals, often causing flooding downstream.

This year, mainly two sections of the Awash River with villages, private and state owned agricultural plantations were affected by inundation. The first area is situated in Zone 3 (lower valley) and the second in Zone 1 (upper valley) of Afar Region. Minor flooding also occurred near Metehara town in neighboring Oromiya Region. The actual flooding destroyed approximately 9,500 ha of cropped farmland, both private and state owned (~ 5,000 ha in Zone 3, ~ 4,000 ha in Zone 1, and ~ 230 ha around Metehara, Oromiya Region). The total area flooded this year is unknown and the number of affected population is also difficult to assess and ranges from 3,000 to 35,000 in Zone 3 and from a few hundred to 50,000 in Zone 1.

This year's (1999) flooding along Awash River was mainly caused by heavy rainfall in the eastern highlands and escarpment areas of North Shewa and Welo and not because of heavy rain in the upper watershed areas (i.e. upstream of the Koka Reservoir). Over the years soil and water run-off in the escarpment areas has steadily increased as a result of deforestation, the most serious environmental degradation in the escarpment areas being caused by overpopulation in the highlands. Tributaries to Awash River such as Kessem, Kebena, Hawadi, Ataye Jara, Mille and Loqiya rivers contributed most to the lowland flooding in Afar

2) - Ethiopia - floods

DHA Geneva information report no.4

4 September 1996

i. Water level in koka reservoir has now been stabilized, and discharge from dam has been reduced. River levels downstream have subsequently fallen significantly. Govt considers that awash basin is out of immediate danger of further flooding.

Upper awash

ii. On 27 august, joint UN assessment team (UNDP/DHA, WFP, and UNICEF) visited Upper Awash area immediately below koka reservoir (towns of Nazareth, wonji, and Metahara) and confirmed that situation there was under control.

iii. Despite widespread inundation, damage to sugar cane and other crops is limited to some 400 hectares, according to govt. at sugar plantation in metahara, about 90 km downstream of Koka dam, work has continued to strengthen protective dikes along 2 km stretch of river.

iv. Though 40,000 people have been displaced in Wonji and Metahara, govt states that their immediate shelter, food, and water needs have mostly been met. Main road to Wonji from Nazareth has been re-opened.

From the above reports release from Koka reservoir and high discharge from the tributaries are being mentioned alternatively. To solve this problem a detailed flood analysis is performed.

The study is carried out by making flood analysis with MIKE 11 in combination with GIS facilities for input data interpretation and processing. This modeling system is used to simulate the propagation of flood along the river network. Once calibrated this modeling system should be able to identify the critical cross sections of the river system where inundation may occur.

At this station, the discharge components will be modeled separately. This will show the contribution of koka release and tributaries to the flooding in this region.

4.2 MIKE 11 hydro-dynamic model

The 1D hydraulic model

Hydraulic models provide a 1D description of the flood propagation along a river network using governing equations based on the mass and momentum conservation principle.

Depending on the purpose of the study and depending on local characteristic of flow, different approximations may be applied at the set of fully dynamic wave equation.

The river network is defined by inserting cross-section and hydraulic parameters. Output is provided at all computational point defined along the river network

The modeling process is carried out in three steps:

- Model construction
- Model calibration
- Model validation

The steep riverbed slopes of the basin forced to use dynamic wave approximation of the Saint Venant equations for the description of the flood wave

MIKE 11 is a modeling package for the simulation of surface runoff, flow, sediment transport, and water quality in rivers, channels, estuaries, and floodplains. The most commonly applied hydrodynamic (HD) model is a flood management tool simulating the unsteady flows in branched and looped river networks and quasi two-dimensional flows in floodplains. MIKE 11 HD, when using the fully dynamic wave description, solves the equations of conservation of continuity and momentum (known as the 'Saint Venant' equations).

The solution to the equations is based on the following assumptions:

The water is incompressible and homogeneous (i.e. negligible variation in density), the bottom slope is small, thus the cosine of the angle it makes with the horizontal may be taken as 1. The wave lengths are large compared to the water depth, assuming that the flow everywhere can be assumed to flow parallel to the bottom (i.e. vertical accelerations can be neglected and a hydrostatic pressure variation in the vertical direction can be assumed)

The flow is sub-critical (super-critical flow is modeled in MIKE 11, however more restrictive conditions are applied)

The equations used are:

Continuity:

$$\frac{\partial Q}{\partial x} + \frac{\partial A}{\partial t} = q$$

Momentum:

$$\frac{\partial Q}{\partial t} + \frac{\partial}{\partial x} \left(\alpha \frac{Q^2}{A} \right) + gA \frac{\partial h}{\partial x} + \frac{gQ|Q|}{C^2 AR} = 0$$

where

Q: discharge, (m³/s)

A: flow area, (m²)

q: lateral inflow, (m²/s)

h: stage above datum, (m)

C: Chezy resistance coefficient, (m^{1/2}/s)

R: hydraulic or resistance radius, (m)

α : momentum distribution coefficient

The four terms in the momentum equation are local acceleration, convective acceleration, pressure, and friction (Source: MIKE 11 online help).

In MIKE 11, a network configuration depicts the rivers and floodplains as a system of interconnected branches. Flood levels and discharges (h and Q) are calculated at alternating points along the river branches as a function of time. It operates on the basic information from the river and floodplain topography, to include man-made features and boundary conditions.

Model construction and setup

Model construction is the process of preparing the data in a correct format and entering them in to a set of input data files required by the model code so that the first model run can be made as the first step in the subsequent model calibration.

Due to the steep slope of the basin, the hydrodynamic wave approximation was used. Owing to their mathematical complexity, exact integration of this equation is impossible.

In order to get solutions numerical methods have to be used: MIKE 11 solves the Saint Venant equations using the finite difference and an implicit numeric scheme, applied between consecutive computational points located at a maximum distance dx defined by the user in the river network editor.

Boundary conditions have to be inserted in the model at each upstream end point of each branch and at downstream end of the river network

The first step in a model application is to collect, analyze and prepare the available data in order to make them readable by the program. Data availability and quality substantially affect the model results. The spatial and temporarily scale of available input data should reflect the purpose of the study and detail requested of model output.

In the present study, a considerable effort in terms of work time was devoted to this modeling phase in order to provide a reliable description of the study area and river network.

Data availability

All available data for the present study are listed below

- Geo referenced definition of Awash river network
- River cross-sections
- Georeferenced location of the cross-sections on Awash main and its tributaries
- Daily discharge data

Definition of the river network

Considerable effort was devoted to provide a reliable description of the Awash River basin network below the reservoir. Starting from the river network definition available in ARC-VIEW format, the co-ordinate of more than 1184 points were manually recreated and imported in the MIKE 11 River Network editor. Connecting them the 9 main branches were defined with a reasonable degree of precision in order to reduce the expected discrepancies. The river network as defined in MIKE 11 is shown in figure (Figure 4.2).

Boundary conditions were provided at each extreme upstream end of all branches. The boundary conditions at the basin outlet were defined as a time series of normal depths computed in function of measured hydrograph at Awash Main River.

An important aspect in MIKE 11 is the definition of the connection describing the confluence. These artificial reaches are not seen by MIKE 11 and in order to prevent problems of discontinuities, the connections have to be as short as possible or the three crosssections involved should have the same elevation (Figure.4.3)

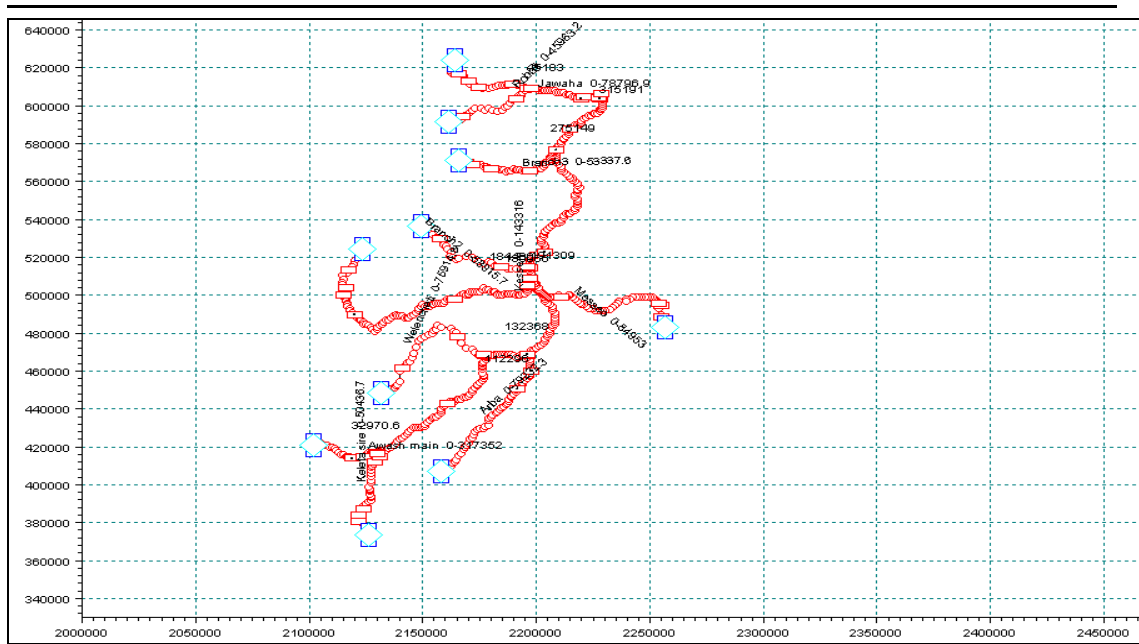


Figure 4.2 The Awash River basin network (vertical and horizontal grids are in UTM)

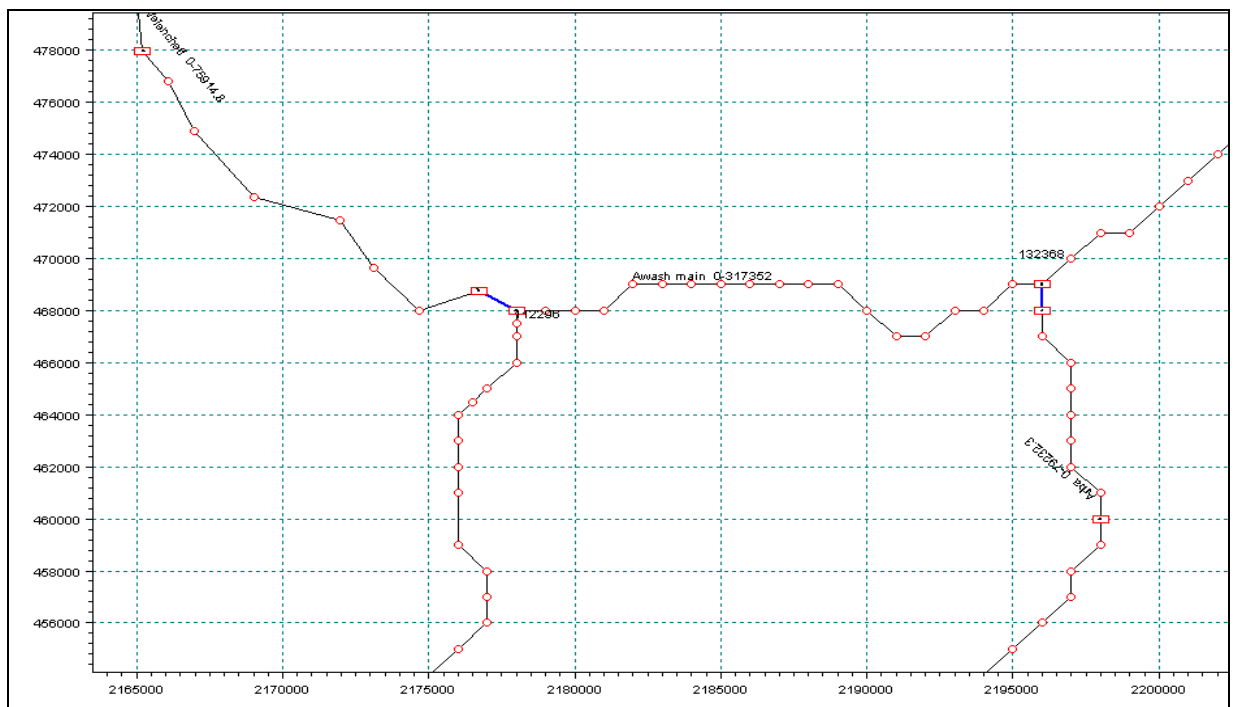


Figure 4.3 Confluence between Awash main, welenchiti, and arba (vertical and horizontal grids are in UTM)

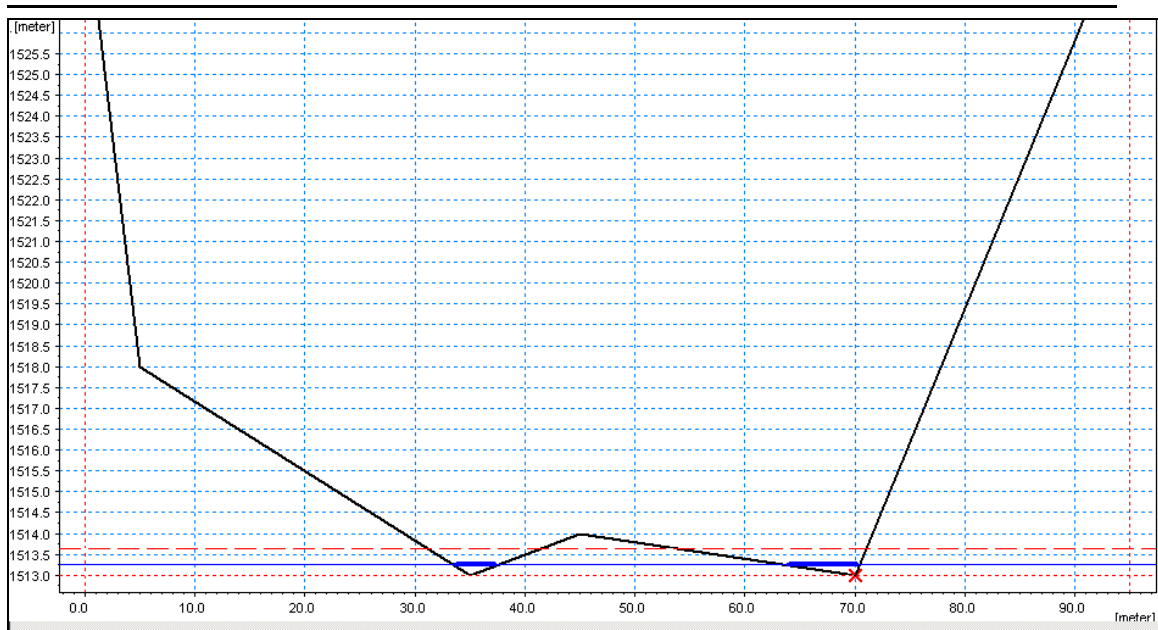


Figure 4.4 Example of a cross-section used to describe the awash main river (horizontal and vertical units are in meter)

The river bed roughness coefficient for Awash and its tributaries were defined according with the value proposed by the official office in Awash basin, Manning coefficient were defined according Chow (1973), counting for the strong mountainous character of the river, the presence of cross section variation and flow observations.

The model setup

The model setup is the process of making the first simulation run and to adjust it for subsequent calibration phase.

Table 4.1 shows the hydrodynamic parameters and variables involved in this modeling phase that normally have to be fixed in order to get the convergence of the numerical scheme used for the computations.

Table 4.1 The hydrodynamic variables involved in the model set up

dt	Time step 30 minutes
dx	Maximum distance between computational points (1000m)
Initial conditions	Water level and discharge
Set of equations	1D

Model calibration

The trial and error method was used for calibration of the present case study implying a manual parameter assessment through a number of simulation runs and a graphical comparison between measured and computed values of daily discharge at the same location.

According to the data availability, the model calibration was carried out comparing the computed and the measured discharge at the main river Awash at MelkaSedi stations located 180km downstream of the reservoir.

