Development of Flood Forecasting Model in Middle Awash River Basin of Ethiopia

Zewde Zinash Mekonnen [*] **Supervisor: Prof. A. W. Jayawardena**[**]
MEE09204

ABSTRACT

Flooding is a common problem in Ethiopia. Frequency and magnitude have increased rapidly in the last few decades. The increase in flooding is a result of climate change as well as land-use change (particularly deforestation). The main impacts of flooding are loss of human lives and properties, destruction of roads and electric systems that result great economical loss for the country. Another notable consequence of flooding is crop destruction and subsequent malnutrition. Large-scale irrigated agriculture of the country is concentrated along the Awash River. Irrigation development in this river basin is quite advanced and is located in the flood plains on either side of the Awash River. Therefore, high economic damage occurs during flooding along this river. It is estimated that in the Awash Valley, almost all of the area (200,000 - 250,000 ha) delineated for irrigation development is subjected to flood during high flows of the river (Kefeyalew, 2003). Currently, the authorities responsible for river basin management identify the effectiveness of an early flood warning system for the middle Awash River basin. The main objective of this study is to develop flood forecasting model to give an early warning for the people who are likely to be affected in Middle Awash River basin. The specific objective of this study is to examine the suitable model for the study area. The selected models to be tested for this study are; i) stochastic type (ARMA) model ii) Statistical method and iii) Deterministic type model i.e. Local Approximation approach. For this a detailed time series analysis for hydrological and meteorological data is made before model application. Flood frequency analysis is performed and floods of different return periods are identified. In this study one step ahead prediction was made using a nonlinear Local Approximation method, which is deterministic type model, and applied to some daily stream flow data. For comparison, stochastic type (ARMA) model and statistical method of forecasting were implemented. Forecasting accuracy of all the proposed methods was tested by different statistical indices. The results show that flood forecast in Middle Awash River is more convenient to be modeled by using Local Approximation approach than stochastic type (ARMA) model approach and statistical method.

Keywords: *Flood forecasting model, Stochastic (ARMA) model, statistical method, deterministic model, Local Approximation approach, Flood frequency analysis.*

INTRODUCTION

Flood forecasting and early warning system is an important tool to give appropriate reliable information of the incoming flood to the vulnerable community. In Ethiopia, flood forecasting and early warning system is not well advanced. Presently the Government established flood control management to minimize flood risk in Awash River basin. Nevertheless, there still are problems to increase flood damage in this area. The study area covers the total catchment of 31183 km^2 and located in middle part of Awash River. In this stretch of the Awash River, flooding is common during the rainy season. The area under Middle Awash River basin (Figure 1) is well known for irrigation activity. In the Middle Valley, the Awash River is joined by its major tributary river Kessem. The river Kessem (a total catchment area of 3113km^2) drains huge amount of water from the Western highland of the drainage system. So, this joining with the Awash River poses flood risks to the crop plantations in lower plain area.

[*] Head of Addis Ababa Regional Office under Hydrology Department, Ministry of Water Resources, Ethiopia
[**] Research and Training Advisor, International Centre for Water Hazard and Risk Management (ICHARM), Japan

In a recent rainy season flooding event, the whole 10,000 ha farm of Amibara project has been inundated together with standing crops in the farm land. Figure 1.shows location of Middle Awash River basin along with discharge measuring station.

