ADDIS ABABA UNIVERSITY

SCHOOL OF GRADUATE STUDIES

Stochastic Simulation of Streamflow and Hydrologic Drought Analysis (Case Study: Upper Blue Nile Basin).

A thesis submitted in partial fulfillment of the requirements for the degree of Masters of Science degree in Civil Engineering.

By

Yohannes Hagos Subagadis

Dr. Semu Ayalew Moges/Thesis Advisor

June, 2009

ADDIS ABABA UNIVERSITY

SCHOOL OF GRADUATE STUDIES

Stochastic Simulation of Streamflow and Hydrologic Drought Analysis (Case Study: Upper Blue Nile Basin).

By

Yohannes Hagos Subagadis

Approved by Board of Examiners:

1.	Dr. Seleshi Bekele Awulachew		29/7/2009
	Advisor	Signature	Date
2.	Dr. Yilma Seleshi		30.07.2009
	Internal Examiner	Signature	Date
3.	Dr. Habtamu Hailu		31.07.2009
	External Examiner	Signature	Date
4.	Ato Habtamu H.		03/08/2009
	Chairman	Signature	Date

Abstract

This thesis is concerned with the problem of stochastic simulation of streamflow and hydrological drought analysis in the upper Blue Nile basin (Abbay).Thirteen stations with long record period ranging from 35 to 45 years were used for the analysis. Two sets of time series models named PARMA (p,q) and Thomas-Fiering were tested and inter-compared for the stochastic simulation of monthly streamflow , and a suitable model were selected.

PARMA models of low order are found to be most appropriate for all stations than Thomas – Fiering model. The performance of the model ranges from R^2 value of 0.893 at Gilgel Abbay near Merawi station to 0.33 at Abbay near Bahirdar station. The Thomas-Fiering model gives similar results; its R^2 values ranging from 0.87 at Gilgel Abbay near Merawi station to 0.32 at Abbay near Bahirdar station. But both models did not give acceptable results for the gauging station at Abbay near Bahirdar .In general, best fit PARMA (p,q) models was found to perform better in most of the gauging stations than Thomas-Fiering model considered in this study. Therefore, long sequence of synthetic flow used for drought analysis in the study were generated using PARMA (p,q) model.

Hydrological drought parameters namely, drought duration (D) and drought severity (S) were derived and analyzed from historic and long sequence of synthetically generated data at Gilgel Abbay near Merawi and Guder near Guder stations. The threshold method has been used to identify drought parameters in the streamflow series. Both drought severity and duration were found independent without a trace of trend within it. Accordingly the important drought characteristics were determined and the suitable probability distribution for each parameter was arrived at after studying different probability models. The use of the probability curves thus derived has been also illustrated.

It can be concluded that the duration and severity of drought in the basin doesn't show any trend of increasing or decreasing. However, in the event of other circumstances like climate change, the trend may evolve differently. Future studies may investigate the impact of climate change in the severity and duration of drought.

ACKNOWLEDGEMENT

I would like to express my wholehearted gratitude to my advisors, Dr. Semu Ayalew Moges, for his priceless support in supervising my work and providing me with important reference materials, for his invaluable guidance, constructive comments and his encouragement from early stage of my study.

My sincere appreciation goes to all staff members of Ministry of Water Resource, particularly Hydrology department, and Library, for their appreciable support in providing me hydrological data, digitized map and other reference materials.

I wish to extend my gratitude to all staff members of Addis Ababa University, who helped me one way or the other, in carrying out my research through remarkable encouragement, advice, material support and collaboration in every aspect. I would also like to thank my family and Aster A. for their unconditional love and continuous support of my academic career.

Finally, I owe whatever I have accomplished to God without Whom I could accomplish nothing.

Table of Content

List of Tables

List of Figures

Abbreviations

ARMA - Autoregressive moving average

ACF - Autocorrelation function

PACF- Partial autocorrelation function

PARMA- Periodic Autoregressive moving average

AICC- Akakia information criteria corrected

BIC- Bayes information criteria

SIC- Schwarz Bayesian criteria

WN- White nose

SSE- Sum of squares of errors

n- Size of sample

P- Order of autoregressive

q- Order of moving average

k- Sum of p and q

Φ- Periodic autoregressive parameter

Θ- Periodic moving average parameter

w- Number of seasons

$Y_{\nu,\tau}$ - Streamflow process for year ν and season τ

$\varepsilon_{\nu,\tau}$ - Uncorrelated nose term

δ^2_{τ} - Standard deviation

S - Hydrological drought severity

D- Hydrological drought duration

1. INTRODUCTON

1.1 Background

Stochastic simulation of streamflow time series has been widely used for solving various problems associated with the planning and management of water resources systems for several decades. Typical examples are the determination of a reservoir capacity, determining the risk of failure of dependable capacities of hydroelectric systems, evaluation of adequacy of a water resource management strategy under various potential hydrologic scenarios, hydrological drought analysis, and for many other purposes (Salas *et. al*, 1985; Salas, 1993).

Hydrologic drought properties (such as duration and severity) of various return periods, for example, are needed to assess the degree to which power generation ,Agriculture, water supply and so on will be able to cope with future droughts and, accordingly, to plan alternative water supply strategies. They can be determined from the historical record alone by using nonparametric methods but, because the number of drought events that can be drawn from the historical sample is generally small, the "historical" drought properties have a large degree of uncertainty. Consequently, the stochastic models are used here to generate long series of data so that adequate characteristics of the drought can be captured. Therefore, this study focuses on selecting appropriate stochastic model to generate long time series data and use the data to evaluate the hydrological drought condition on the Upper Blue Nile Basin.

1.2 Problem Description

Due to population growth and expansion of agriculture and industrial sectors, the water demand has been increased in many parts of the world. Many other factors such as climate change and contamination of water supplies have contributed in the scarcity .The flood and drought events have been experienced with higher peaks and severity level.

Drought has long been associated with Ethiopia also, with records indicating such conditions as far back as 250 BC. Moreover, the frequency of drought has been increasing more recently. In the 1970s and 1980s, droughts typically occurred, on average, once per decade; presently, droughts are anticipated to occur about once in every three years (USAID, 2003).

These facts and other related problems thus magnifies the need for integrated planning and management of water resource systems, including the ability of the system to face period of water shortages.

Stochastic simulation of hydrologic processes like streamflow, has become standard tool for analyzing many water related problems. Simulation signifies the behavior of the underlying process so that realistic representations of it can be made. Stochastic simulation enables one to obtain equally likely sequences of hydrologic processes that may occur in the future , which are used for estimating drought properties , such as drought duration , severity and intensity at the key points in the water supply system among others.

An understanding of the drought conditions in time and space is fundamental to a wide range of water management problems. Drought occurs in virtually all climatic zones .However; its characteristics vary significantly from one region to other .Statistical techniques dealing with the duration aspect of drought are reasonably well developed, whereas techniques for severity aspects are less satisfactory and require considerable improvement and refinement (Panu&Sharma, 2002). Therefore, the main goal of the research presented here is to fit a stochastic model to represent a given monthly river flow data, estimate parameters, check for goodness of fit to the data, and possibly to use the fitted model for hydrological drought analysis.