The accuracy criteria were chosen to be focalized on a good simulation of the peak flow that is widely the most critical variables in a flood assessment study. According with the purpose of the present analysis, the calibration has been done using as input data for the model that discharge data between 1990 and 1993 (Figure 4.5)

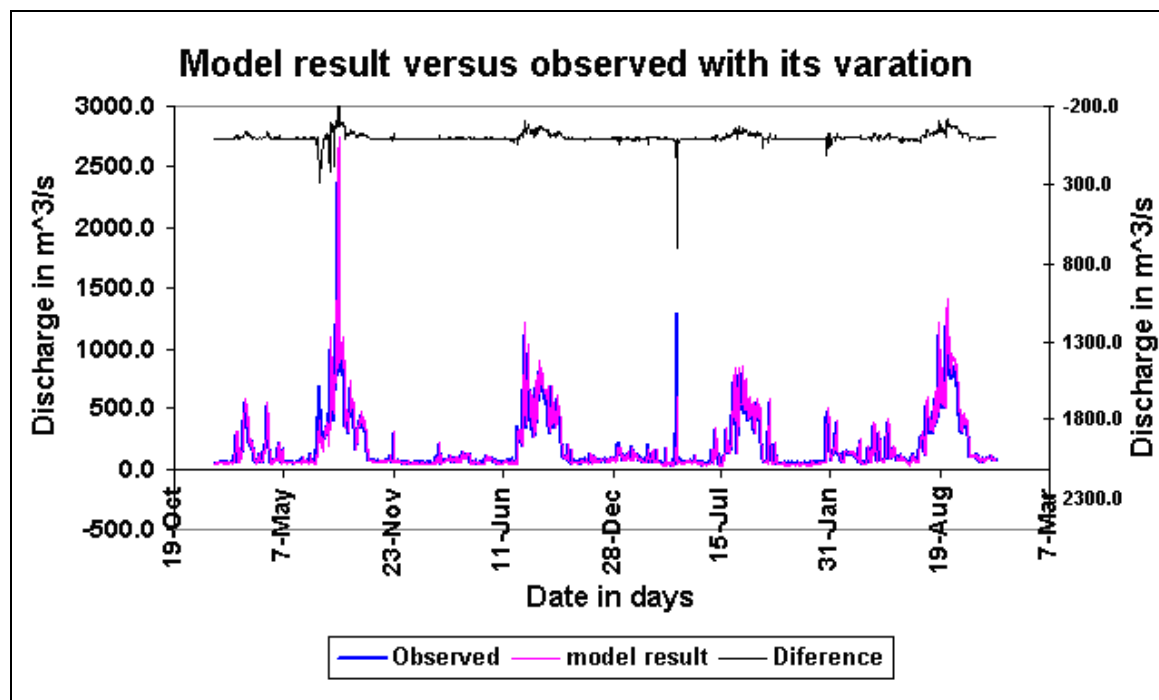


Figure 4.5 Model versus measured result at MelkaSedi station 1990-1993

The parameters involved in the calibration procedure were the manning coefficient a reasonable first coarse set of values were found out according with the physical characteristic system.

Table 4.2 The manning coefficient fixed with the model calibration

Manning coefficient	M^{1/2}/s
Awash main	0.033
Welencheti	0.035
Kessem	0.036
Keleta sire	0.033
Arba	0.033
Jawaha	0.033
Manning coefficient	M^{1/2}/s
Robbit	0.036
Branch 3	0.035
Branch 2	0.028
Messeo	0.025

Based on the analysis of Mike 11 model it can be seen that calculated and observed discharges are in good agreement considering the recession and rising limb of hydrographs. Whereas analysis of the difference between calculated and observed discharge shows an overestimation of model results in the high flow and under estimation of low flows. During wet season the channel or the stream conveyed a portion of peak flood the rest will flow over the bank and conveyed in the flood plain this discharge is not measured by the gauge station which is situated in the river reach and this may be one of the reasons that shows difference in discharge station downstream of the tributaries and summation of discharges in the upstream. During low flow it needs further verification about the water exchange between the river and the bank of the river because during low flow in the stream there may be a flow from the bank of the river to the channel in which the model do not consider during calculation

Model validation

Model validation, an iterative process linked to model calibration, is another important step in the model building sequence. It is implemented by fitting model with information other than that used in model calibration process in order to test the applicability of the model to other time step period. In this study, the simulation result of year 1994 is checked against the observed data. During the model validation, the parameter i.e. Manning are not adjusted anymore because the result seems good and it is depicted in Figure 4.6

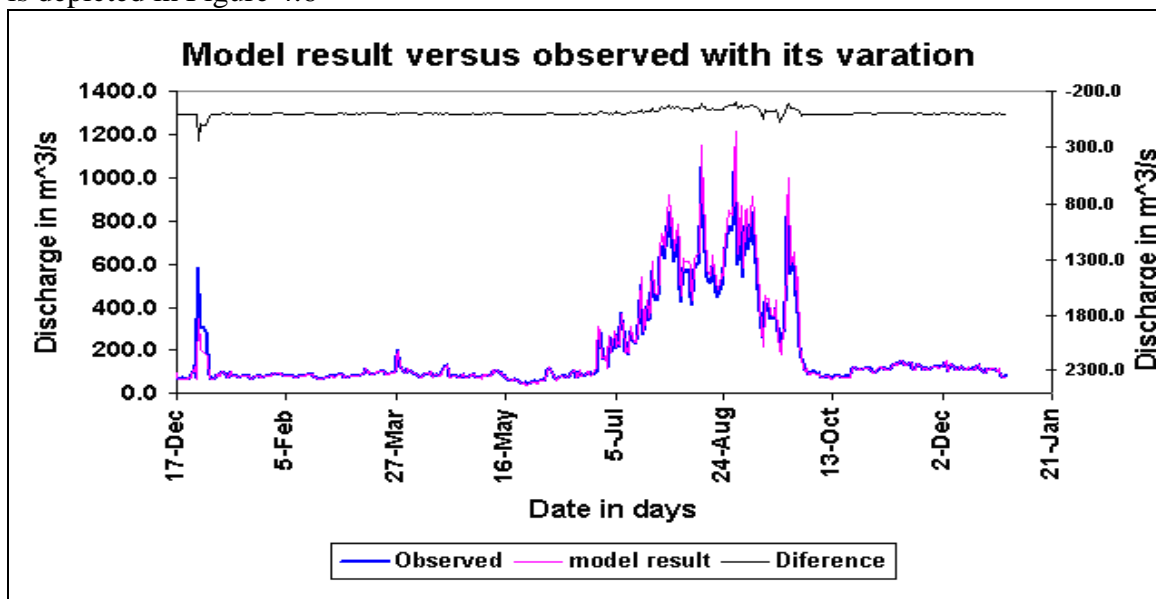


Figure 4.6 Model versus measured result at MelkaSedi station in 1994

4.3 Results analysis

Contribution to flooding of koka reservoir release and tributaries with MIKE 11 1D model successfully prepared, flood analysis can be done. First the critical location for flooding is determined and then the contributor to the flooding are modeled separately to asses their significance. And based on this critical section using different discharge values i.e. combining the discharge of tributaries and release discharge of reservoir,

only release discharge from reservoir and only tributary discharges are taken in the model as boundary conditions in order to see their significance in contributing to the flood.

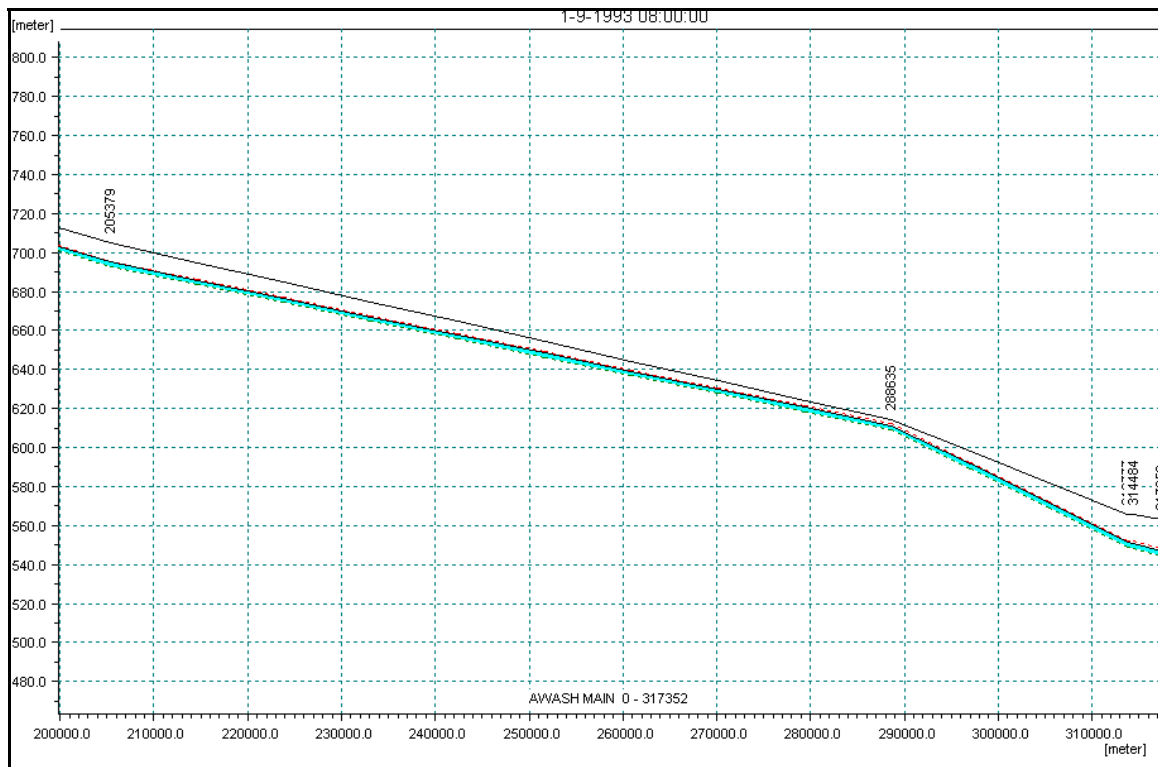


Figure 4.7 Water level result only using release discharge from the reservoir around MelkaSedi station

Ignoring the tributaries discharge (making $Q=0$) and using only release discharge from Koka to see the water level in the main river around MelkaSedi and to this its effect in contribution of flooding during tributaries discharge hydrograph.

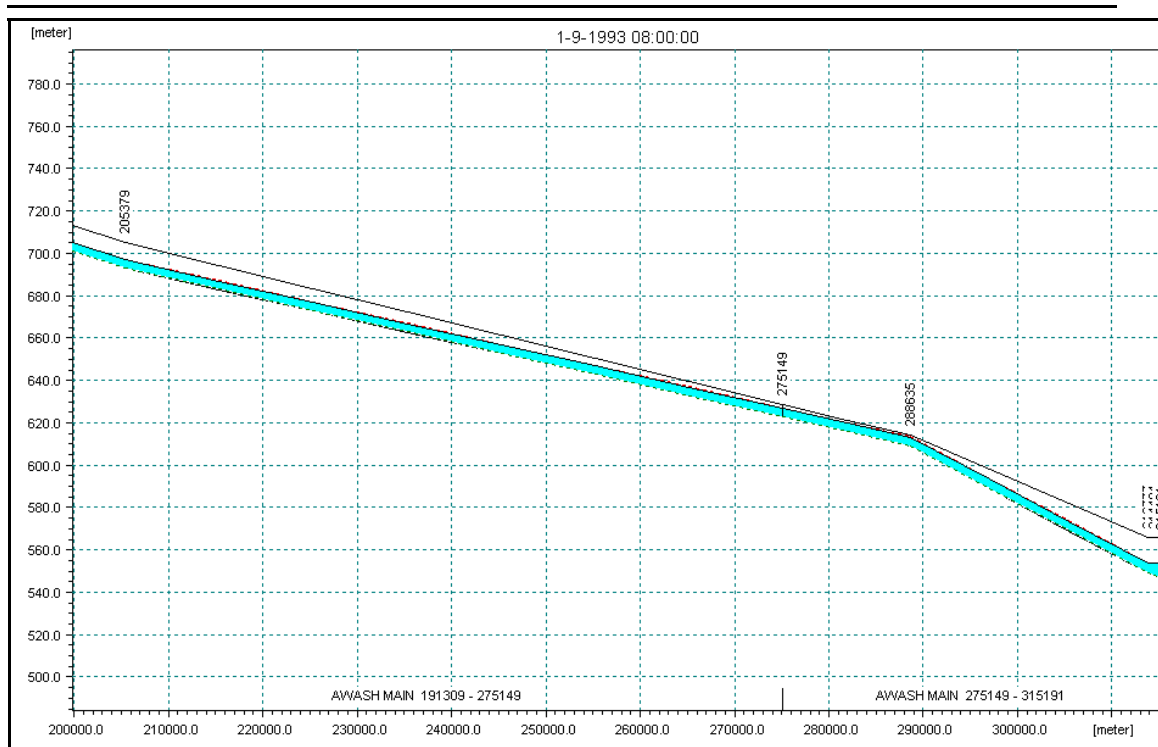


Figure 4.8 Water level profile only tributaries discharge

From Fig 4.8, it can be seen that if the Koka reservoir release is set to zero and using only the discharges of tributaries, it is highly critical at the identified crosssections. This shows that without the flow from the reservoir it reaches maximum capacity of the conveyance considering some level as free board.

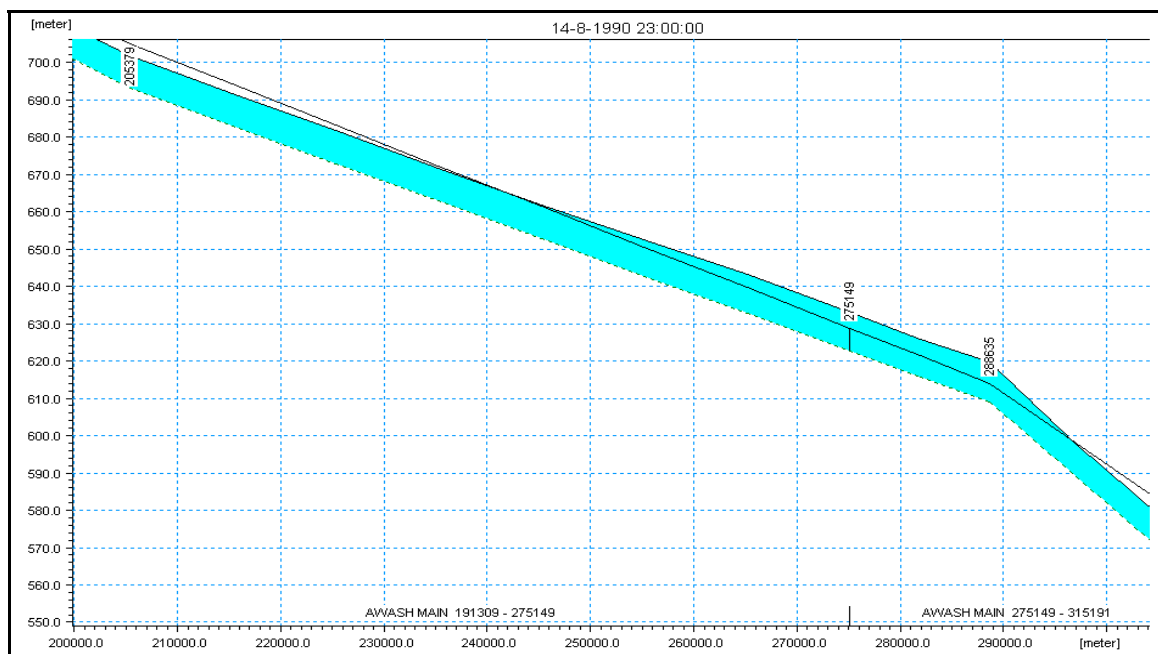


Figure 4.9 Water level profile with combined discharges, 1990, around MelkaSedi stations

From Fig 4.9, it is clearly showed that the combined discharge can not always be conveyed by the channel unless it is supported by the flood plain at that critical section and downstream and upstream reach of the sections.