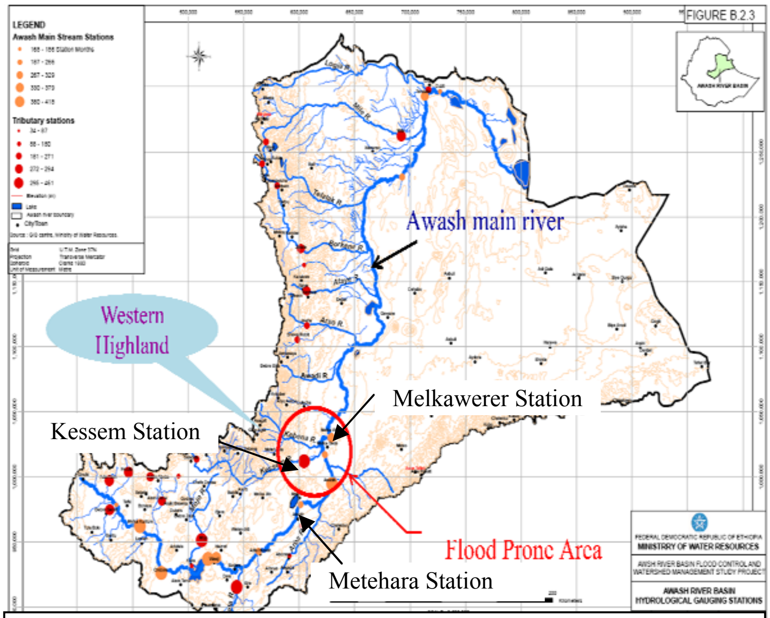

Figure 1. Location of Middle Awash River basin flood prone area along with discharge measuring stations

DATA

The hydrological department under the Ministry of Water Resources in Ethiopia is responsible for hydrological data (flow data) collection, processing, analysis and dissemination to users. The meteorological data (rainfall data) collected and analyzed by National Meteorology Service Agency of Ethiopia. The collected flow and rainfall data at different gauging stations in Awash River basin are utilized for calibration (training) and testing application for the proposed models by splitting into two parts. Table 1 summarizes the data that are used in this study.

Table 1: Summary of used data in model calibration (training) and testing

Data Source	Type of Data	Total Data Range	Station Name	Calibration		Testing	
				From	To	From	To
Hydrology department	Maximum monthly flow (m^3/s)	1969-2003	Kessem (Q_{29})	1969	1998	1999	2003
		1985-2002	Melkawerer (Q_p)	1985	1998	1999	2002
		1985-2002	Metehara (Q_{31})	1985	1998	1999	2002
National Meteorological Service Agency	Monthly rainfall (mm)	1985-2002	Awaramelka	1985	1998	1999	2002
		1985-2002	Melkakuntur	1985	1998	1999	2002
		1985-2002	Metehara	1985	1998	1999	2002

THEORY AND METHODOLOGY

The adopted methodology consisted of the following analysis:
Historical data i.e. rainfall and flow data of the selected stations in the study area are used for this research. Hydrological data analysis i.e. missing data, trend and periodicity of the data series are checked before model application. The data series trend is detected by using:

1) Regression test for linear trend and the significance of trend is tested by using Student's ***t*** test. The result of the trend analysis which is tested by Student's ***t*** test is verified by MINITAB software.
2) Mann-Kendall test for trend and the significance of trend is tested by using the test statistic *Z* at α level of significance, the tested significance level of α is 5% or 0.05 (Yasemin Ezber et al, 2007).

Periodicity of the data series is identified by using Autocorrelation analysis and the detected periodicity pattern of the data series is removed by using the equation $\frac{(y-\bar{y})}{s_t}$ (Maidment, 1992) before model application. Frequency analysis is performed based on McCuen (1941) in order to determine flood magnitude of different return period. The fitness of the data series with the frequency distribution methods i.e. Normal, Log-Normal and Gumbel distribution methods are tested by using

Chi-Square goodness of fit test technique. In this study flood forecasting model developed by using the selected three type of approaches i.e. stochastic type (ARMA) model ,statistical method and deterministic type model (Local Approximation) approaches. The standard statistical measures i.e. Pearson correlation coefficient (CC), Relative Root Mean Square Error (RRMSE), Mean Absolute Error (MAE) and coefficient of determination (R^2) was used to evaluate the accuracy of prediction for all methods. The adopted flood forecasting model for the study area described as follow:

- **Deterministic model using Local Approximation approach**

According to Jayawardena & Lai Feizhou,1994, forecasting can be done by considering the time series in a piecewise manner. The dynamics of a time series $X_1, X_2, X_3, ..., X_n$ are fully captured or embedded in the *m*-dimensional phase space. For the purpose of describing the dynamics of the system that generates the time series, a vector Y_t of the time series $X_1, X_2, ..., X_n$ can be defined as:

$Y_t : \{X_t, X_{t-\tau}, X_{t-2\tau}, ..., X_{t-(m-1)\tau}\}$ Where, 'τ' is a delay time and '*m*' is the dimension of the vector.