1.3 Objective of the Research

The main objective of this research is to develop a stochastic hydrological model to represent the flows of the Upper Blue Nile river basin system and generate data applicable for hydrological drought analysis.

Specific objective

- Diagnosis of applicable stochastic model structure for monthly time series of selected tributaries in the upper Blue Nile river basin (Abbay).
- Analysis of hydrologic drought properties (such as duration and severity) of various return periods in the basin.

- Determine the suitable probably distribution for each drought parameter (duration and severity) among different probability models.
- Investigate the dependence between duration and severity of hydrological drought if any.

1.4 Research Questions

1. What is a suitable stochastic model applicable in the Upper Blue Nile tributaries?
2. What is the largest drought duration between probable droughts and what is its severity?
3. What is the most severe drought between the probable drought events?
4. Is there any dependence between drought duration and severity?

1.5 Significance of the research

Persistent incidences of drought in Ethiopia, in most cases, have resulted in famine and devastating subsidiary economic activities. Beyond its short-term effects, such as the loss of human life, livestock and wildlife, lie long-term social and environmental impacts. The increase in water demands for municipalities, hydropower and agriculture coupled with increased threat of global warming requires a better understanding of drought characteristics and sustainable use of water resources. River systems are most stressed in low flow periods and thus, an understanding of drought conditions in time and space is fundamental to a wide range of water management problems.

The result of this study can thus be used as an input for informed decision making from a basin perspective which later could be developed to regional scale. As an example, Synthetic flows are vital to construct hydrological series such as the full time series, the flood and drought series. Understanding of specially the evolution of extreme flow sequences provides information for Decision Support System (DSS) to understand the possible occurrence of consecutive drought / wet years from which one set of DSS scenario can be formulated

1.6 Description of the Study Area

The study area, Upper Blue Nile river basin, is known as Abbay in Ethiopia which lies in the western part of Ethiopia, between $7^0 45'$ and 12^0 45'N, and $34^0 05'$ and $39^0 45'$ E as shown in Fig 1.1. It is the largest in terms of volume of discharge, second largest in terms of area in the country. The river contributes over 50 % of the long-term river flow of the Main Nile (Conway, 2000). The basin covers an area of 175,000 km^2 having 14 sub basins. High population growth and environmental degradation, limited resources, and inadequate land use and water policies have increased Ethiopia's vulnerability to drought disaster.

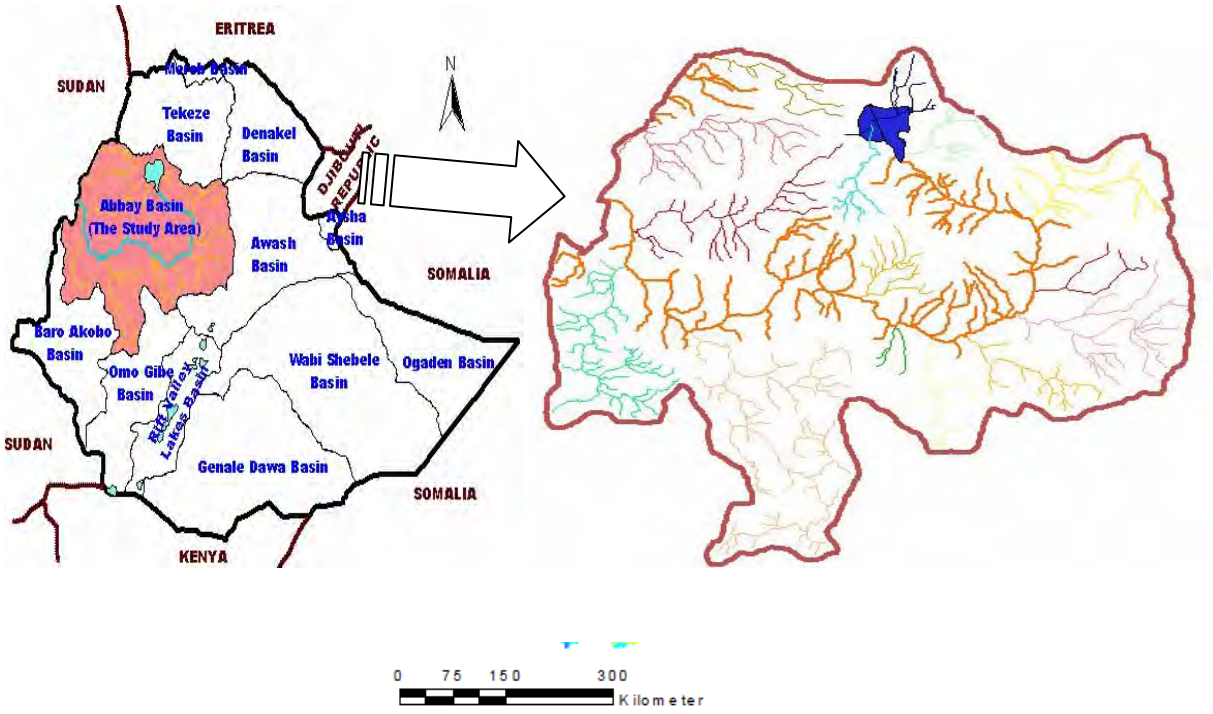


Fig1.1: Map of the Ethiopian's River Basins with the study basin

The annual rainfall over the basin decreases from the south - west (>2000 mm) to the north -east (around 1000 mm), with about 70 per cent occurring between June and September (Conway, 2000). Conway (1997) explained that the Blue Nile is characterized by severe seasonality with average annual flow of about 50 km^3 measured at the basin outlet at El Diem station near the Sudan-Ethiopia border. More than 80% of this flow occurs during the flood season between July and September, while only 4% of that flow occurs during the driest period January-April. In the

basin, annual mean PE and rainfall range, with increasing elevation, from1800mm to 1200mm and 924mm to 1845mm, respectively.

In this study, on the bases of the record length, its continuity and suitability for streamflow analysis threaten stations were considered for the analysis. Generally the distributions of the gauging stations used in this study are shown in Figure 1. 2.