4.4 Conclusion and Discussion

Conclusion

The part of the study area downstream of Koka reservoir has been modeled successfully with MIKE11. Specially the rising and falling of the reaches of hydrograph show good agreement in the validation. The flood analysis with MIKE 11 shows that the cause of the flood in the Awash basin is the combined effect of both the release discharge from the reservoir and high discharge from the upstream tributaries. In 1990, the flood was mainly due to high flow in tributaries and the other years flood was due to combined flow from reservoir and tributaries. It is known that the stream channels do not convey the extreme peaks of flow unless it supported by flood plain.

Discussion

As it is shown in the MIKE11 the water level of the simulated flow is much higher than the top bank level of the river during wet seasons this indicates the conveyance of the main river is not enough to convey or store the flow during wet season this may be due to sediment load which will decrease the conveyance or due to the higher flow. On the other hand, the measured result in some of the station shows lower value of this wet season data than its upstream stations this is due to the overflowing of the river above staff gauges. Visual observation on the flood events indicate that a higher rainfall records do not accompany the peak flow even this case is valid by the historical flood event data.

The model result versus measured result of low flow needs further verifications because the model result do not consider the bank exchange or flow exchange between bank of river and stream flow on the other hand in actual situations depending on the geological and soil conditions there will be a flow exchange between stream flow and bank of river.

5. Reservoir operation

5.1 Analysis set-up

The purpose of the present work is to develop reservoir operation strategies that will reduce flooding downstream of the koka reservoir without affecting power generation potential of the dam.

First, a reservoir model is built. Because no data on the design of the release structures in the dam and on the inflow to the reservoir is present, additional step to prepare the model are needed.

- i. Generate inflow time series with conceptual rainfall runoff model and inverse modeling from release time series
- ii. Make assumption on the control structure in the dam
- iii. Prepare reservoir model

Second release schedule scenarios that reduce flooding downstream of koka reservoir are developed using discharge time series analysis and the MIKE 11 model from chapter four.

Third, the effects of the release scenarios on the reservoir water levels are determined with the reservoir model and downstream flooding with MIKE 11

Fourth the risks for power generation and the effects on are discussed

5.2 Reservoir inflow time series

5.2.1 GR3J: a daily three-parameter rainfall-flow model

Model description

GR3J (which stands for modèle du Génie Rural à 3 paramètres Journalier) is a daily lumped 3-parameter rainfall-flow model. It belongs to the family of soil moisture accounting models family. A full description of the model can be found in (<http://dataserv.cetp.ipsl.fr/AIMWATER/reports/userRequirements.pdf>)and(<http://www.antony.cemagref.fr/webqhan/projets%20themes/Hydrologie/Code%20fortran.htm>)

The model needs daily precipitation and Evaporation data . P (precipitation) is an estimate of the area catchment rainfall that can be computed by any interpolation method from available rain gauges. E can be a long-term average value, which means that the same PE series is repeated every year.

All water quantities (input, output, internal variables) are expressed in mm, by dividing water volumes by catchment area, when necessary. All the operations described below are relative to a given time step and correspond to a discrete model formulation (obtained after integration of the continuous formulation on the time step).

Model parameters

In the GR3J model, four parameters have to be defined:

X1: groundwater exchange coefficient	(mm)
X2: maximum capacity of the production store 330 mm (given)	(mm)
X3: One-day ahead maximum capacity of the routing store	(mm)
X4: time base of unit hydrograph UH1	(days)

All four parameters are real numbers. X2 and X3 are positive, X4 is greater than 0.5 and X1 can be either positive, zero or negative. In GR3J X2 is fixed, and other 3 have to be optimized

Model construction and setup

Five year discharge data of upstream station at MelkaHombole and area/average precipitation of rainfall is taken as the input data for the model and one year data for warm-up the model and the other four year data for calibration purpose is used in the study for predicting inflow to the reservoir and then using this data to extrapolate for the whole basin for predicting inflow. The comparison of result of the model with measured data is depicted below

Table 5.1 Parameter values for calibration

Parameters	Parameter value
X ₁	1.5
X ₂	330mm
X ₃	250mm
X ₄	2days

Calibration is made by taking different value of ground water exchange coefficient and one day ahead maximum capacity manually and based on the calibration result the model predicts the peak value well but it predicts bad the low value.

From the general view of model building, sensitivity analysis is one of the important processes in providing a reliable model. Via this process, we can understand how a model responds to parameter changed and, in particular, to identify the impact that various parameters and process have on the computed response. In this study based on the background knowledge, experiences and results from model calibration, it can be seen that the groundwater exchange coefficient and one-day ahead maximum capacity of the routing store are key parameters that need to be justified for good model performance. However, sensitivity analysis hasn't been carried out for both above parameters in this study. One reason behind this is that sensitivity analysis is not crucial process considering the primary objective of the study.

Result discussion

The model result shows that it predicts the high flow well but it highly exaggerates the low flow especially after the peak during recession limb (Figure 5.1). This indicates

that the ground water exchange coefficient parameter is highly sensitive and needs to be optimised. Of course this coefficient value is determined based on the FAO soil database from Global Soil Data Products to relate the soil unit and the corresponding soil property. Therefore it will not give it will not give feasible result. Therefore this model cannot be used for producing inflow time series for the Koka reservoir. Because it produce the peaks reasonably well it could be further developed to use for inflow forecasting.

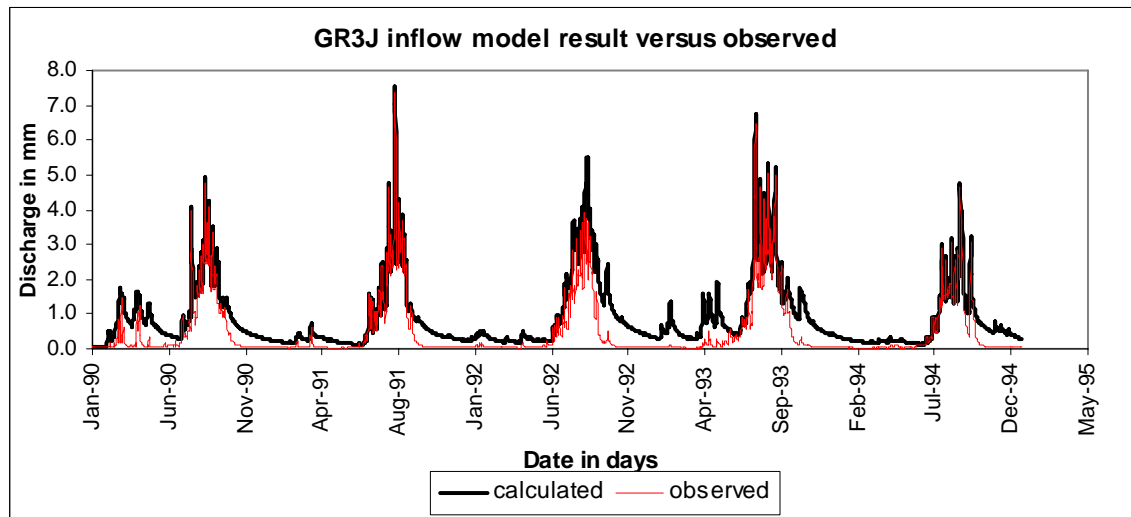


Figure 5.1 GR3J model result at MelkaHombole station 40 km from the reservoir

5.2.2 Inverse modeling

From the GR3J model result at MelkaHombole inflow volume from 8470 km² catchment area and measured result it is seen that especially in low flow the model result exaggerates the values that means the input to the model does not express the real catchment exchange properties of the soil or the bank of the river and the catchment. Taking this in mind, the downstream measured out-flow and upstream measured loss mainly due to evaporation are taken as input for inverse modeling of inflow and the result is compared in volume wise. From the comparison of volume of 5 years of outflow data from inverse modeling and measured result it indicates that the catchment below MelkaHombole station contributes to the reservoir in an average value of 15%. That means the inflow from total catchment above Koka reservoir is calculated by 1.15 times measured data at MelkaHombole at 40 km station. The time series is driven by lagging the MelkaHombole station time series and then multiplying by 1.15 each daily discharge (Figure 5.2).

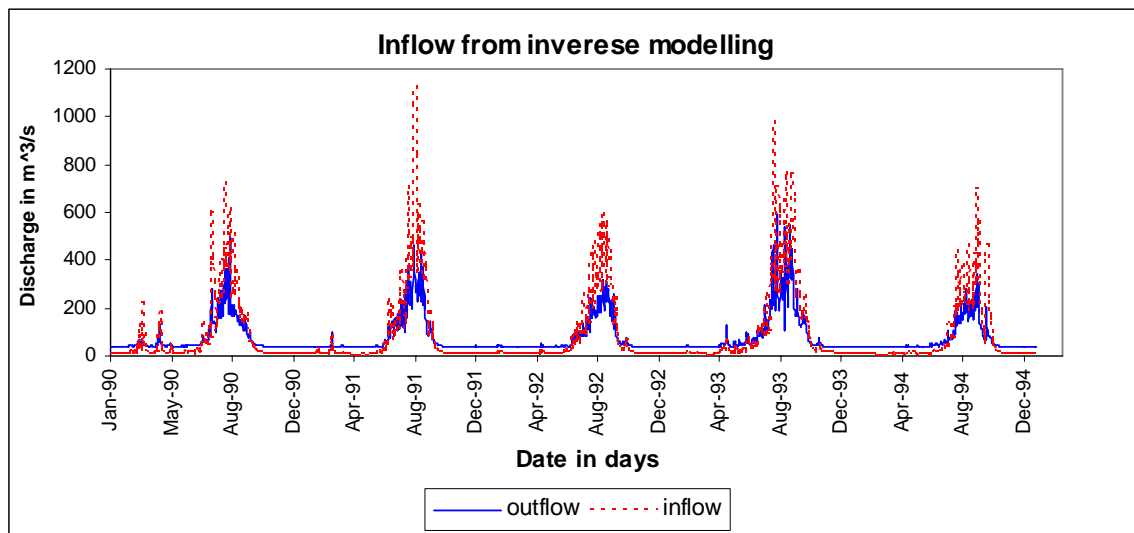


Figure 5.2 Inflow Time series from inverse modeling and measured outflow and losses from Koka

Discussion: The outflow loss from the Koka reservoir station and derived inflow from the inverse modeling are in good agreement in timing of peak. Of course, due to the effect of the reservoir, the peak outflow is lower and it has also high low flow due to the regulated release to fulfill downstream purposes. The volume calculated from the inverse modeling and from the koka, release is also equal value.

5.3 Reservoir model

Based on the above model result of inflow and taking some assumption on missing data like bottom outlet diameter, minimum water level, minimum power requirement and existing operation rules a reservoir model is built. This data are assumed based on theoretical background.

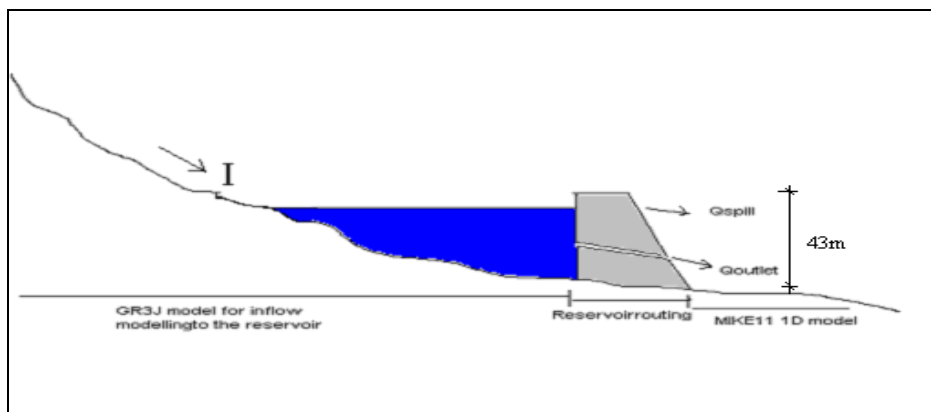


Figure 5.3 idealization of reservoir operation

5.3.1 Reservoir operation

The operation of a single reservoir for a single function does not present many analytical problems, but the same is not true when a reservoir fulfils a number of potentially conflicting objectives or where several reservoirs are operated

conjunctively. Maximizing the benefit of one or more objectives of a reservoir, needs balance information on the likely resource against the multiple demands on it, while maintaining the safety of the dam in times of severe flood.

In the reservoir with controlled outflow, gates are used for the purpose of regulating flow through the outlet structure. The gates are operated following established rules and this rule determines the relation between inflow, outflow and storage. In this study, the only data available are the inflow hydrograph (upstream boundary condition), initial outflow and storage (initial condition) and physical and operational characteristics are assumed on theoretical background and experience.

Since real reservoir usually have a non-linear storage-outflow relation, the non-linear properties of the storage-outflow relation is determined using the following formula:

$$S = f(O)$$

$$S = KO^n$$

A general hydraulic outflow formula is:

$$O = C_d ZH^Y$$

In which O is outflow

C_d is coefficient of discharge

Z is area of a conduit or length of spillway

Y exponent of rating

Theoretical value of discharge coefficient and rating exponent are determined using hydraulic principles.

Based on this an equation for outflow from Koka reservoir is established with assumption on bottom outlet diameter and position of outlet. And using storage induction method flow routing is undertaken. The following tables are prepared for routing

- i. Elevation versus storage based on topographic information with some modification
- ii. Elevation versus outflow based on hydraulic properties
- iii. Storage versus outflow based on the above information
- iv. Storage induction versus outflow

Formula for routing

$$\frac{2S_{n+1}}{dt} + O_{n+1} = I_n + I_{n+1} + \frac{2S_n}{dt} - O_n - \bar{O}_r$$

In which S storage

O outflow

I inflow

\bar{O}_r target release

Based on this routing it is possible to generate a release outflow, which can be operated in reservoir outlet in such a way as to absorb much of the flood and release less in August and release more water in July in advance and September with controlled rate, which does not cause flooding downstream.

The principle of reservoir operation should be applied with to keep net system storage proportionally distributed (shared), water supply demands and ending water level same as or better than starting water levels.

The Reservoir is operated based on giving high priority to power generation and risk of overtopping as well with several constraints.

Constraints applied for the Reservoir operation are as follows:

1. Monthly water level \geq Operational minimum water level
2. Monthly water level $<$ Operational maximum water level
3. Percentage monthly effective capacity equal with determined maximum capacity
4. Final water level of reservoir \geq Initial water level
5. Discharge outflow reservoir \geq water demand downstream

5.4 Reservoir release scenario's

High release from Koka occurs between July and September. With the peaks mostly in August, release scenario's aim at reducing the peak discharge from Koka reservoir.

From the discharge time series different scenarios are derived based on the power requirement schedule and flooding hazard downstream. Taking in to consideration considering the data unavailability and age of reservoir 30% of its volume is taken as sediment volume from upstream area in addition to dead storage volume so a total volume of 0.8 billion m^3 is considered as minimum volume for the analysis. Based on this scenarios are derived from the existing release volume by shifting August release discharge in advance and increase release discharge in July and September keeping total volume of water released with the existing conditions. The released scenarios are shown in Figure 5.4.