The basic idea in the local approximation method is to break up the domain in to local neighborhoods and fit parameters in each neighborhood separately. For example, $Y_t : \{X_t, X_{t-\tau}, X_{t-2\tau}, ..., X_{t-(m-1)\tau}\}$ has a neighbor:

$Y_{t-1} : \{X_{t-1}, X_{t-1-\tau}, X_{t-1-2\tau}, ..., X_{t-1-(m-1)\tau}\}$,
$Y_{t-2} : \{X_{t-2}, X_{t-2-\tau}, X_{t-2-2\tau}, ..., X_{t-2-(m-1)\tau}\}$,
$Y_{t-3} : \{X_{t-3}, X_{t-3-\tau}, X_{t-3-2\tau}, ..., X_{t-3-(m-1)\tau}\}$,
....etc. where, Y_t is an m-dimensional vector and X_t is a scalar.

Forecasting is done on the basis of the behavior of the series in the neighborhood of the vector Y_t which contains the current value X_t. The neighborhood is defined as a domain which will contain the vector Y_t in the interior, and a chosen number of other vectors $Y_{t'}$ with ***t'*< *t*** from all possible values of ***t'*** in the periphery. It can be thought of as an m-dimensional space, in which Y_t is an m-dimensional point in the interior, and $Y_{t'}$ are similar peripheral points.

The vectors $Y_{t'}$ are chosen so as to give the closest neighborhood. Calculating the distance between two points i.e. vector Y_t and neighborhood's vectors $Y_{t'}$ is the criterion to define the closeness of the neighborhoods. The smallest distance means the closest neighbors. For distance calculation between two vector points; the Euclidean norm was adopted in this study. The Euclidian distance between the two vectors is determined by:

If $m = 2$, $\|Y_t - Y_{t'}\| = \sqrt{(X_t - X_{t'})^2 + (X_{t-\tau} + X_{t'-\tau})^2}$

If $m = 3$, $\|Y_t - Y_{t'}\| = \sqrt{(X_t - X_{t'})^2 + (X_{t-\tau} + X_{t'-\tau})^2 + (X_{t-2\tau} + X_{t'-2\tau})^2}$
…etc., for all ***t'* < *t*** (1)

The number of the chosen vectors ($Y_{t'}$), *K*, are determined by trial and error through experimentation.

After time *l*, the nearest points $Y_{t'}$ would have moved to new positions. The best forecast of X_{t+l} would then be the weighted average of the values of $X_{t'+l}$ with the weighting being related in some form to the distance of $Y_{t'}$ from Y_t. If equal weights are used for the *K* neighbors, then the forecast can be given by:

$$X_{t+l} = \left(\frac{1}{K}\right) * \sum_{i=1}^{K} X_{t'+l} \quad (2)$$

where, X_{t+l} is forecast value at leadtime l, l =1 day in this study, $X_{t'+l}$ is the selected neighbours flow values at lead time l, i = 1,...,K *and* K is the number of neighbours considered in the model. This study considers equal weights for the K neighbors. Therefore, the above mentioned equation (2) was applied to forecast the incoming flood in the study area. In this approach forecasts were made for the selected station (Kessem station), with a record length of 12000 data points for calibration (training) and 80 data points for comparison, based on the dynamic system of the flow data and the past history.

RESULTS AND DISCUSSION

Trend analysis:

The results of Regression test for linear trend (Table 2) verified by MINITAB software and Mann-Kendall trend test (Table 3) shows that there were no significant trends for the selected stations in the study area.

Table 2 Regression test result for linear trend

Station Name	Type of Data	from	To	Calculated Student's *t* value	Student's *t* value from table
Kessem	Flow	1969	2003	-0.21793	1.976227
Metehara	Flow	1985	2002	0.0920296	1.983608889
Melkawerer	Flow	1985	2002	0.07952062	1.983608889
Awaramelka	Rainfall	1985	2002	0.04775697	1.983608889
Melkakuntur	Rainfall	1985	2002	-0.442052	1.983608889
Metehara	Rainfall	1985	2002	-0.75379486	1.983608889