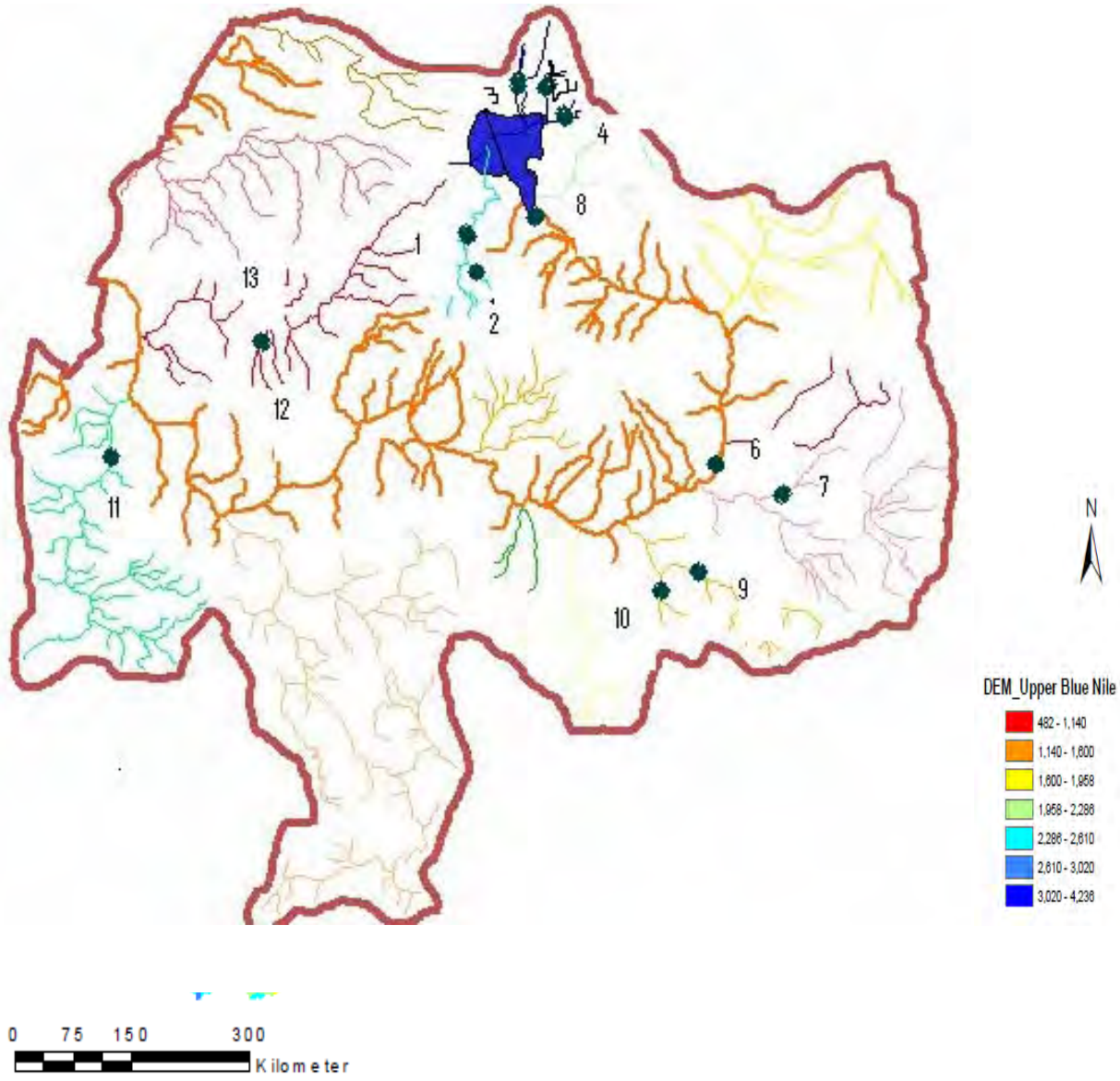


Fig1. 2: Distributions of the gauging stations used in the basin

2. LITERATURE REVIEW

2.1 General

Time series analysis and modeling is an important tool in hydrology and water resources. It is used for building mathematical models to generate synthetic hydrologic records, to determine the likelihood of extreme events, to forecast hydrologic events, to detect trends and shifts in hydrologic records, and to fill in missing data and extend records. Synthetic river flow series are useful for determining the dimensions of hydraulic works, for flood and drought studies, for optimal operation of reservoir systems, for determining the risk of failure of dependable capacities of hydroelectric systems, for planning capacity expansion of water supply systems, and for many other purposes (Salas *et. al*, 1985; Salas, 1993). For example, hydrologic drought properties (severity and duration) of various return periods are needed to assess the degree to which a water supply system will be able to cope with future droughts and, accordingly, to plan alternative water supply strategies.

Drought parameters can be determined from the historical record alone by using non-parametric methods but, because the number of drought events that can be drawn from the historical sample is generally small, the “historical” drought properties have a large degree of uncertainty. Other alternatives for finding drought properties include using stochastic models (such as PARMA models) that can represent the underlying river flows, simulating long records of such river flows, and then deriving droughts properties from the simulated samples based on the theory of runs .

2.2 Characteristics of Hydrological Time Series

The structure of hydrological time series consists of mainly one or more of these four basic structural properties and components (Salas, 1993):

- Over years trends and other deterministic changes (such as shifts in the parameters). In general, natural and human induced factors may produce gradual and instantaneous trends and shift in hydrological time series. See, for example, Salas (1993) for a detailed discussion of trends and shifts in hydrological data and their removal.
- Intermittency in the processes, mainly consisting of the hydrology of intermittent sequences of zero and non-zero values.
- Seasonal or periodic changes of days, weeks or months within the annual cycle. Periodicity means that the statistical characteristic changes periodically within the year. For example, in hydrologic data concerning river flows, we expect high runoff periods in the summer and low flow periods in the winter. Thus the river flow correlations between winter months may be different from the correlations between summer months.
- Stochasticity or random variations.

The rivers used as case studies in this research are perennial rivers and it does not reveal any significant trend or shift. Therefore, it is considered only periodicity and stochasticity in the modeling of the river flow series.

Although basic hydrologic processes evolve on a continuous time scale, most analysis and modeling of such processes are made by defining them on a discrete time scale. In most cases, a discrete time processes is derived by aggregating the continuous time process within a given time interval while in other cases it is derived by sampling the continuous process at a discrete point in time. For example, a weekly river flow series can be derived by aggregating the continuous flow hydrograph on a weekly basis while a daily river flow series can be the result of simply sampling the flows of a river once daily or by integrating the continuous flow hydrograph on a daily basis. In any case, most time series of hydrological processes are defined at hourly, daily, weekly, monthly, bimonthly, quarterly, and annual time intervals. In this research, the analysis is done using monthly time scale.It must be emphasized at this point that the term period (also used interchangeably with season) applies for any time resolution smaller than one year.

The plot of a hydrological series gives a good indication of its main characteristics. For example, the monthly and annual flows of the gauging station can give an indication. In general, most hydrological time series exhibit periodic variations in mean, variance, covariance and skewness. The periodicity in mean, as can be seen from Figure 2.1 may be easily observed in the plot of the underlying hydrologic time series. However, the periodicity in higher order moments is not so easy to observe and usually requires further mathematical analysis.

Autocorrelation analysis can also be used to identify cycles or periodic components of hydrological time series. The non-decaying nature of the correlation function can be observed, indicating a correlation between flows at 12 months' lag.

Spectral analysis has also been used to detect cycles in hydrological data (Yevejevich,1966 and Salas, 1993).