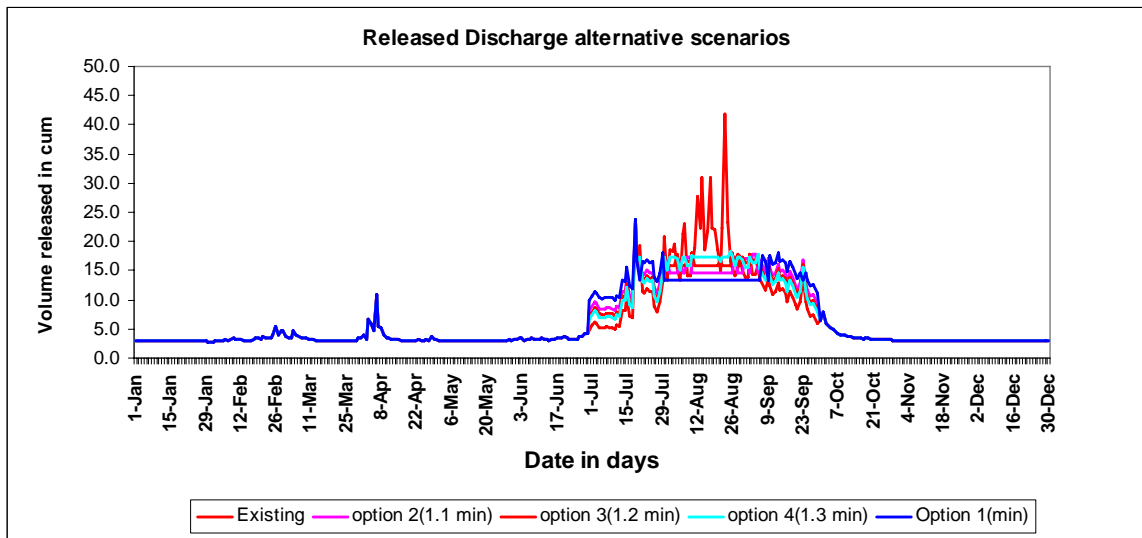


Figure 5.4 Different scenarios (minimum of August as reference and then 1.1* minimum etc)

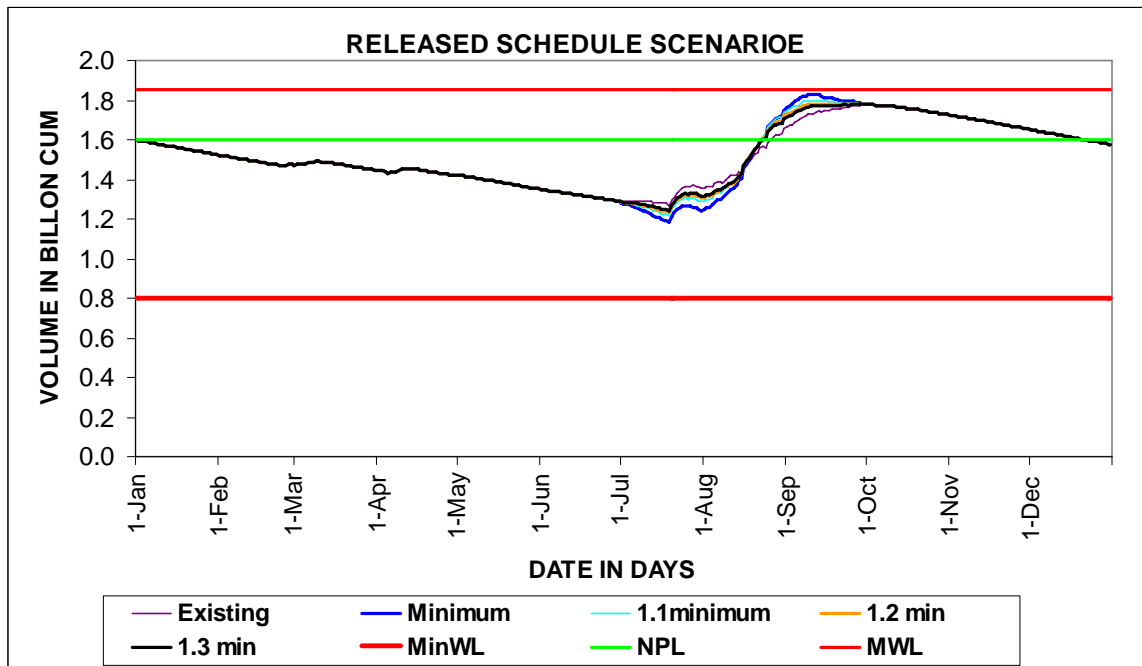


Figure 5.5 release schedule scenarios checked with minimum volume of water level for power generation

The scenario with the maximum flood reduction within the constraints is determined through the iteration process described in the flow chart (Figure 5.6)

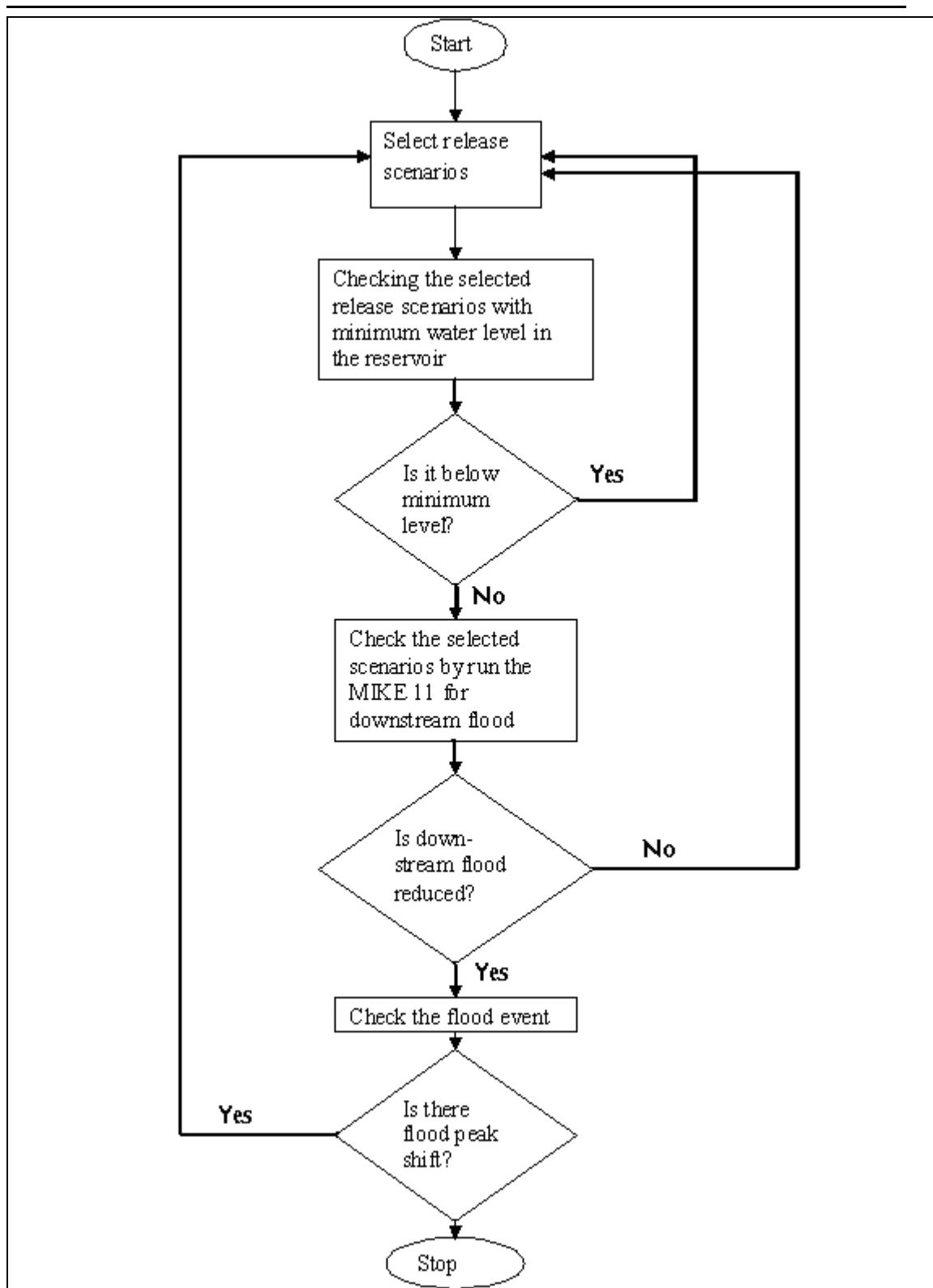


Figure 5.6 Flow chart for advance release scenario

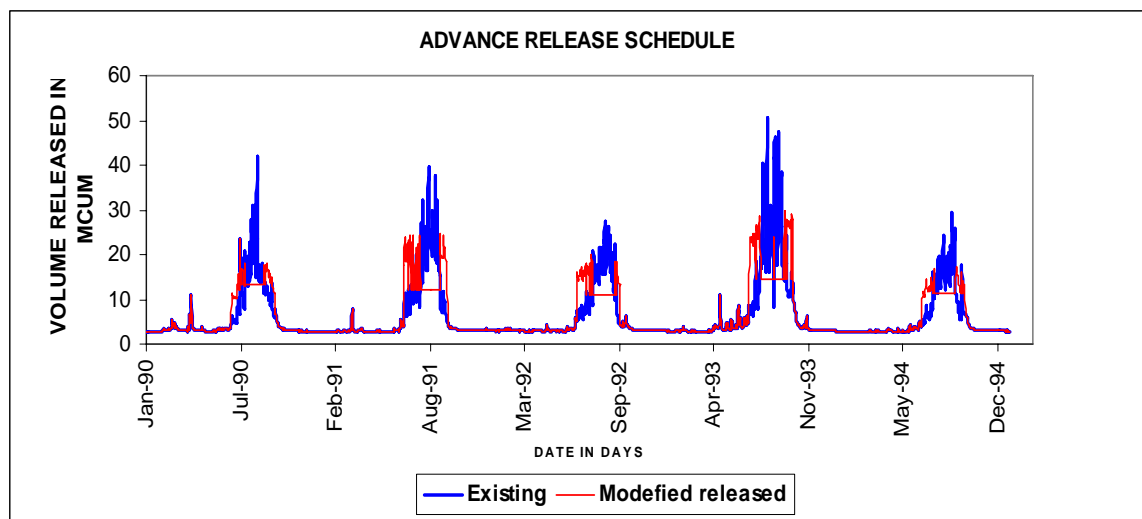


Figure 5.7 five-year advance released schedule

The above four scenarios (Figure 5.4) are checked by MIKE 11 model for their effect in downstream flood control.

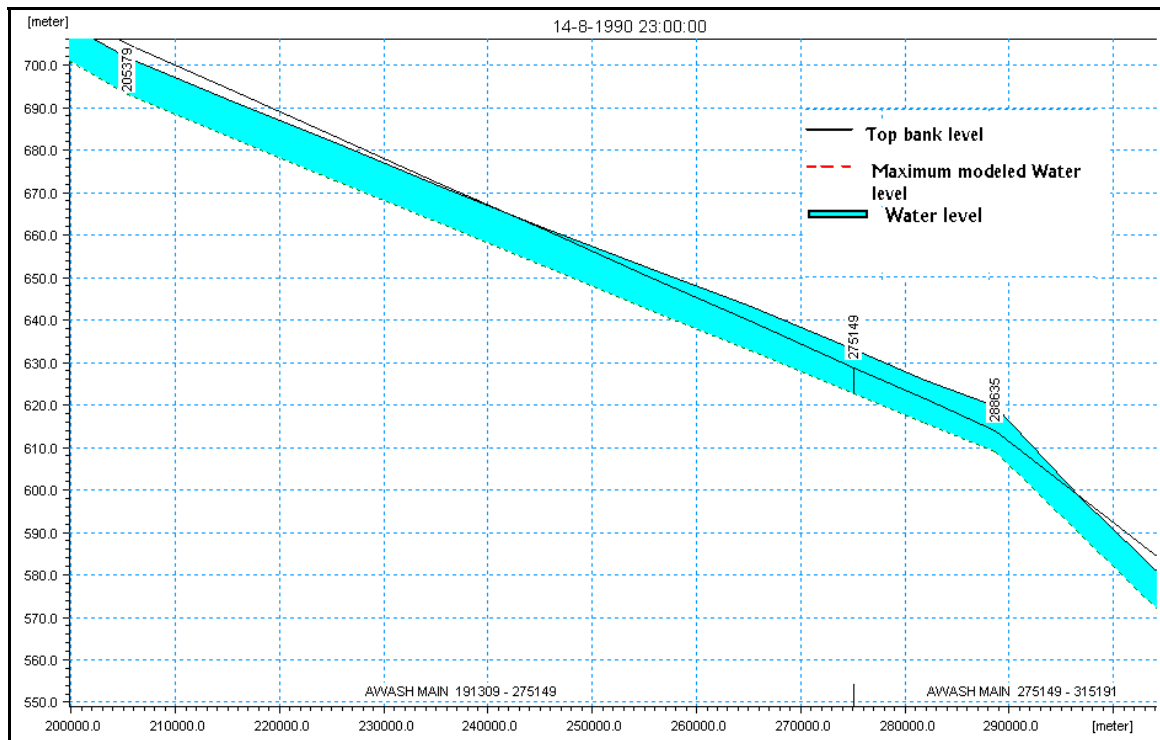


Figure 5.8 a) Mike 11 water level profile with advance release, 1990, around MelkaSedi

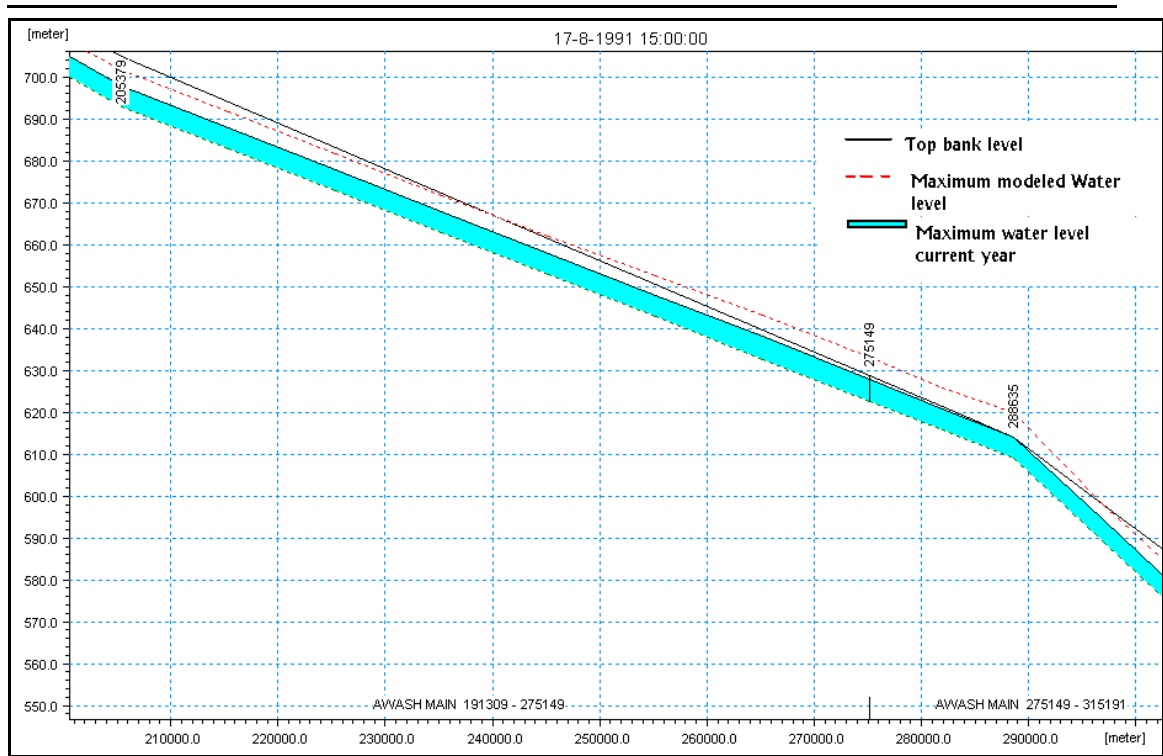


Figure 5.8b) Mike 11 water level profile with advance release, 1991, around MelkaSedi