Table 3 Mann-Kendall trend analysis test result

Station Name	Type of data	From	To	Mann-Kendall test result	Remark
Kessem	Flow	1969	2003	S = -2906 Z = -1.010709 Variance = 8261142	Negative trend Trend is not significant
Metehara	Flow	1985	2002	S= 1248 Z= 1.174414 Variance = 27433	Positive trend Trend is not significant
Melkawerer	Flow	1985	2002	S= 504 Z= 0.473 Variance = 27401	Positive trend Trend is not significant
Awaramelka	Rainfall	1985	2002	S= 113 Z= 0.1060056 Variance = 1116293	Positive trend Trend is not significant
Melkakuntur	Rainfall	1985	2002	S= -250 Z= -0.2358322 Variance = 1114789	Negative trend Trend is not significant
Metehara	Rainfall	1985	2002	S= -912 Z= -0.8607882 Variance = 1120067	Negative trend Trend is not significant

Flood frequency analysis:

Fitness of the annual maximum flow of Kessem station data with Gumbel and Normal distribution were performed in order to calculate the extreme flood magnitude with its recurrence period. Based on the Chi-square goodness of fit test results Normal distribution (Chi-value = 1.33) is best fit for the data set than Gumbel distribution (Chi-value = 2.10).

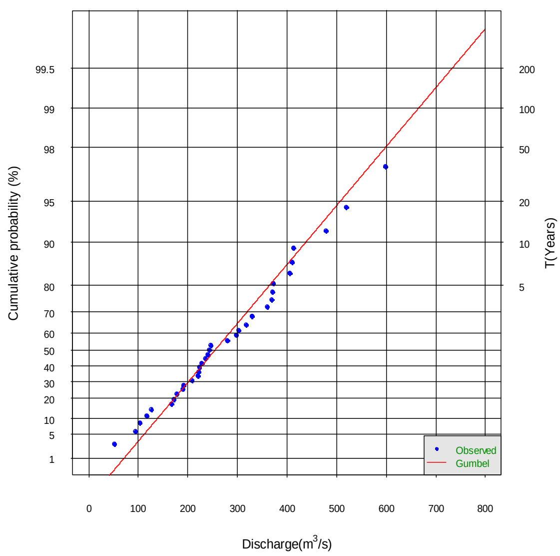

Figure 2 Gumbel distribution plot of Kessem station

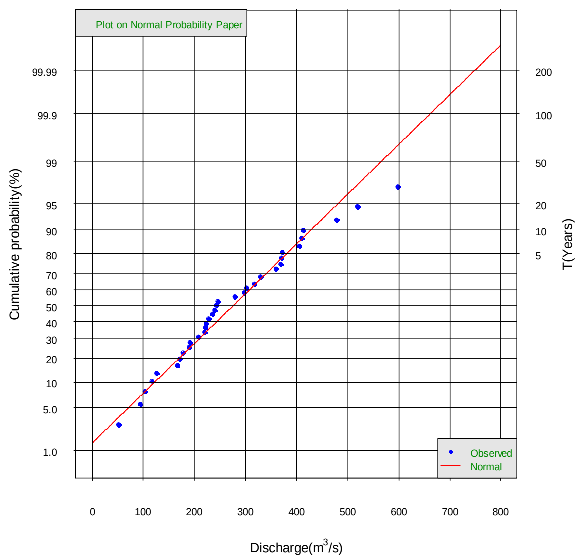

Figure 3 Normal distribution plot of Kessem station

Table 4 Summary table for determined parameters, statistical indices and developed model for the selected forecasting methods:

Description	Types of Models that are examined in this study		
	Stochastic Type Model, AR(1) Model	**Statistical method**	**Deterministic Type Model (Local Approximation Approach)**
Determined Parameters	$\phi_{1,1} = 0.2997$	X = 0.28006	Embedding Dimension $'m' = 22$
		Y = 0.172723	Number of neighbors considered in the model $'K' = 3000$
		Z = 0.565947	
Developed Model	$\hat{Z}_t(l) = 0.2997 * Z_t(l-1)$	$Q_p(t) = 0.28006 * Q_{29}(t-1) + 0.172723 * Q_{31}(t-1) + 0.565947 * RF_{avg}.(t)$	$X_{t+1} = (\frac{1}{K}) * \sum_{i=1}^{K} X_{t'+1}$
Statistical indices			
Calibration	CC=0.827	CC=0.747	
	R^2=0.6835	R^2=0.5576	
	MAE=0.033914	MAE=13.0557	
	RRMSE=1.006851	RRMSE=0.590434984	
Testing	CC=0.839	CC=0.7676	CC=0.979
	R^2=0.6937	R^2=0.5889	R^2=0.9582
	MAE=1.24572712	MAE=23.07417	MAE=0.049001713
	RRMSE=1.12656291	RRMSE=0.690053199	RRMSE=0.21140138