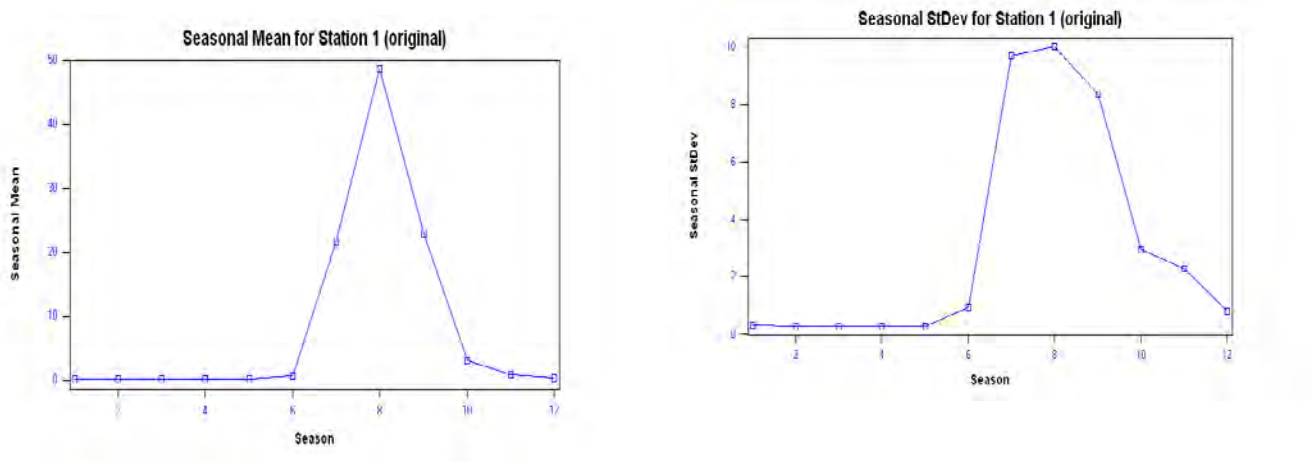


Figure 2.1: Plot of Seasonal mean and standard deviation of Gilgel Abby near Merawi.

2.3 Modeling and Simulation of Hydrological Series

A number of approaches have been suggested for modeling hydrological time series defined at time intervals less than a year (Salas, 1993). The common procedure in modeling such periodic river flow series is first to standardize or filter the series and then fit an appropriate stationary stochastic model to the reduced series (Salas, *et al.*, 1980). However, standardizing or filtering most river flow series may not yield stationary residuals due to periodic autocorrelations. In these cases, the resulting model is misspecified (Tiao and Grupe, 1980). Periodic models can, therefore, be employed to remove the periodic correlation structure. An important class of periodic models useful in such situations consists of periodic autoregressive moving average (PARMA) models, which are extensions of commonly used ARMA models that allow periodic parameters. The PARMA modeling procedure involves iterative steps of model identification, parameter estimation, model diagnosis and fitting the residuals (noise) with a probability distribution function (pdf). The opposite process to stepby- step modeling is the use of models to generate (simulate) new samples or a long sample of the process. One starts with the random noise and its pdf by generating its sample(s).Then generate the corresponding data samples by using the fitted PARMA model.

2.4 Review of previous studies

Early studies by Hazen (1914) and Sudler (1927) showed the feasibility of using statistics and probability theory in analyzing river flow sequences. Hurst (1951) in investigating the Nile River for the Aswan Dam project, reported studies of long records of river flows and other geophysical series, which years later tremendously impacted the theoretical and practical aspect of time series analysis of hydrologic and geophysical phenomena. Barnes (1954) extended the early empirical studies of Hazen and Sudler and introduced the idea of synthetic generation of stream flow by using a table of random numbers. However, it was not until the beginning of the 1960's that the formal development of stochastic modeling started with introduction and application of autoregressive models for annual and seasonal stream flows .

Since then, extensive research efforts have been made towards improving the earlier concepts and models, providing physical justification of some models, introducing alternative models, developing and applying improved estimation procedures as well as fitness tests for such models, and studying their impact in water resources systems planning and management. The concept of periodically correlated processes was introduced by Gladyˇshev(1961). He gave a formal definition of periodic stationarity for a general periodic process and showed that a necessary and sufficient condition for such stationarity is the stationarity of the so-called "lumped" vector process. The first application of periodic time series models seems to have been by hydrologists Thomas and Fiering (1962). They used the lag-one autoregressive (AR) model for modeling monthly stream flow. Since then there have been many discussions and summaries about periodic time series models .

For example, Noakes *et al*. (1985) demonstrated the superiority of periodic autoregressive models among several other competitors in forecasting thirty monthly river flow time series. Troutman (1979) studied some properties of the periodic autoregressive models using a related multivariate autoregressive representation. Tiao and Grupe (1980) showed how periodic autoregressive moving average models may be misspecified as homogeneous models. Because of the seasonal parameters, the estimation of periodic models is more difficult than that of the homogeneous models. Pangano (1978) dealt with moment estimation of parameters in periodic autoregressive (PAR) models. He showed that estimates of the seasonal parameters obtained by using the seasonal Yule-Walker equations possess many desirable properties, including

maximum asymptotic efficiency under normality. Salas *et al*. (1982) suggest estimating the parameters of PARMA models using the seasonal Yule-Walker equations. Vecchia (1985b) proposed an algorithm for the maximum likelihood estimation for periodic ARMA models. Li (1988) developed an algorithm for exact likelihood of periodic moving average models. Anderson, Meerschaert and Vecchia (1999) developed the innovations algorithm for estimation of PARMA model parameters. Lund and Basawa (2000) examined the recursive prediction and likelihood evaluation techniques for PARMA models, which are extensions of Ansley (1979) for ARMA series. Shao and Lund (2004) studdied correlation and partial autocorrelation properties PARMA time series models. Additional review of literature is given at the beginning of the remaining chapters.

Concerning the previous studies on the basin the author finds no scientific work done in the basin. This could probably form the first hydrologic time series analysis study in the basin.

2.5 PARMA Models and Their Parameter Estimation

2.5.1 General

A time series is a set of observations (X_t), each one being recorded at a specific time *t*. For example, (X_t) can be river flow (daily, monthly, etc.) measurements at time *t*. In general, a collection of a random variables, indexed by *t* is referred to as a stochastic process. The observed values of a stochastic process are referred to as a realization of the stochastic process.

A mathematical model representing a stochastic process is called a stochastic model or time series model. The model consists of a certain mathematical form or structure and a set of parameters. Such models are built to "reproduce" or to "resemble" the main statistical characteristics (mean, standard deviation, skewness, autocorrelation, range and run) of the time series. Several stochastic models have been used for modeling hydrological time series in general and stream flow time series in particular (Salas *et al.*, 1981). Unfortunately, the exact mathematical model of a hydrological time series is never known. The exact model parameters are also never known, they must be estimated from limited data.

2.5.2 Mathematical Formulation of PARMA Model

Stationary ARMA models have been widely applied in stochastic hydrology for modeling of annual time series where the mean, variance, and the correlation structure do not depend on time. For seasonal hydrologic time series, such as monthly series, seasonal statistics such as the mean and standard deviation may be reproduced by a stationary ARMA model by means of standardizing the underlying seasonal series. However, this procedure assumes that season-to-season correlations are the same for a given lag. Hydrologic time series, such as monthly streamflows, are usually characterized by different dependence structure (month-to-month correlations) depending on the season (e.g. spring or fall). Periodic ARMA (PARMA) models have been suggested in the literature for modeling such periodic dependence structure. A PARMA(p,q) model may be expressed as (Salas, 1993):

$$Y_{\nu,\tau} = \sum_{i=1}^{p} \phi_{i,\tau} Y_{\nu,\tau-i} + \varepsilon_{\nu,\tau} - \sum_{j=1}^{q} \theta_{j,\tau} \varepsilon_{\nu,\tau-j} \qquad 2.1$$

Where, $Y_{\nu,\tau}$ represents the streamflow process for year ν and season τ. For each season, τ, this process is normally distributed with mean zero and variance δ^2_τ (*Y*). The $\varepsilon_{\nu,\tau}$ is the uncorrelated noise term which for each season is normally distributed with mean zero and variance δ^2_τ (ε). The $\{\Phi_{1,\tau}\ldots\ldots\Phi_{p,\tau}\}$ are the periodic autoregressive parameters and the $\{\theta_{1,\tau}\ldots..,\ \theta_{q,\tau}\}$ are the periodic moving average parameters. If the number of season s or the p eriod is ω, then a PARMA(p,q) model consists of ω number of individual ARMA(p,q) models, where the dependence is across seasons instead of years.