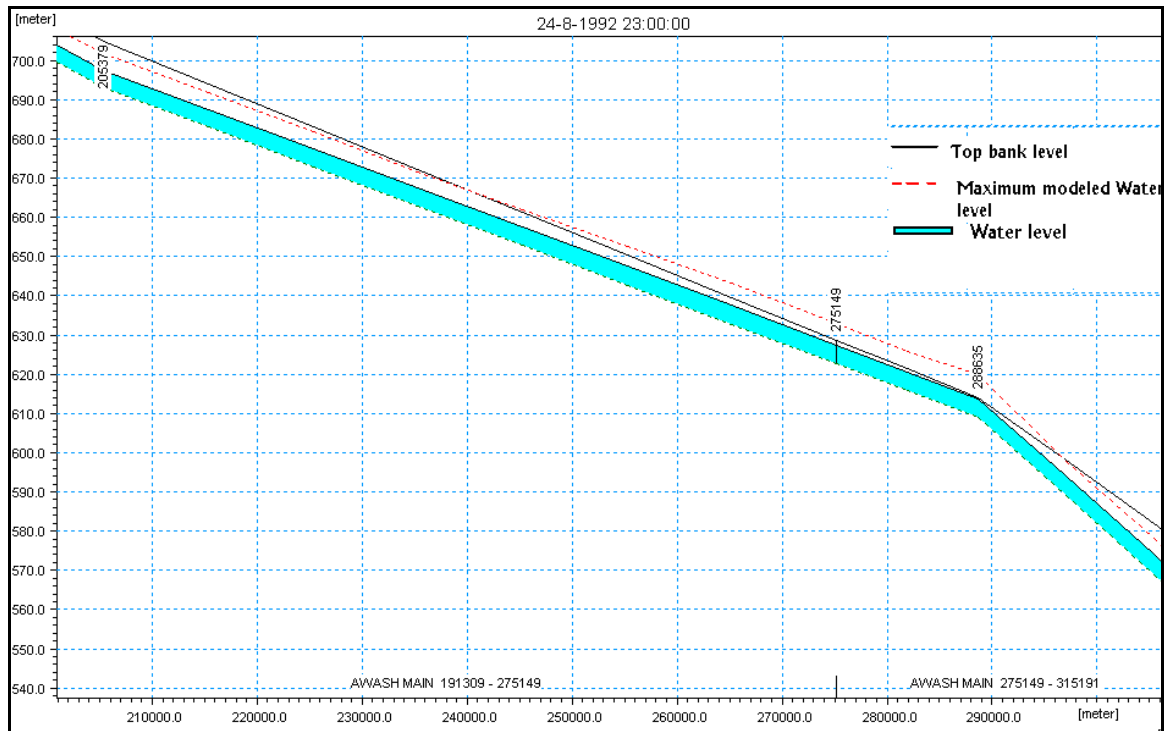


Figure 5.8c) Mike 11 water level profile with advance release, 1992, around MelkaSedi

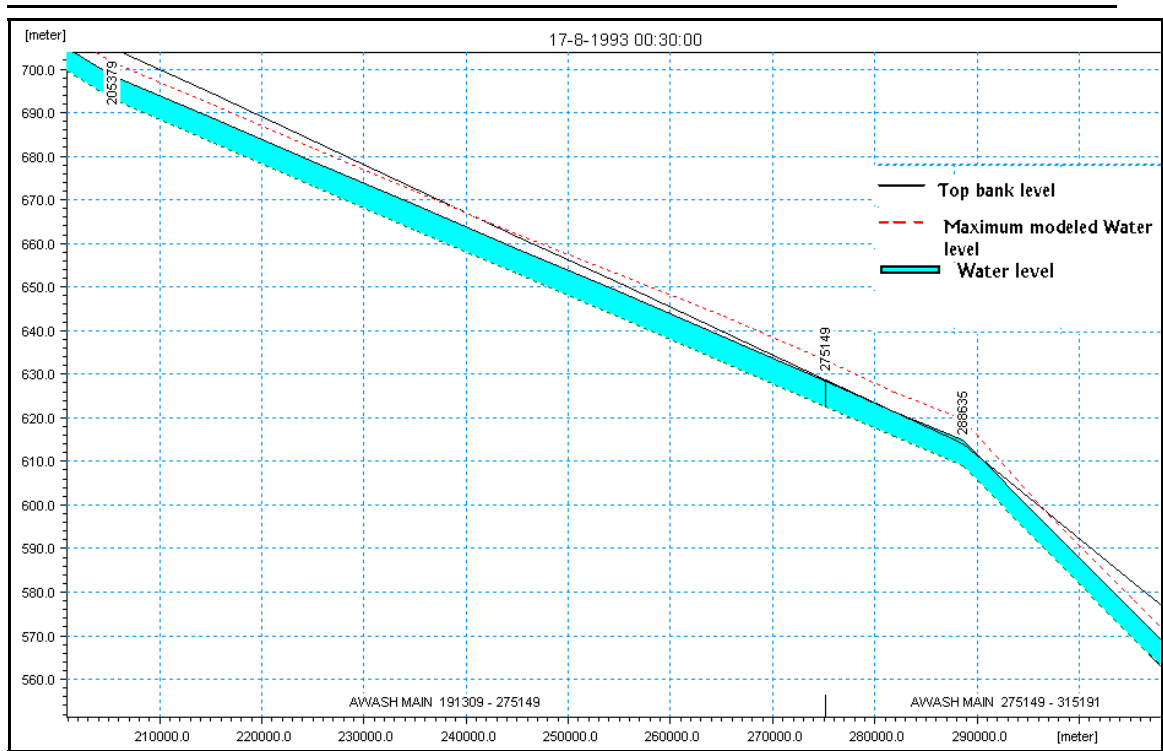


Figure 5.8d) Mike 11 water level profile with advance release, 1993, around MelkaSedi

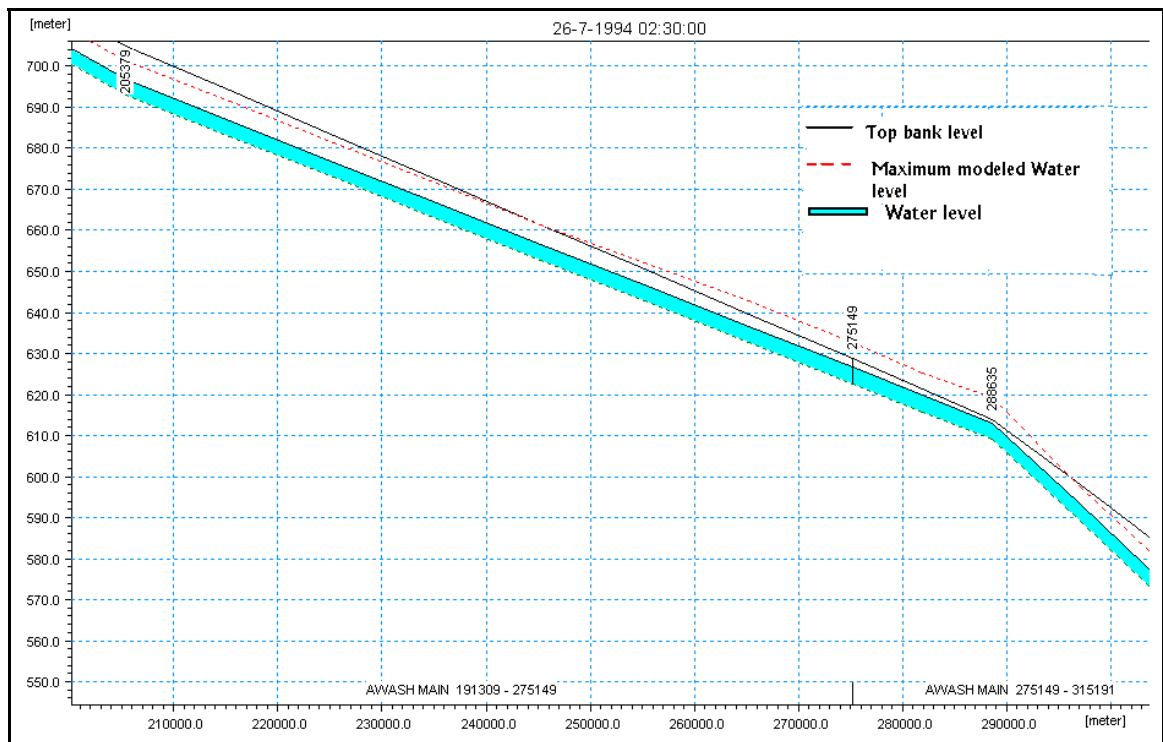


Figure 5.8e) Mike 11 water level profile with advance release, 1994, around MelkaSedi

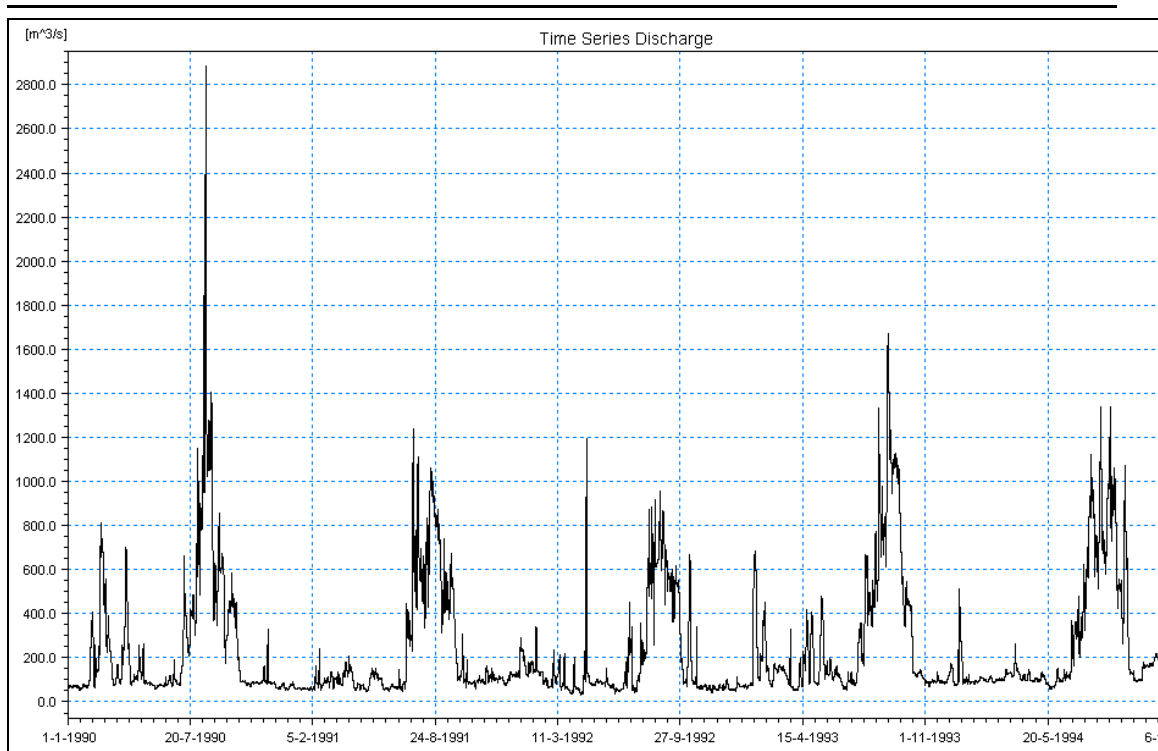


Figure 5.9 Discharge time series at critical cross section with existing release (around MelkaSedi)

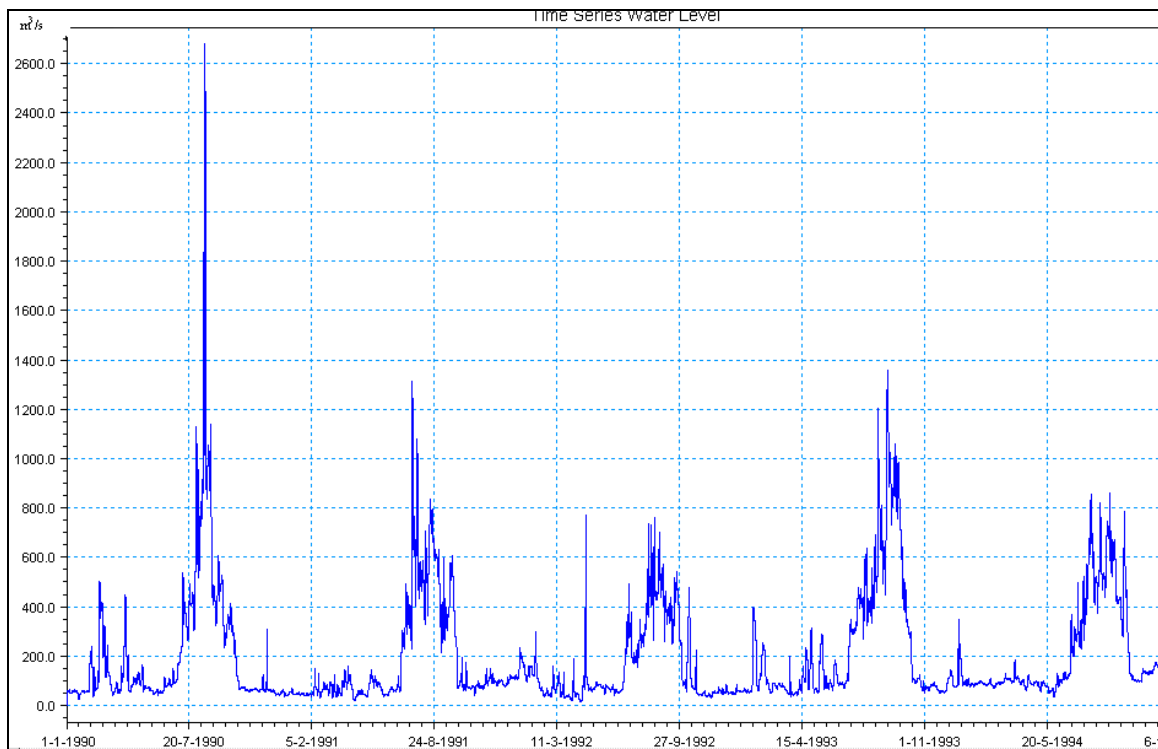


Figure 5.9 Discharge time series around MelkaSedi with advance release

RESULT: In Figure 5.5, the rule curve of reservoir operation is presented with the four alternative release schedules. The optimal one selected with the criteria of maintaining the effective volume released and keeping the water level always above the minimum water level required for power generation. Towards the end of the year, the water

level should be restored to that before the operation of the reservoir. The result will be a reduction in the flooding in downstream of the reservoir. Here another important criteria in addition to the reduction of the flood is that it will not shift the flood to July or September. By applying the criteria above, the release schedule using a minimum release in August and releases in July and September determined by multiplying 3% of the existing release is selected as best option for flood management.

From MIKE 11 result, it shows that in year 1990 already the tributaries discharge is dominate to flooding and the other years around 280 m³/s flow reduction on the peaks are obtained using advance release(Fig 5.7, Fig 5.8, Fig 5.9, Fig 5.10).

DISCUSSION: As been described previously, technically there are many problems occurring in operating the rule curve of reservoir operation. The main problem is the hardly predictable climate condition by the time of operating the reservoir in order to full fill the water needs. Discharge inflow into the reservoirs is almost too difficult to be predicted, due to the effect of land use change along with the population and economic growth. Land use change will clearly affect the hydrograph pattern.

Inflow forecasting allows for more flexible reservoir operation. The GR3J model that has been developed can be enhanced to provide reliable inflow forecasts.

5.5 Conclusion

The existing operation is mainly based on power generation and safety of the reservoir. The principle of following heavy rainfall runoff from the upstream catchment cause large inflow discharge to the reservoir at times where the water level reaches or close to the maximum excess water is therefore discharge without considering the peak hydrograph downstream in addition to tributary discharges. This aggravates flooding downstream and increases the construction cost of structural measures for flood defence. From the above result, it can be concluded that by managing the reservoir operation system it is possible to reduce the flooding risk downstream in combination with structural measures at economic cost. The existing release schedule of the reservoir should be changed without risking the power generation. Further reduction in the discharge released in August and more released in July will shift the flooding hydrograph from August to July.

Here it is concluded that to limit flooding in addition to structural measures for flood defence in the basin, the water level in the reservoir could be lowered in advance, provided the inflows can be forecasted.