The time series and scatter plot of stochastic type model for calibration and testing are shown below:

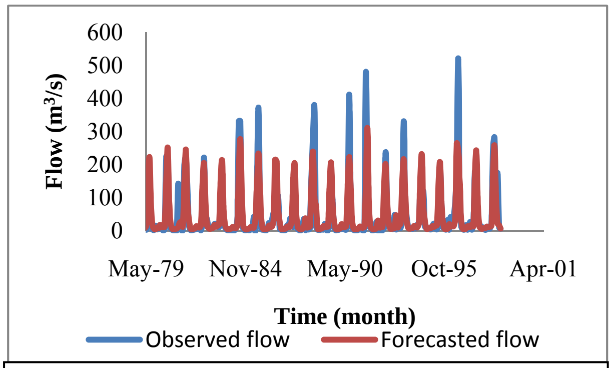

Figure 4 Time series plotting of observed and forecasted flow for calibration (1969-1998).

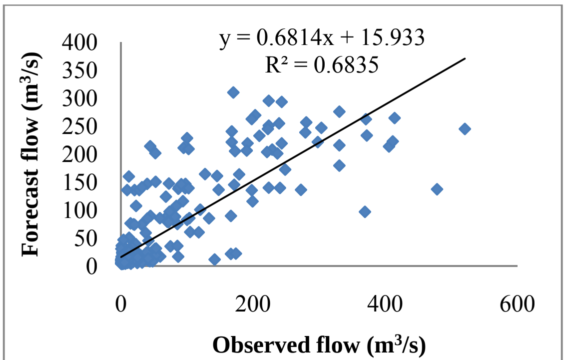

Figure 5 Scatter plotting of observed and forecasted flow for calibration (1969-1998).

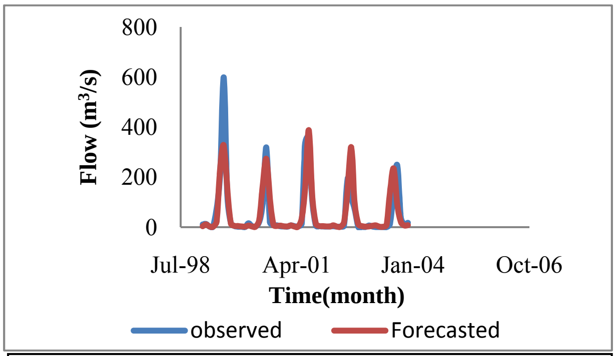

Figure 6 Time series plotting of observed and forecasted flow for testing (1999-2003).

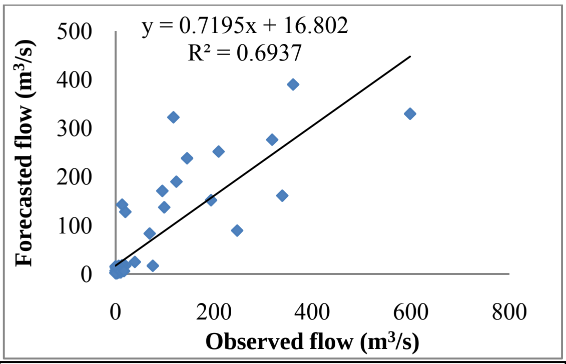

Figure 7 Scatter plotting of observed and forecasted flow for testing (1999-2003).

The scatter plot of the developed model used by statistical method for calibration and testing are shown below:

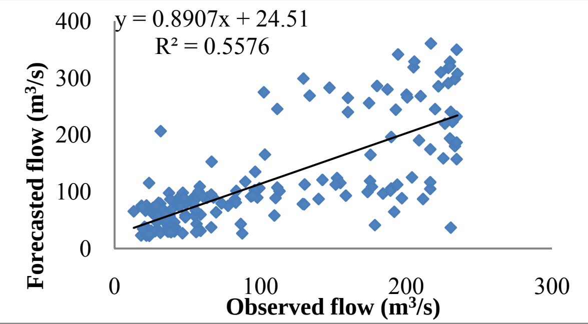

Figure 8 Observed vs. forecasted flow plot of Melkawerer station for calibration (1985-1998).