2.5.3 Parameter Estimation

Parameter estimation methods used in stochastic time series analysis generally fall into three categories:

- Method of moment
- Method of maximum likelihood and
- Method of least squares

The method of moments is based on taking as many moment equations as the number of parameters, substituting the population moment by the sample moments, and solving the equations simultaneously for parameters.

In the method of maximum likelihood, the likelihood function is first determined (this function is a function of the parameters given the observations), then the function (or its logarithm) is maximized and the parameters corresponding to such a maximum are the maximum likelihood estimators.

In the method of moment and least squares, the parameters which minimize the sum of square residuals $\sum \varepsilon_t^2$ are the least squares estimators; The method of moment and lest square method may be used in parameter estimation of low order PARMA (p,q) models (Salas et al., 2007). As an illustration the moment estimators for the PARMA (1, 0) (eqa.2.2-2.5) and PARMA (1,1)(eqa.2.6-2.9) models using the method of moment(MOM) are shown below (Salas et al.,1982).

-PAR(1) $$Y_{\upsilon,\tau} = \phi_{1,\tau} Y_{\upsilon,\tau-1} + \varepsilon_{\upsilon,\tau} \qquad 2.2$$

$$\hat{\mu}_{\tau} = \bar{Y}_{\tau} \qquad 2.3$$

$$\hat{\phi}_{1,\tau} = \left(\frac{S_{\tau}}{S_{\tau-1}}\right)\upsilon_{1,\tau} \qquad 2.4$$

$$\hat{\sigma}^{2}_{\tau}(\varepsilon) = S^{2}_{\tau} - S^{2}_{\tau-1}\upsilon^{2}_{1,\tau} \qquad 2.5$$

-PARMA(1,1) $Y_{\nu,\tau} = \phi_{1,\tau} Y_{\nu,\tau-1} + \varepsilon_{\nu,\tau} - \theta_{1,\tau}\varepsilon_{\nu,\tau-1}$ 2.6

$$\hat{\phi}_{1,\tau} = \frac{m_{2,\tau}}{m_{1,\tau-1}} \qquad 2.7$$

$$\hat{\theta}_{1,\tau} = \hat{\phi}_{1,\tau} + \frac{S^{2}_{\tau} - \hat{\phi}_{1,\tau}\hat{m}_{1,\tau}}{\hat{\phi}_{1,\tau} S^{2}_{\tau-1} - m_{1,\tau}} - \frac{\phi_{1,\tau+1}\hat{S}^{2}_{\tau} - m_{1,\tau+1}}{(\hat{\phi}_{1,\tau} S^{2}_{\tau-1} - m_{1,\tau})\hat{\theta}_{1,\tau+1}} \qquad 2.8$$

$$\hat{\sigma}^{2}_{\tau}(\varepsilon) = \frac{\phi_{1,\tau+1}\hat{S}^{2}_{\tau-1} - m_{1,\tau+1}}{\hat{\theta}_{1,\tau+1}} \qquad 2.9$$

In a similar manner, the Least Squares (LS) method can be used to estimate the model parameters of PARMA (p,q) models. In this case, the parameters Φ's and θ's are estimated by minimizing the sum of squares of the residuals defined by

$$F = \sum_{\nu=1}^{N}\sum_{\tau=1}^{\omega}\varepsilon^{2}_{\nu,\tau} \qquad 2.10$$

Where ω is the number of seasons and *N* is the number of years of data. For the PARMA (p,q) model, the residuals are defined as

$$\varepsilon_{\nu,\tau} = Y_{\nu,\tau} - \sum_{i=1}^{p}\phi_{i,\tau} Y_{\nu,\tau-1} + \sum_{j=1}^{q}\theta_{j,\tau}\varepsilon_{\nu,\tau-j} \qquad 2.11$$

Once the Φ's and θ's are determined the seasonal noise variance $\sigma^{2}_{\tau}(\varepsilon)$ can be estimated by

$$(1/N)\sum_{\nu=1}^{N}\varepsilon_{\nu=\tau}^{2} \qquad 2.10$$

For this study both method of moment and least square methods are used for parameter estimation.

2.6 Thomas–Fiering model

The monthly flow series are non-stationary and therefore complicated mathematical models are employed in their simulation. The first model that appeared in the hydrology literature for the generation of synthetic monthly flow sequences is that due to Thomas & Fiering (1962). Basically, this model is of a Markovian nature with periodic parameters, namely, the monthly means, standard deviations and the lag-zero crosscorrelations between successive months. In its simplest form the model consists of twelve regression equations, one for each month. The model implicitly allows for the non-stationarity observed in monthly flow data.

The Thomas–Fiering model is given by

$$X_{i,j} = \bar{X}_j + b_j(X_{i,j-1} - \bar{X}_{j-1}) + \varepsilon_t \sigma_j \sqrt{(1-r_j^2)} \qquad 2.11$$

Where, $\bar{X}_j$ and $\bar{X}_{j-1}$ are the mean flows in seasons j and j-1 respectively. σ_j is the standard deviation of flow in the season j.

r_j is the correlation between flows in the j^{th} and (j-1)th seasons. b_j is the regression coefficient of flow in the j^{th} season with flow in the (j-1)th season. ε_t is the normal random variate. $x_{i,j}$ is the flow generated in the j^{th} season of the i^{th} year, with j = 1, 2, ...,m and i = 1, 2, ... If a season is a month, then m = 12; for a weekly model, m = 52;

for a daily model, m = 365.

It may be noted that in the Thomas–Fiering model the flow in any season j is a sum of three terms. The first term is the mean flow in that season. The second term is the regressed component on the flow in the previous season. The third term is a random component to reflect the variance. This model accounts for the persistence only up to lag-1. Hence, the Thomas–Fiering model may be viewed as a non-stationary first-order AR model. This model can be applied directly in the form given in Equation (2.10), drawing the random variate ε_t from a normal distribution with zero mean and unit variance [N(0,1)] provided that , the flow in all seasons is normally distributed.

2.7 Hydrological Drought Analysis

2.7.1 General

Obtaining drought properties is important for planning and management of water resources system. For example, the design of water supply capacity of a given city may be based on meeting water demands during a critical drought that may occur in a specified planning horizon (Frick *et al.*, 1990). Moreover, the estimation of return periods associated to severe droughts can provide useful information in order to improve water systems management under drought condition.