6. Flood Forecasting

6.1 Set up analysis

In chapter four, it was shown that the tributaries discharge is significant or even critical for floods. In addition to structural measures to reduce flooding, forecasting and early warning systems are necessary. The objective is using the rainfall data and discharge data known now and before, of tributaries and the main river to predict the discharge 3 days ahead at MelkaKunture station with special concern about flood prediction. The steps are:

- i. Input data to the model is prepared applying statistical parameters and separating the data in wet season and full year data
- ii. Forecasting using M5 model tree
- iii. Replicate with same input with ANN

6.2 Data preparation

Correlation analysis based on the area average rainfall

The correlation of the predicted downstream discharge QMS $t+3$ (MelkSedi station) with area average effective rainfall (RE), moving average of area average effective rainfall (RE Mov t), upstream and tributary discharge (Qa, QAM, QMesso) with different time lags has been analyzed. This analysis is based on two types of data sets, one being the continuous full year data and the other one is the flood season data.

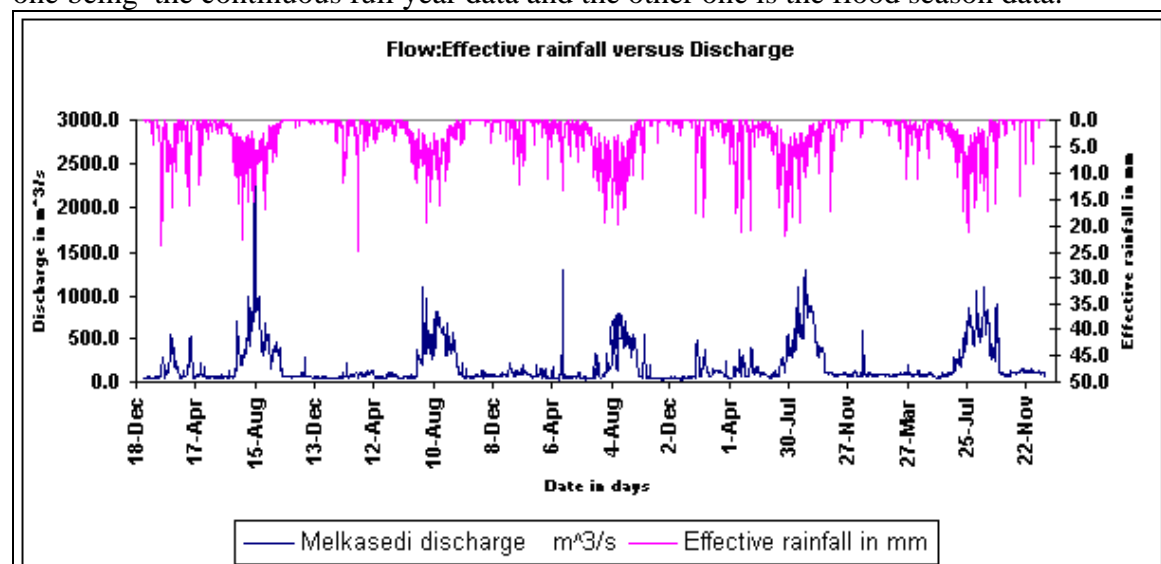


Figure 6.1 Effective rainfall versus discharge at MelkaSedi station

Here effective rainfall is calculated by subtracting evaporation from the area rainfall in the catchment

Table 6.1 The correlation coefficient of relevant variables and time lags with predicted discharge of full year data downstream of koka reservoir

	Time lag in days					
	0	1	2	3	4	5
QA	0.78	0.76	0.74	0.72	0.69	0.67
QAM	0.47	0.44	0.44	0.43	0.43	0.42
QMesso	0.62	0.59	0.57	0.55	0.56	0.55
RE	0.50	0.49	0.48	0.49	0.48	0.45
REMov2	0.55	0.54	0.54	0.54	0.52	0.50
REMov3	0.58	0.58	0.571	0.57	0.56	0.54
REMov4	0.60	0.60	0.60	0.59	0.58	0.57
QMS	1.0	0.97	0.93	0.91	0.87	0.84

Table 6.2 The correlation coefficient of relevance variables and time lags with predicted discharge of wet season data downstream of koka reservoir

	0	1	2	3	4	5
RE	0.11	0.10	0.09	0.14	0.12	0.09
RE mov2t	0.13	0.11	0.14	0.16	0.13	0.09
RE mov3t	0.13	0.15	0.16	0.16	0.12	0.10
RE mov4t	0.15	0.16	0.17	0.16	0.15	0.13
QA	0.57	0.53	0.50	0.46	0.41	0.38
QAM	0.30	0.24	0.26	0.23	0.24	0.21
QMESO	0.49	0.44	0.43	0.39	0.41	0.40
QMS	1.0	0.96	0.93	0.90	0.87	0.84

Correlation of whole year data,(1990-1994)

As shown in Table 6.1, the correlation of the continuous year data is summarized as following;

The maximum correlation coefficient of upstream main river QA at Awash station and predicted discharge at MelkaSedi is 0.78 and time lag is 0 day because it is near to the predicted station.

The maximum correlation coefficient of upstream main tributary at AweraMelka (Kessem River) station and predicted station is 0.47 and time lag 0 days.

The maximum correlation coefficient of effective precipitation and predicted station is 0.50 and time lag 0 days

The maximum correlation coefficient of 2 days moving average effective precipitation and predicted station is 0.55 and time lag 0 days

The maximum correlation coefficient of 3 days moving average effective precipitation and predicted station is 0.58 and time lag 0 days

The maximum correlation coefficient of 4 days moving average effective precipitation and predicted station is 0.60 and time lag 0 days

The maximum correlation of moving average of precipitation of different days are arranged from 0.55 -0.60, the time lag is 0 days, among which the maximum value is four day moving average and the discharge at MelkSedi correlation coefficient

Selecting training, testing and validation data sets

The hydrological characteristics of the 1990-1994 time series are shown in table 6.3. The selection of the training and testing data set is according to the statistical characteristic of year maximum discharge, year rainfall depth, runoff depth and the rainfall runoff coefficient derived in the table. Both the maximum and minimum value should be included in the training data set. The ratio annual rainfall depth and annual runoff depth is used as the arid index, the average value of 5 year is 0.18.

Table 6.3 The hydrological characteristic of 5 year time series

Year	Maximum discharge (m ³ /s)	Total annual rainfall (mm)	Annual total runoff (mm)	Rainfall runoff ratio
1990	2455.7	1027.0	191.25	0.18622
1991	1082.8	861.8	160.66	0.18642
1992	1299.7	1143.2	213.98	0.18718
1993	1298.7	1138.6	212.07	0.18626
1994	1099.0	807.1	150.53	0.18652
Average	1534.225	1042.65	194.49	0.18652

Table 6.4 The selected training and testing data type

Model type	Type		
	Training	Testing	Validation
M5 model tree	1990-1993	1994	
ANN	1990-1992	1994	1993

6.3 Inputs selection

System state parameter selection

Any modeling tool requires accurate identification of input and systems variables. As the objective of the model is to predict the discharge downstream, it involves the physical process of rainfall-runoff yield and runoff routing. All the system state parameters related to it should be considered, such as rainfall, evaporation, soil characteristic, upstream discharge of the main river and tributaries etc.

In this study besides rainfall, one of upstream main river discharge and two tributaries and the discharge of the predicted station known before is used.

The following consideration should be mentioned

-
- The data driven model work with total rainfalls and total flows, there is no necessity to apply loss functions and base flow separation techniques as in conventional hydrological approaches. However, the rainfall losses have to be considered by using proper parameters.
 - Evaporation data has not been used as input to the model in this study. It is due to the following reason: evaporation has close relation with rainfall, and it is not very important for flood forecasting comparing to rainfall and then it is used as loss in precipitation.

Input variable separation and selection

Having chosen the rainfall and the discharge as system state parameters, the next question is how to process them as efficient input variables to the model, to let model discover the relation between them

Conceptually, there are two types of rainfall-runoff generation. One is the runoff yield due to the accumulative rainfall after the soil moisture content is saturated; the other one is the runoff due to intensive rainfall, which is bigger than the infiltration rate. Therefore, there should be different formula to describe the different runoff yield types.

In this study the daily rainfall, moving average of daily rainfall and discharge of upstream of predicted stations together with their corresponding time lags are used as input variables. It is supposed that the daily rainfall data and the soil moisture content the moving average rainfall can implicitly (the antecedent rainfall) express the rainfall intensity.

The problems here are:

- How many days rainfall should be used to calculate the days moving average rainfall in order to represent the soil moisture state correctly?
- How to express the relative rainfall intensity compared with the soil moisture?
- What is the threshold value to distinguish the two different runoff yield types? In other words, how to obtain the threshold value when the soil moisture is saturated, and what is the actual value?

This can be solved by analyzing each flood event using the classical hydrological method using depletion curve and doing base flow separation, in order to find losses of rainfall and rainfall-runoff response lag time. Then the flood events can be classified in to different classes with each class having its corresponding parameters. The advantage of these approaches is that it gives a deep understanding of the catchments and could be considered as the preferred one. However, obviously this approach is very labor intensive and time consuming, requires field's studies and cannot be followed in this study.

An alternative way of choice of how many days moving average rainfall should be used in this study is according to correlation analysis, and appropriate choice of time lags.

M5 model trees forecasting using full-year data

According to the correlation analysis, the selected inputs are:

The 4 day moving average precipitation including today and the before today (REmov4t, RE mov4t-1 and RE mov4t-2), one day precipitation and the upstream discharge of today and before today of tributaries at Aweramelka (river kessem) and Messeo and upstream main river station Awash station and predicted station discharges are selected.

Table 6.5 The corresponding correlation coefficient of input variables and their corresponding time lags with the predicted discharge

S.No	Variable name	Correlation coefficient	Time lags (days)
1	REt	0.502	1
2	REmov4t	0.604	1
3	REmov4t-1	0.601	2
4	REmov4t-2	0.602	3
5	QAt	0.78	1
6	QAMt	0.47	1
7	QMES	0.49	1
8	QMSt-1	0.93	2
9	QMSt	0.97	1

Predicted result

The different selection of the pruning factor value can determine the different accuracy and complexity of the generated model tree. The same as ANN, there is over fitting problem in the M5 model tree algorithm, which can be solved by cross-validation. The best pruning factor value is that gives the highest accuracy of the testing data set

For the full-year data, the best pruning factor is 4 and there are 19 leaf nodes in the generated M5 model tree, corresponding to 19 linear equations. The root node in the trees and the most important splitting attributes is QMS (MelkaSedi). The following most important attributes is REMov4t, QMES (messeo) that are in the first and second level branch of the trees and appears in the sub branches frequently. The predicted performance showed that the low flow and some peak value are predicted well (Fig 6.3,6.4) the root mean square error in training and testing are 120 and 98m³/s respectively (Table 6.6)

```

M5 pruned regression tree:
(using smoothed predictions)

QMSt <= 136.05 :
|  REmov4t <= 1.35 :
|  |  QMSt <= 52.5 : LM1 (191/5.78%)
|  |  QMSt > 52.5 :
|  |  |  QMEST <= 5.65 :
|  |  |  |  QMSt-2 <= 58.7 : LM2 (52/6.708%)
|  |  |  |  QMSt-2 > 58.7 : LM3 (77/14.509%)
|  |  |  QMEST > 5.65 : LM4 (328/21.369%)
|  REmov4t > 1.35 : LM5 (351/62.378%)
QMSt > 136.05 :
|  QMSt-2 <= 487.9 :
|  |  REmov4t <= 1.95 :
|  |  |  QMSt <= 294.2 : LM6 (50/25.964%)
|  |  |  QMSt > 294.2 : LM7 (18/83.756%)
|  |  REmov4t > 1.95 :
|  |  |  QMSt <= 299.05 : LM8 (104/109.039%)
|  |  |  QMSt > 299.05 : LM9 (144/101.206%)
|  QMSt-2 > 487.9 :
|  |  QMEST <= 68.05 :
|  |  |  REmov4t <= 4.35 :
|  |  |  |  QAMt <= 19.7 : LM10 (8/42.923%)
|  |  |  |  QAMt > 19.7 :
|  |  |  |  |  REmov4t-2 <= 4.45 : LM11 (2/9.794%)
|  |  |  |  |  REmov4t-2 > 4.45 : LM12 (4/25.492%)
|  |  |  REmov4t > 4.35 :
|  |  |  |  QAMt-1 <= 20.15 :
|  |  |  |  |  QAt <= 540 : LM13 (14/44.335%)
|  |  |  |  |  QAt > 540 : LM14 (5/40.393%)
|  |  |  QAMt-1 > 20.15 : LM15 (32/59.718%)

```

Figure 6.2 the model tree generated using full year data

Linear Regression Model

$$QMSt+3=5.2719*REt+17.2823*REmov4t-16.7052*REmov4t-1+7.7514*REmov4t-2-0.4032*QAMt-1+0.3084*QAt+0.6703*QMEST+0.2438*QMSt-1+0.1665*QMSt-2+5.4881$$

Model tree rule

LM1=5.2719*REt+17.2823*REmov4t-16.7052*REmov4t-1+7.7514*REmov4t-2-
0.4032*QAMt-1+0.3084*QAt+0.6703*QMEST+0.2438*QMSSt-1+0.1665*QMSt-2+54.6551

LM2=5.2719*REt+17.2823*REmov4t-16.7052*REmov4t-1+7.7514*REmov4t-2-
0.4032*QAMt-1+0.3084*QAt+0.6703*QMEST+0.2438*QMSSt-1+0.1665*QMSt-2+59.7516

LM3=5.2719*REt+17.2823*REmov4t-16.7052*REmov4t-1+7.7514*REmov4t-2-
0.4032*QAMt-1+0.3084*QAt+0.6703*QMEST+0.2438*QMSSt-1+0.1665*QMSt-2+71.7351

LM4=5.2719*REt+17.2823*REmov4t-16.7052*REmov4t-1+7.7514*REmov4t-2-
0.4032*QAMt-1+0.3084*QAt+0.6703*QMEST+0.2438*QMSSt-1+0.1665*QMSt-2+85.0829

LM5=5.2719*REt+17.2823*REmov4t-16.7052*REmov4t-1+7.7514*REmov4t-2-
0.4032*QAMt-1+0.3084*QAt+0.6703*QMEST+0.2438*QMSSt-1+0.1665*QMSt-2+109.067

LM6=5.2719*REt+17.2823*REmov4t-16.7052*REmov4t-1+7.7514*REmov4t-2-
0.4032*QAMt-1+0.3084*QAt+0.6703*QMEST+0.2438*QMSSt-1+0.1665*QMSt-2+105.5401

LM7=5.2719*REt+17.2823*REmov4t-16.7052*REmov4t-1+7.7514*REmov4t-2-
0.4032*QAMt-1+0.3084*QAt+0.6703*QMEST+0.2438*QMSSt-1+0.1665*QMSt-2+219.2603

LM8=5.2719*REt+17.2823*REmov4t-16.7052*REmov4t-1+7.7514*REmov4t-2-
0.4032*QAMt-1+0.3084*QAt+0.6703*QMEST+0.2438*QMSSt-1+0.1665*QMSt-2+273.1939

LM9=5.2719*REt+17.2823*REmov4t-16.7052*REmov4t-1+7.7514*REmov4t-2-
0.4032*QAMt-1+0.3084*QAt+0.6703*QMEST+0.2438*QMSSt-1+0.1665*QMSt-2+412.6798

LM10=5.2719*REt+17.2823*REmov4t-16.7052*REmov4t-1+7.7514*REmov4t-2-
0.4032*QAMt-1+0.3084*QAt+0.6703*QMEST+0.2438*QMSSt-1+0.1665*QMSt-2+237.9685

LM11=5.2719*REt+17.2823*REmov4t-16.7052*REmov4t-1+7.7514*REmov4t-2-
0.4032*QAMt-1+0.3084*QAt+0.6703*QMEST+0.2438*QMSSt-1+0.1665*QMSt-2+312.3881

LM 12=5.2719*REt+17.2823*REmov4t-16.7052*REmov4t-1+7.7514*REmov4t-2-
0.4032*QAMt-1+0.3084*QAt+0.6703*QMEST+0.2438*QMSSt-1+0.1665*QMSt-2+494.8631