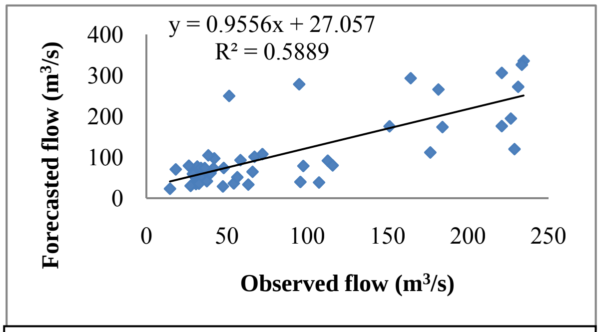

Figure 9 Observed vs. predicted flow plot of Melkawerer station for testing (1985-1998).

The time series and scatter plot of the developed deterministic model using Local Approximation approach for testing are shown below:

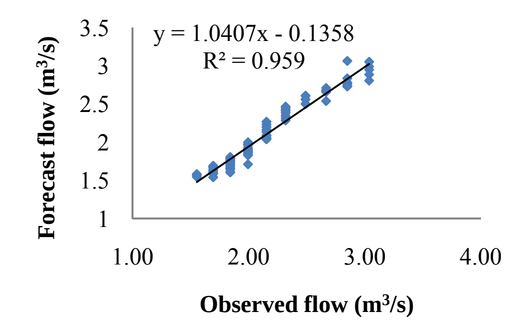

Figure 10 Observed flow vs. forecasted flow plot for Kessem station flow data

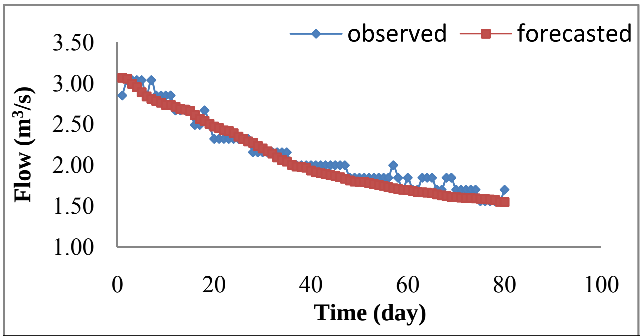

Figure 11 Observed vs. Forecasted flow time series plot of Kessem station data.

CONCLUSIONS

Forecast accuracy of all forecasting model approaches tested by statistical indices. There is a convincing statistical evidence to believe that flood forecasting in the study area is better modeled by Local Approximation method (equation 2) than ARMA model and statistical method.

RECOMMENDATION

As far as this method is still in the early stage of development it is difficult to get references with detail information except limited number of journal papers. So, the determination of embedding dimension '*m*' and number of neighbor value was done by using iteration for calculation. Therefore, Future study for embedding dimension '*m*' determination method should be conducted to improve the accuracy of the forecast value.

ACKNOWLEDGEMENT

I would like to express my gratitude to my supervisor Prof. Dr. A. W. Jayawardena, Research and Training Advisor, ICHARM, PWRI, Tsukuba, Ibaraki, Japan for his overall supervision, invaluable suggestions and comment in every aspect of this work.

REFERENCES

Jayawardena, A. W. & Lai Feizhou (1994) Analysis and Prediction of Chaos in Rainfall and Stream Flow Time Series. J. Hydrol.153. 23-52.

Kefeyalew, A. (2003) Integrated flood management in Ethiopia. http://www.apfm.info/pdf/case_studies/Ethiopia.pdf.

Maidment, D. R. (1992) Handbook of Hydrology, McGRAW-HILL, INC., United States of America.

McCuen, R. H. (1941) Hydlologic Analysis and Design, Pearson Prentice Hall, United States of America, 3rd edition.

Yasemin Ezber, Omer Lutfi Sen, Tayfun Kindap and Mehmet Karaca, (2007) Climatic effects of urbanization in Istanbul: a statistical and modeling analysis, International Journal of Climatology, 27: 667–679