It is difficult to give a universal definition since there exists widely diverse views about the interpretation of droughts among the scientists of different disciplines. In a broader sense, a drought is defined as a deficit of water in time. It can be divided into three categories meteorological, hydrological, and agricultural drought-which may occur as a result of precipitation, river flow, and soil moisture deficits, respectively (see for example, Dracup *et al*. 1980; Sen 1980, etc). In this study the emphasis is on hydrological drought, which is defined as uninterrupted sequences of river flows below a threshold level. It has three components duration, intensity, and severity. Since river flow sequences are stochastic variables, the corresponding drought properties are random and must be described in probabilistic terms.

Yevjevich (1967) proposed the theory of runs as a major tool to use in objectively defining droughts and studying their statistical properties. Many studies of drought properties followed Yevjevich's definition of droughts. For example, probabilistic behavior of drought properties has been derived analytically, assuming a given stochastic structure of the underlying hydrological series (Cancelliere and Salas, 2004).

Hydrologic droughts can be determined from the historical record alone by using nonparametric methods but, because the number of drought events that can be drawn from the historical sample is generally small, the "historical" drought properties have a large degree of uncertainty. Other alternatives for finding drought properties include using stochastic models that can represent the

underlying river flows, simulating long records of such river flows, and then deriving droughts properties from the simulated samples based on the theory of runs.

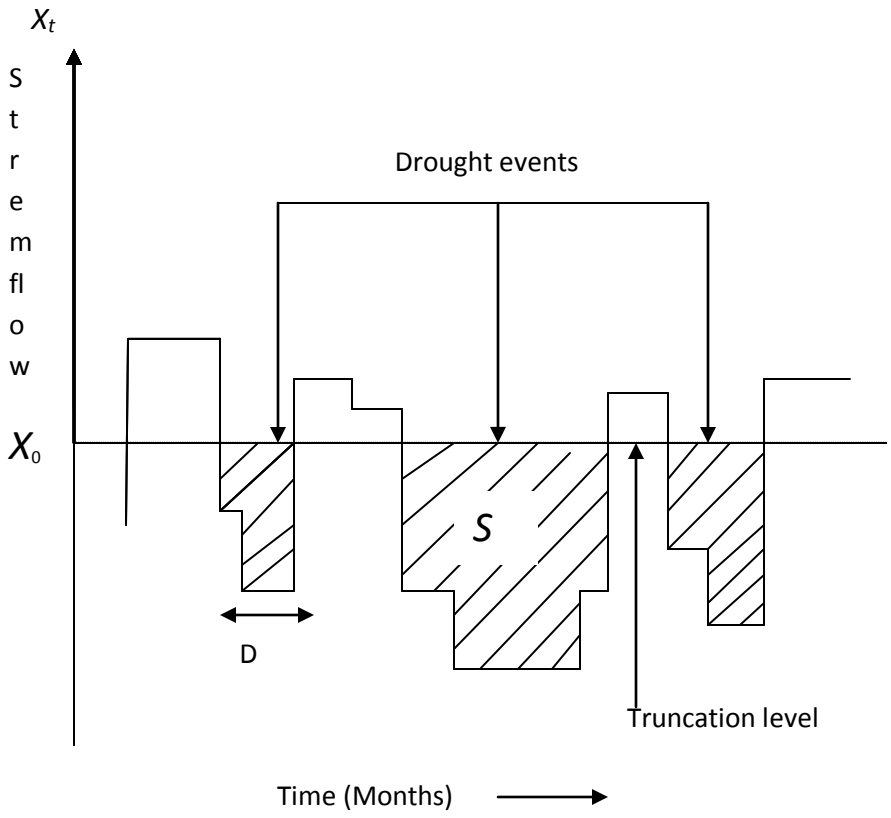


Figure 2.3: Definition sketch of drought parameters

2.7.2 Time resolution of the data series

Choosing an appropriate concept to study droughts depends also on the time resolution of the available data and vice versa the most favorable time resolution depends on the purpose and outline of the study, the characteristics of the streams under study, the methods one wants to apply, and the available computing tools. A daily time series of course contains more detailed information about the stream's discharge and about drought events, but also discharge series with a larger time interval can be favorable for various reasons. In the past, hydrological drought studies have been based on anything from daily up to annual time series. In general, local scale data records often have a resolution of days or months and local studies are preferentially based

on high resolution data, whereas studies with a larger spatial coverage and/or temporal extent are often based on time-aggregated seasonal or annual date (Stahl & Hisdal, 2004).

The droughts themselves are prolonged periods and a single drought event can last a couple of days, several years or any time period in between. Therefore when choosing the appropriate time resolution, an important aspect to consider is the duration of the studied droughts. In a semi-arid climate a dry period might only considered to be a drought when it lasts for more than a year while in a humid climate some streams might never experience multi-year droughts. In the first case, a statistical description of long multi-year droughts might be conducted much more easily when based on annual data, especially when it comes to identify return periods of the observed drought events, and a time series with annual resolution might be favorable.

2.7.3 Methods of characterizing hydrological droughts

Various lecturers provide diverse methods of characterizing hydrological droughts (Salas, 1993 &FriendNile, 2004). Most frequently noticed methods include the following.

2.7.3.1 Flow duration curve (FDC) and percentiles

The flow duration curve (FDC) displays for all observed discharge values the percentage of time during which higher discharge values are observed. As such it plots the discharge above its exceedance frequency (Figure 5.2). A discharge value which is exceeded in x % of the time is referred to as the x-percentile of the FDC, Q_x.

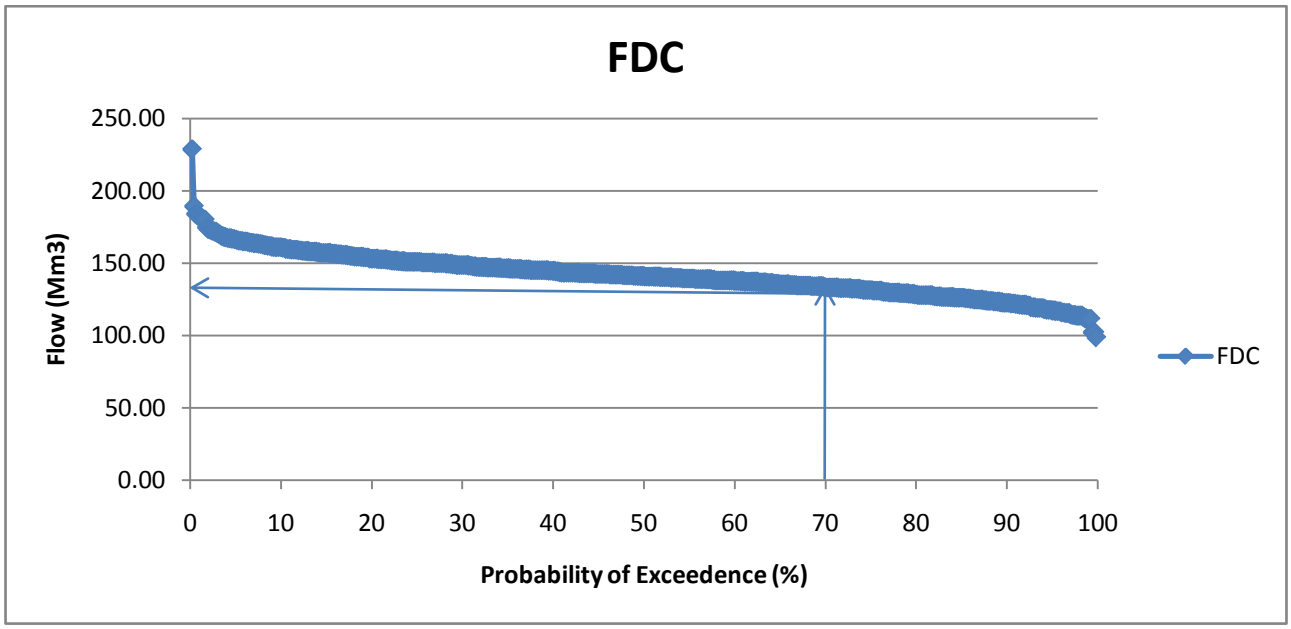


Figure 5.2: Flow duration curve of Gilgel Abbay near Merawi indicating Q70 flow.