LM13=5.2719*REt+17.2823*REmov4t-16.7052*REmov4t-1+7.7514*REmov4t-2-
0.4032*QAMt-1+0.3084*QAt+0.6703*QMEST+0.2438*QMSSt-1+0.1665*QMSt-2+407.8024

LM14=5.2719*REt+17.2823*REmov4t-16.7052*REmov4t-1+7.7514*REmov4t-2-
0.4032*QAMt-1+0.3084*QAt+0.6703*QMEST+0.2438*QMSSt-1+0.1665*QMSt-2+618.3881

LM15=5.2719*REt+17.2823*REmov4t-16.7052*REmov4t-1+7.7514*REmov4t-2-
0.4032*QAMt-1+0.3084*QAt+0.6703*QMEST+0.2438*QMSSt-1+0.1665*QMSt-2+597.2787

LM16=5.2719*REt+17.2823*REmov4t-16.7052*REmov4t-1+7.7514*REmov4t-2-
0.4032*QAMt-1+0.3084*QAt+0.6703*QMEST+0.2438*QMSSt-1+0.1665*QMSt-2+593.1924

LM17=5.2719*REt+17.2823*REmov4t-16.7052*REmov4t-1+7.7514*REmov4t-2-
0.4032*QAMt-1+0.3084*QAt+0.6703*QMEST+0.2438*QMSSt-1+0.1665*QMSt-2+646.3181

LM 18=5.2719*REt+17.2823*REmov4t-16.7052*REmov4t-1+7.7514*REmov4t-2-
0.4032*QAMt-1+0.3084*QAt+0.6703*QMEST+0.2438*QMSSt-1+0.1665*QMSt-
2+2033.8551

LM19=5.2719*REt+17.2823*REmov4t-16.7052*REmov4t-1+7.7514*REmov4t-2-
0.4032*QAMt-1+0.3084*QAt+0.6703*QMEST+0.2438*QMSSt-1+0.1665*QMSt-2+791.8247

Table 6.6 the predicted result test of model trees using full year data input

Performance criteria	Training	Testing
Number of rules	19	19
Correlation coefficient	0.8463	0.88
Mean absolute error	60.55	50.9
Root mean square	120.1	97.7
Relative absolute error	38.25	33.9
Root relative square error	53.69	46
Total number of instance	1457	362

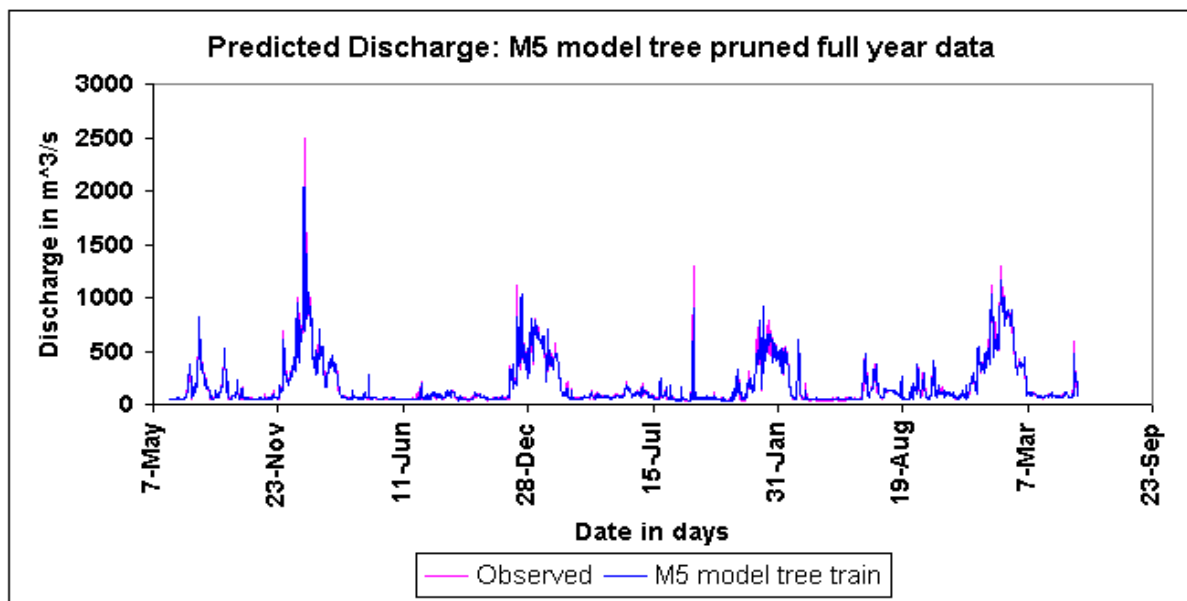


Figure 6.3 observed versus predicted in training using four (1990-1993) year data

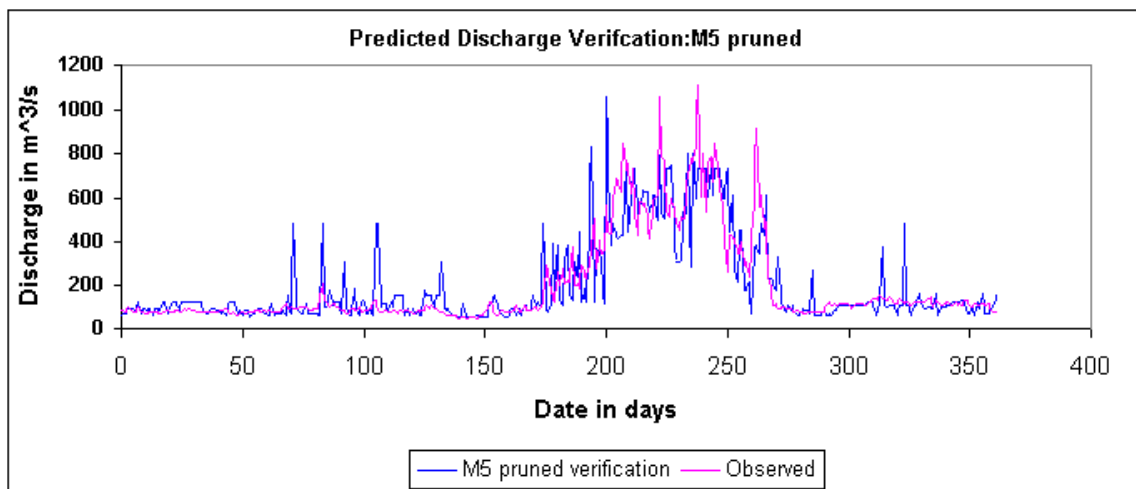


Figure 6.4 Observed versus predicted in verification data(1994)

Table 6.7 The predicted result test of model trees using wet season data input

Performance criteria	Training	Testing
Number of rules	25	25
Correlation coefficient	0.85	0.848
Mean absolute error	84.6	99.11
Root mean square	149.13	160
Relative absolute error	38.04	40.78
Root relative square error	53.6	52.3

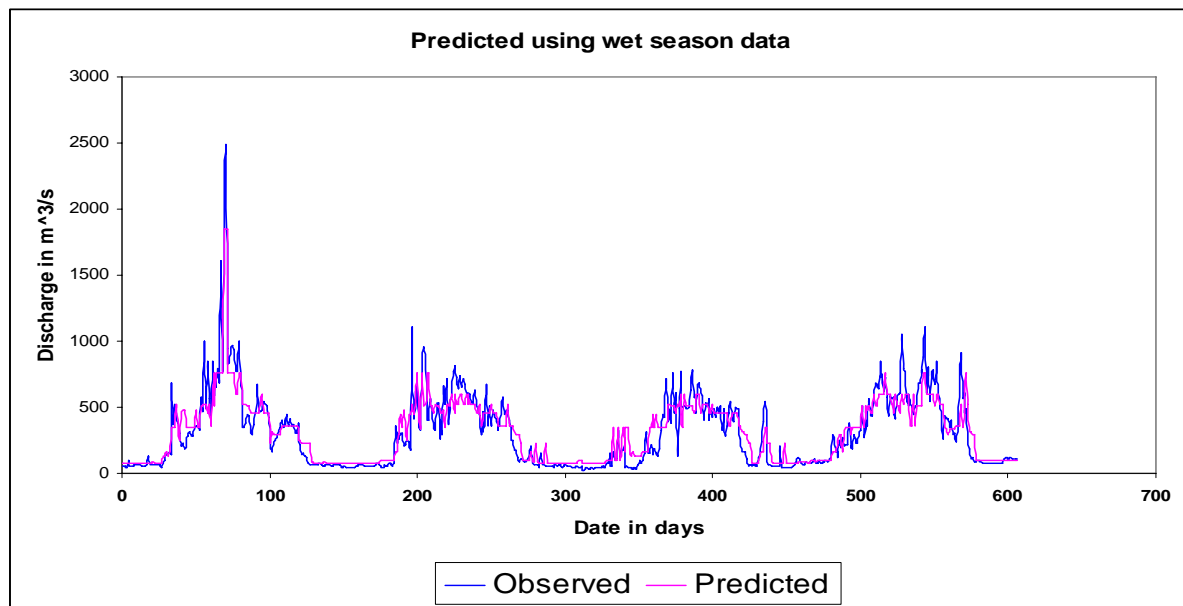


Figure 6.5 Observed versus predicted in training using wet season data (1990-1993)

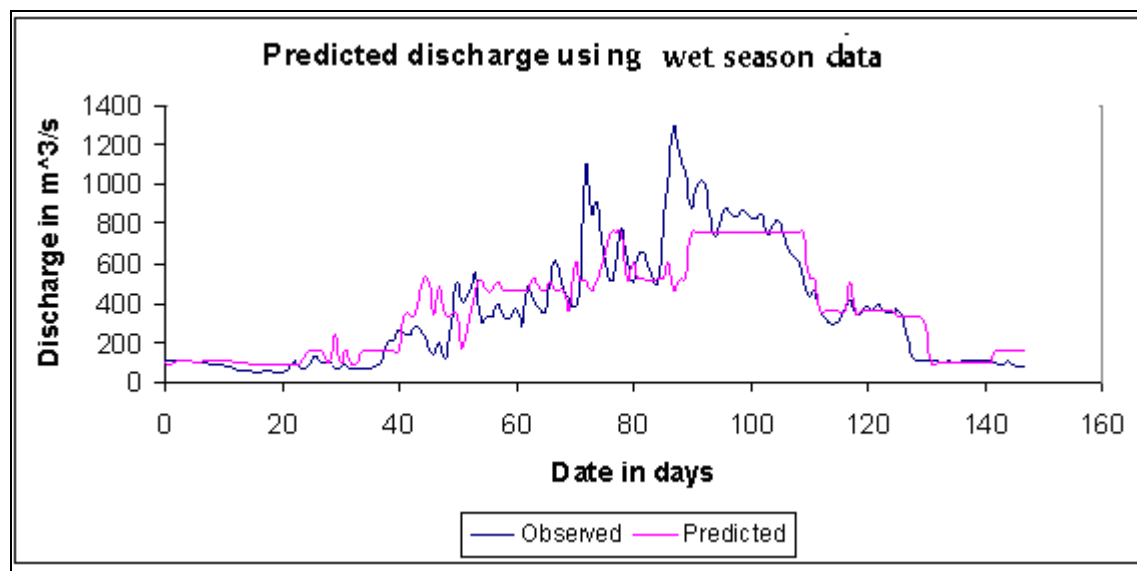


Figure 6.6 Observed versus predicted in test data using wet season data (1994)

Discussion:

Using the full year data as input the flood prediction is good in training but in verification in addition to amplitude change there is also phase shift between peaks. Peaks are not predicted well with high classification error. Based on this result, it was tried to use wet season data as input in order to improve the model performance but the result shows no improvement in peak prediction rather the shift of phase increases in verification (Figure 6.5, 6.6, table 6.7). In general, an overall reasonable prediction in training data set and in verification data set there is a shift of peak flows in addition to lowering its value. The error may be reduced using some additional inputs like in some experiments additional inputs were used to help model to automatic distinguish the special data set

- Add discharge changes as input. Because large dQ can represent rapid increase of discharge due to intensive rainfall, thus large dQ can represent intensive rainfall and flood upstream: positive and negative dQ value can tell the raising limb and recession part discharge of hydrograph
- Add the difference of rainfall and moving average rainfall as an indicator of relative rainfall intensity.

Conclusion: With the above result, it can be said that M5 model tree in this data set predicts with considerable error for warning system most of the generated rules has higher classification error as well. Therefore to increase the model performance further data classification of the data is needed.

6.5 Flood forecasting using ANN

Inputs and the ANN topology

Artificial neural network was used to attempt to replicate the same prediction. The same input and output flood season data used in the M5 model tree (13 inputs and 1 output) were used to the ANN.

For comparison, 5 layer MLPs ANN model architectures used. Hyperbolic tangent and linear activation function were used in the hidden layer and output layer respectively (Figure 6.7).