Low flow indices derived from the FDC are the percentiles which indicate a high frequency of exceedance and therefore present the low flow period of a regime. Common percentiles used as low flow indices are between *Q*95, and *Q*70. They are also frequently chosen as value for the threshold level in drought event definitions (Zelenhasić &Salvai, (1987).

A FDC can be calculated for data with any kind of time step and for any record length. Most commonly the whole period of record is used. The FDC is calculated by assigning each discharge value its rank, (M) in descending order, which means that the largest value gets rank 1, and then the values are plotted over p, which is the percentage of data which exceeds a value as it is clearly explained in the following steps.

Step 1: Sort (rank) average daily discharges for period of record from the largest value to the smallest value, involving a total of n values.

Step 2: Assign each discharge value a rank (M), starting with 1 for the largest daily discharge value.

Step 3: Calculate exceedence probability (P) as follows:

$P = 100 * [M / (n + 1)]$

P = the probability that a given flow will be equaled or exceeded (% of time)
M = the ranked position on the listing (dimensionless)
n = the number of events for period of record (dimensionless)

2.7.3.2 Threshold level method

The threshold level method originates from the theory of runs as introduced by Yevjevich in 1967 (Smakhtin, 2001). Yevjevich originally defined droughts as periods during which the water supply does not meet the current water demand, and both the water supply, $I(t)$ as well as the water demand, $D(t)$ were expressed as time series with the same temporal resolution (Figure 5.3). A time series of drought events is then obtained through the series of uninterrupted sequences of negative values of the supply-minus-demand time series, $Y(t)$.

Each uninterrupted sequence of negative values constitutes one drought event.

$$Y(t) = I(t) - D(t)$$

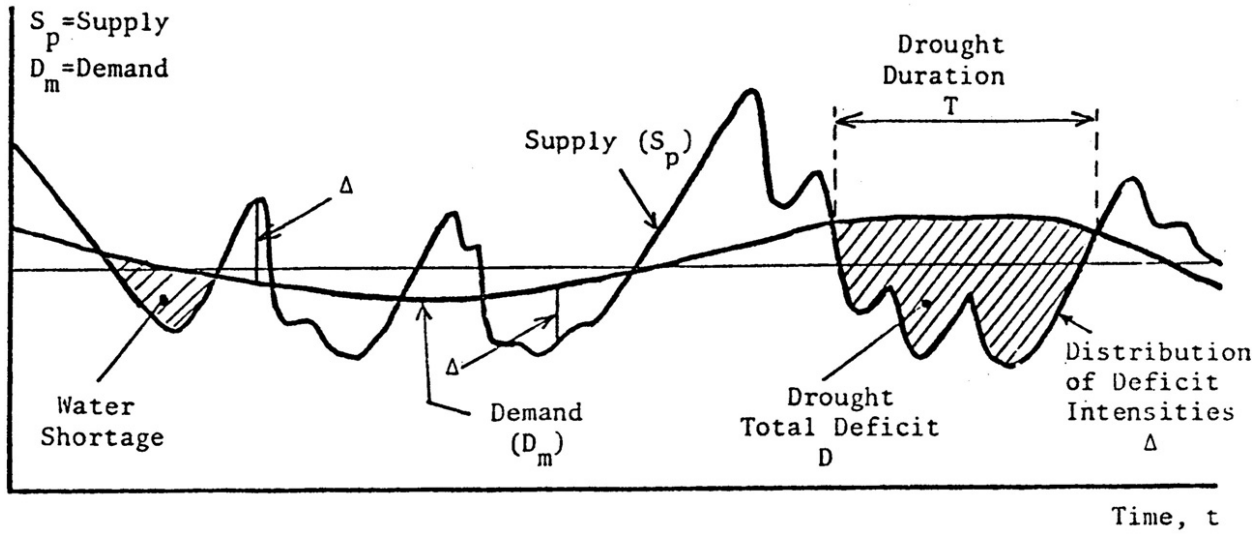


Figure 5.3 Supply-minus-demand series for the definition of droughts (Yevjevich, 1983).

The statistical problem arising from this definition was the complexity of the supply-minus demand series. Usually $I(t)$ is a stochastic or periodic-stochastic process and $D(t)$ a trend periodic-stochastic process. This results in the $Y(t)$ series as a trend-periodic-stochastic process, with periodicity and stochasticity from both, supply and demand time series and trend mainly coming from the demand time series (Yevjevich, 1983). And according to Yevjevich (1983) "the application of the theory of runs to complex trend-periodic-stochastic processes has not yet been developed to a degree of a reliable current use." In order to be able to describe the supply-minus-

demand series in a statistically correct way, Yevjevich simplified the concept by applying a constant demand (Yevjevich, 1983). The demand is now represented by a threshold level, Q_0 and droughts are defined as periods during which the discharge is below the threshold level.

The application of a threshold level allows both, defining drought events as periods with discharge below normal as well as identifying periods with insufficient water supply for a specific demand. In the latter case the threshold level is set equal to the discharge demand, but in the former case it is considered to represent 'normal' conditions, which means that it can be chosen more on study. It is however common to apply as a low flow index an objective and comparable value for the threshold level. Frequently used threshold levels are between$Q95$ and $Q70$ (Zelenhasić &Salvai (1987).

2.7.4 Identification of Drought Parameters

In this study drought events (drought duration D, and drought deficit volume or severity S) were identified using the threshold level method with constant Q0, both at the yearly and monthly bases. From frequently used threshold level values Q70 for yearly drought extraction and Q80 for monthly drought extraction is used here for both stations. While identifying the drought events the following assumptions are made.

1. Streamflow deficits exceeding the threshold level deficit of any year/month have been included in the evaluation of critical drought
2. The drought events are assumed to be independent. It can happen that some of the generated droughts would be very close to each other and, therefore, mutually dependent. To cope with the problem of dependence between these drought events, without significant influence on the accuracy of the final results, all droughts with deficits S_i, i = 1, 2, …, satisfying the condition S_i < 0.01 times the maximum deficit are neglected and are not considered in this study.
3. If the time period separating droughts (intervening time) is less than or equal to 2 years and the surplus volume during this intervening period is very small (<0.01 times the maximum deficit of either of the droughts), then these droughts are assumed to be continued.