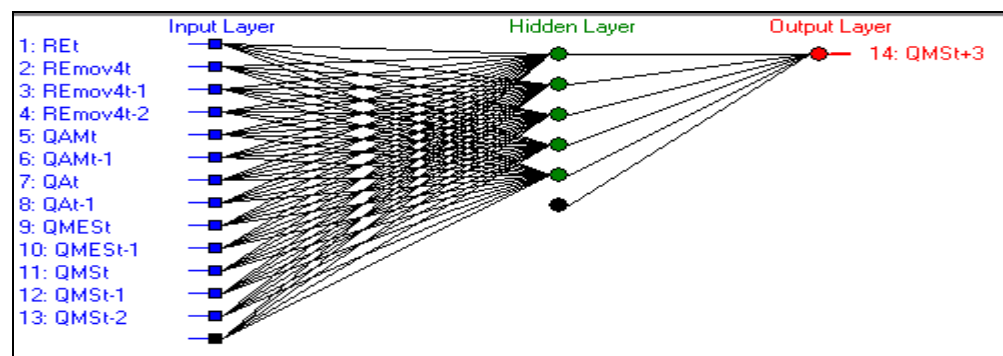


Figure 6.7 The model architecture

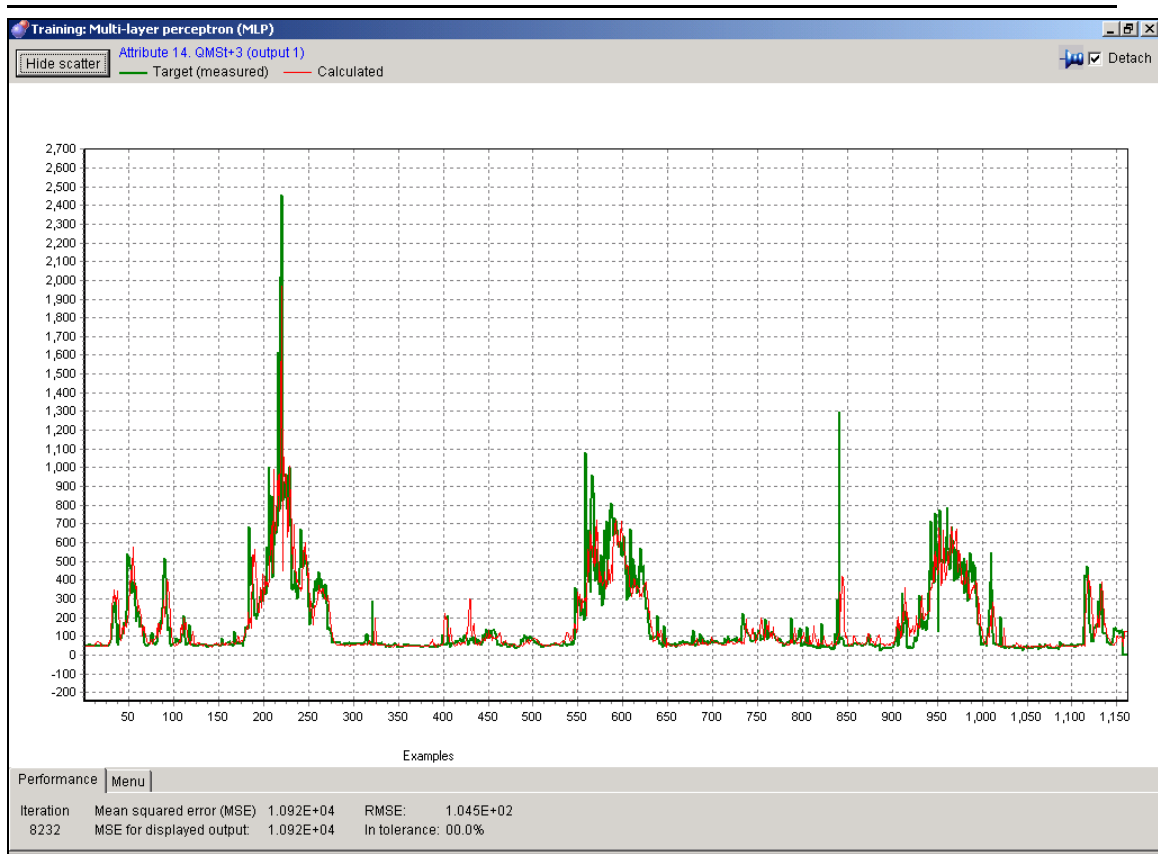


Figure 6.8 Training prediction out put

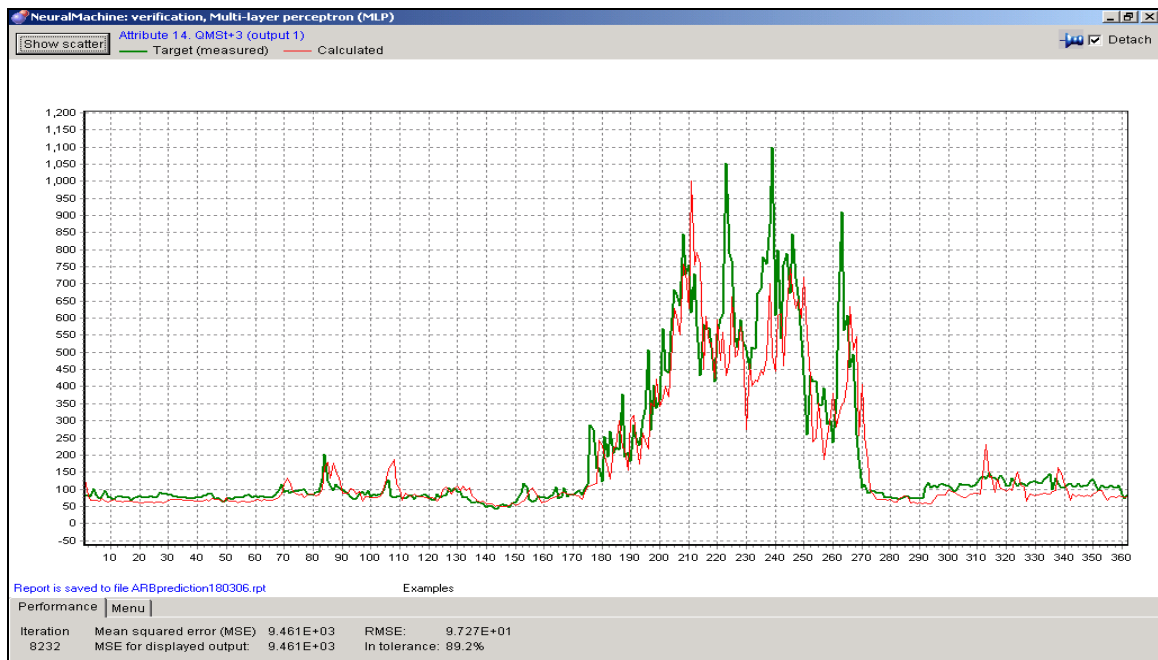
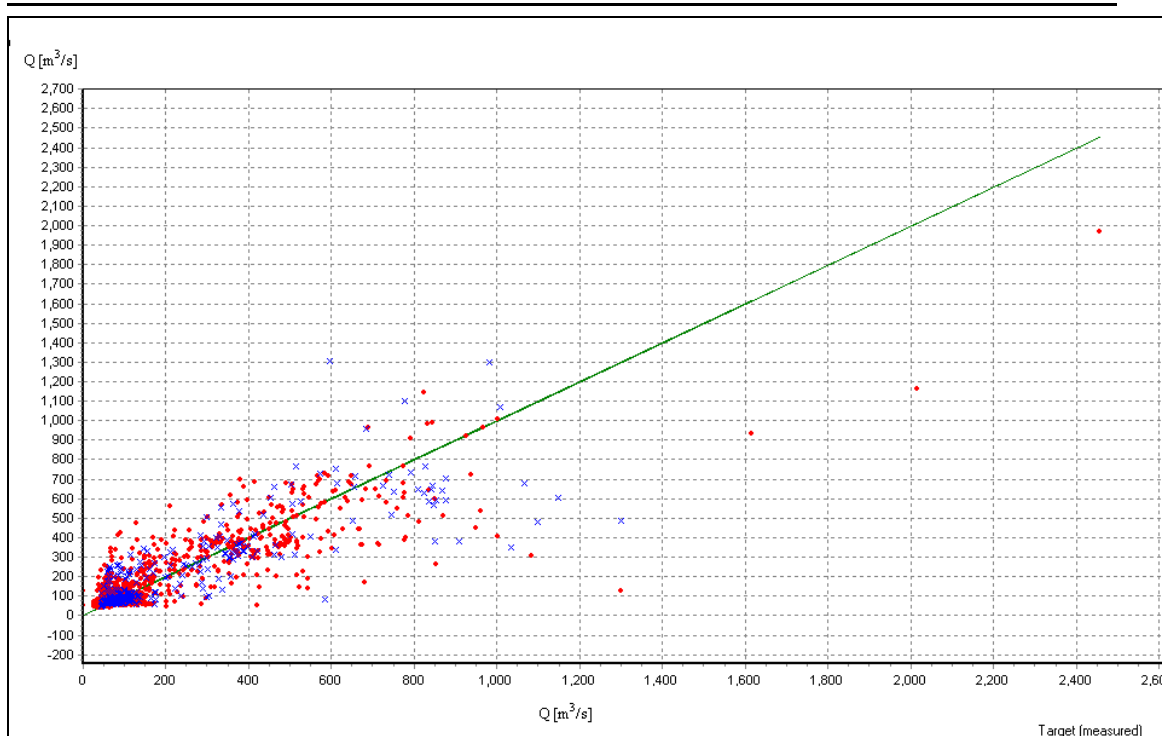


Figure 6.9 Verification prediction out put



- Train data
- X cross validation

Figure 6.10 Scatter plot of observed and predicted

Discussion:

As shown in above figures (Figure 6.8,6.9,6.10),the ANN prediction overall result is better as compared to M5 model tree. Comparing the hydrograph of M5 model tree and ANN, it can be seen that the maximum error occurs at the same flood events. It is known that ANN are capable of identifying usable relationship between discharge and antecedent rainfalls.M5 mode tree are easier for practioner to understand

Conclusion: From the result analysis, it can be concluded that the error is not because of the sample size of the data, but because of the attributes selection.Therofere it supports the idea by adding, some new attributes it is possible to increase prediction accuracy

6.6 Error analysis

Error due to the input variables

As mentioned above, independent of what model used-M5 model tree or ANN, for some flood events the predicted results are rather poor. The flood events with maximum prediction error occurred at 1990and 1993 for training data and 1994 for testing data, for training and testing time series respectively. As shown in table 6.5.Figure 6.4 and Figure 6.5, after analyzing the relation between the flood event and its associated antecedent rainfall and upstream discharge, the error can be classified in to 3 basic types, although some of them can happen at the same time:

-
- 1) **Error 1** predicted discharge is much higher than the observed: this happens due to the model cannot distinguish the flood due to a very intensive and short duration rainfall with long duration medium intensive rainfall. This type of error is caused by medium intensive and long duration rainfall; the antecedent rainfall has much influence on the runoff generation. If intensive rainfall occurred after a long period of dry series, although there may be 1 or 2 rainy days during the long dry series, then the predicted runoff is much higher than observed. On the other hand, if the intensive rainfall follows a long duration of rainfall, although there may be 1 or 2 dry day during the long rain event. In this case the soil is more or less saturated which means there is not much rainfall losses, the flood discharge is higher, but the predicted discharge will be much lower than observed

 - 2) **Error 2** phase shift of predicted discharge accompanying with amplitude error, so that the predicted peak is lower than the observed. This type error usually happens at floods caused by a long duration of rainfall, although there may be 1 or 2 dry day during the long rain events, the soil moisture content is nearly saturated. However, model cannot distinguish them from error 1.

 - 3) **Error 3** the predicted flood discharge is much lower than the observed, due to flash flood happened at tributaries, whose discharge is not included in the input. The error 3 is due to the heterogeneous distribution of rainfall-runoff yield characteristics among different tributaries and sub catchment, which cannot be represented by the parameter of area average rainfall

7. Conclusions and Recommendations

This study was carried out to identify the causes of flooding in Upper and Middle Awash, and to discuss the advance release from Koka reservoir to reduce flooding in downstream of the reservoir in the basin from a joint operational perspective with the downstream tributaries. The aim was to reduce flooding downstream of Koka reservoir in order to protect irrigation farms, which has a direct impact in poverty alleviation in the country. In addition, the forecasting of floods to give warning for the people to reduce disaster is evaluated.

7.1 Conclusions

Findings from the present study are:

Flood analysis

Flooding in Awash basin is the combined effect of high rainfall in the highlands of the catchment through tributaries and high release discharge from Koka reservoir during the wet season especially in August that are unplanned or made without calculating the risk to the downstream flood prone area.

Reservoir operation

Advance releases from Koka reservoir can reduce the peak discharge downstream by 280/300m³/s. Koka reservoir with the present capacity and inflow volume from the upstream catchment is capable of satisfying the downstream demand without any risk in power generation by changing the existing operation rule of the reservoir together with forecasting the inflow

Flood forecasting

Flood forecasting model that includes the discharges of the tributaries of the Awash river can reduce damages and loss of life. ANN model performs better than Model tree. MLP models with 3-day lead-time simulates single events well, but forecasting of several consecutive events needs further improvements and need incorporating process knowledge in it.

Results confirmed that the benefit from joint operation of reservoir and both structural and non-structural measures are solutions in reducing flooding both in cost and environmental considerations in the existing System.

7.2 Recommendations

The following recommendations, which are focused on the problems inherent in the existing operation rule, can be suggested

1. As the advance release discharge increases (by modifying the existing operation more) the downstream flooding also decreases, however, operational conflicts will increase most particularly during dry hydrological sequences and in some years a flood shift from August to some other month will occur. If the frequency of water-use conflicts continue to increase, the use of a real-time operational model could be considered. To adopt such a system requires concurrent commitment in the area of institutional and data management support
2. The study showed that applying structural measures does not lead to an optimal solution to reduce flooding unless it accompanied by non-structural measures. Therefore, development of a forecasting model for inflow to the system is recommended to enhance its applicability in real-time operation
3. Research is required into the operational use of reservoir and flood plain storage within the flood context with the aim of reducing downstream flood peaks, given a (medium-range) forecast. This requires integration of reservoir operation rules into forecasting models, optimization of reservoir operation and probabilistic forecasting to establish robustness of operation over a range of possible spatial and temporal precipitation and hydrological response scenarios
4. Short term and long term strategies of operating the reservoir are recommended

Short-term strategy: like establishing model or set of model to allow operation policies to be routinely examined

Long-term strategy:

- Extend the use of operational planning model to examine the benefit of inflow forecasting
- Extending the use of operational planning model to assist in preparing a seasonal operating policy which considers start of season reservoir levels and upcoming energy requirement

5. To obtain a higher level of managing floods with respect to all development demands there is a need to make more storage capacity in one of the tributaries for multipurpose objectives. Water availability permits the expansion of the irrigation system as well as power generation to the full development case with adequate reliability.

8. References

- ADF (2003) proposal: Flood control and watershed management study-constructing dikes along a river in Awash (<http://www.afdb.org/pls/portal/docs>)[Accessed 6 August, 2005]
- Achamyeleh.K: Impact of climate change on water resources of Awash basin, Ethiopia (<http://www.int-res.com/articles/cr/12/c012p091.pdf>)[Accessed September 2005]
- De Laat.P.J.M (2001). Workshop on hydrology, Lecture notes, IHE, Delft, the Netherlands
- Ethiopia: Floods [DHA-04: 04-Sep-96] (<http://cidi.org/disaster/96b/0014.html>) [Accessed October, 2005]
- Girma Taddese, Kai Sonder & Don Peden (1999): The water of Awash River basin a future challenge to Ethiopia (<http://www.iwmi.cgiar.org/Assessment/files/pdf/publications/WorkingPapers/WaterofAwashBasin.pdf>)[Accessed September, 2005]
- Maskey.S. (2004), modeling uncertainty in flood forecasting, PhD thesis, Unesco-IHE, the Netherlands
- MoA, (2000): Report on sustainable development strategies
- MoWR,(1998):Water resource strategies and policy report
- Reader (2005): MDG summary task force reports, internal Unesco-IHE use only
- Price, R.K, (2000): Hydroinformatics lecture notes, IHE delft, the Netherlands
- Solomatine, D.P (2005), Hydroinformatics lecture notes, IHE delft, the Netherlands
- UNDP-EUE (1999):AfarRegionAwashRiverFloods09/99 (<http://www.africa.upenn.edu/Hornet/afar0999.html>,) [Accessed October 2005]
- WMO-NO.168 (fifth edition 1994): Guide to hydrological practice data acquisition and processing, analysis, forecasting and other applications

Supportive References

- Beven, K.J. (2001) Rainfall-Runoff modeling, the primer, johnwiley&sons, LTD, UK

-
- Chi,P.,2001,Regional flood frequency analysis for the Huai River using artificial neural networks,MSc thesis,IHE,Delft,The Netherlands
- Coulibaly P.,Anctif F.and Bobee B (2000) Daily reservoir inflow forecasting using Artificial networks with stopped training approach, journal of hydrology 230.
- Hall M.J, (2001).How well does your model fit the data? Journal of hydroinformatics, vol.3:49-55
- Linsely.R.K, Kohler.M.A, PaulhusJ.L.H, (1998).hydrology for Engineer, McGraw-Hill, Inc.UK
- Lobbrecht A.H, Y.B Dibeke and D.P.Solomatin, application of neural networks and Fuzzy logic to integrated water management project report, October 2002, Delft
- Marker.M.S and Clark.S.Q, Evaluation of Hydraulic and Hydrologic models for real time flood forecasting use in the Yangtze River catchment, proceedee of 8th national conference on hydraulics in water engineering, Australia
- Reservoir Operation (<http://www.worldbank.org/html/fpd/em/hydro/romdfr.stm>) [Accessed January, 2006]
- Reservoir release forecast model for flood operation of the Folsom project including pre-releases(<http://www.engineering.usu.edu/uwrl/www/faculty/DSB/USSD2004.pdf>) [Accessed January 2006]
- Rodda.J.C, (2000): Hydrological forecasting, dealing with natural disasters. (<http://www.royalsoc.ac.uk/policy/IDNDR/rodde.pdf>)[Accessed November, 2006]
- Water resources management for three series reservoir operation on citarum river basin west java Indonesia (http://www.wrrc.dpri.kyotou.ac.jp/~aphw/APHW2004/proceedings/OHS/56-OHS-A176/56-OHS-A176_Manuscript.pdf)[Accessed January, 2006]
- Jingyi.Z, Hall. M.J., (2002): Regional analysis of flood: a case study from southeast china, Eighth National Hydrology Symposium, British Hydrology society
- Zagona.E, Research Associate, Center for Advanced Decision Support for Water and Environmental Systems, University of Colorado, Boulder, CO; Terrance J. Fulp, Operations Research Analyst, U.S. Bureau of Reclamation, Lower Colorado Region, a general river and reservoir modeling environment (<http://cadswes.colorado.edu/PDF/RiverWare/vegas98.pdf>)[Accessed January,2006]
-