3. DATA ANALYSIS

3.1 Source and Availability of Data

For the purpose of this study monthly streamflow data from thirteen gauging stations in the upper Blue Nile river basin were obtained from the Ministry of Water Resource hydrology department, for this study. The stations were chosen, among those available, for their continuous record and their length of record for hydrological time series analysis. Table 3.1 gives the gauge location, the record length, and the descriptive statistics of streamflow for each station. Stations with area as small as 96 km^2 and as big as 65784 km^2 are included in the analysis. In case of the mean flow of the stations considered in this study it ranges from 159.7 to 17000 Mm^3 as it is shown in the Table 3.1.

Table: 3.1 Gauging stations and their descriptive statistics of monthly streamflow data.

Station ID	Station Name	Latitude	Longitude	Length of record	Area (Km^2)	Mean	StDev
111002	Gilgel_Abbay_Near_Merawi	11.4	37	45	1664	1714	206.5
111003	Koga_Near_Merawi	11.4	37.1	45	244	159.7	76.53
111005	Ribb_Near_Addis_Zemen	12	37.7	43	1664	452.2	149.6
111006	Gumara_Near_Bahir_dar	11.8	37.6	35	1394	1030	349.4
111007	Megeche_Near_Azezo	12.5	37.5	44	462	186.6	87.09
112001	Abbay_Near_Kessie	10.1	38.2	49	65784	17000	6189
112002	Muger_Near_Chancho	9.3	38.7	35	489	249	72.32
112003	Abbay_Near_Bahir_dar	11.6	37.4	44	15321	3796	1420
113002	Fatto_Near_Guder	8.9	37.7	45	96	73.24	16.76
113005	Guder_At_Guder	9	37.1	45	524	391.8	80.2
113012	Gudla_Near_Dembecha	10.6	37.5	44	242	315	86.1
114001	Diddesa_Near_Arjo	8.7	34.4	45	9981	3768	908.6
116005	Main_Beles_Near_Metekel	11.3	36.4	42	3431	1435	489

3.2 Data considerations

The data from the Upper Blue Nile River Basin are in this section evaluated based on their suitability for hydrological time series analysis and drought analysis. This includes two main aspects: (1) the quality of the data and (2) properties of the time series, such as trends and seasonality. In addition, one has to make sure that all used data series have been treated in the

same way and that the chosen series are really comparable. The most basic example would be that they all should be in the same unit and averaged over the same time interval. Here all data is monthly data in Mm^3.

3.2.1 Data Quality

To receive a correct and appropriate data series a proper quality control includes checks for the length of time series, accuracy and continuity, and it starts with looking at the processes of measuring and collecting the data (Mosley & McKerchar, 1993). For the data series used in this study little information about the collection and any possibly done pre-processing of the data was available. Inaccuracies and problems in the data can be caused in an endless number of ways and they are not always as easily identifiable as barely missing data. It could be human mistakes when measuring or processing the data, the measuring device itself could cause problems or there could be reasons in nature which can lead to wrong conclusions if they are unknown.

In this study single values causing sudden drop downs or peaks which were inconsistent with the rest of the data series were removed and it was interpolated instead. For example the time series data of Koga near Merawi station had two sudden peak values and these were removed and interpolated.

3.2.2 Filling in the Missing data

For stochastic stream flow modeling, systematic stream gauging records of sufficient length are required to warrant statistical parameters as the basis for determination .The gauging stations of upper Blue Nile river basin which are used in this study have experienced one or more missing values. As a result, regression analysis was used to fill the missing monthly data with satisfactory correlation coefficient (that is with R^2 value greater than 0.55) .The correlation was done based on neighboring station and geographical proximity.

3.2.3 Data properties

3.2.3.1. Stationarity

A hydrologic time series is stationary if it is free of trends, shifts, or periodicity. Slow changes can complicate an analysis considerably, no matter whether they are a naturally occurring trend or actually an error in the data. It is therefore important to know about a trend in a time series to avoid any wrong conclusions or interpretations.

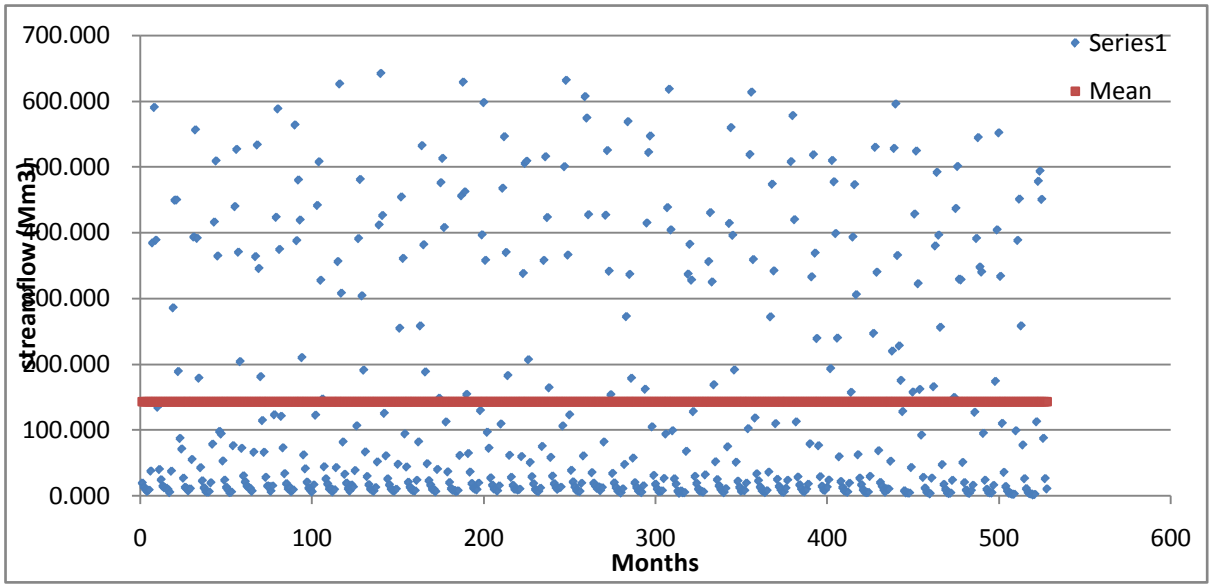


Figure 3.1: Average monthly flow (Mm^3) for Gilgel Abbay near Merawi indicating seasonal stationarity pattern.

In this study the stationarity of the data are tested by looking at the plot of time series data and its fluctuation on the mean value of the series as it is on Fig 3.1as a result periodic stationarity is observed for all monthly time series data of the stations considered.

3.2.3.2. Seasonality (Periodicity) aspect

Hydrologic series defined at time intervals smaller than a year (such as monthly series) generally exhibits distinct seasonal (periodic) patterns due to astronomic cycle (Maidmant *et al* 1993). In most analysis and modeling of hydrologic time series , testing for seasonality in statistics is done by using simple procedures , mostly by observing the plot of the statistic under consideration versus the season τ .Fig 3.2 plots the mean $\grave{Y}_\tau$ versus τ =1,….,12 . The plot suggests that the $\grave{Y}_\tau$'s

during the low flow season are quite different from those of high-flow season. Thus, even though some of the Ỳ's in the low-flow season are similar to each other, one would conclude overall that $\grave{Y}_\tau$ is a seasonal (periodic) statistic. A similar argument can be made in relation to seasonality in other statistics such as like autocorrelation and spectrum of the series.

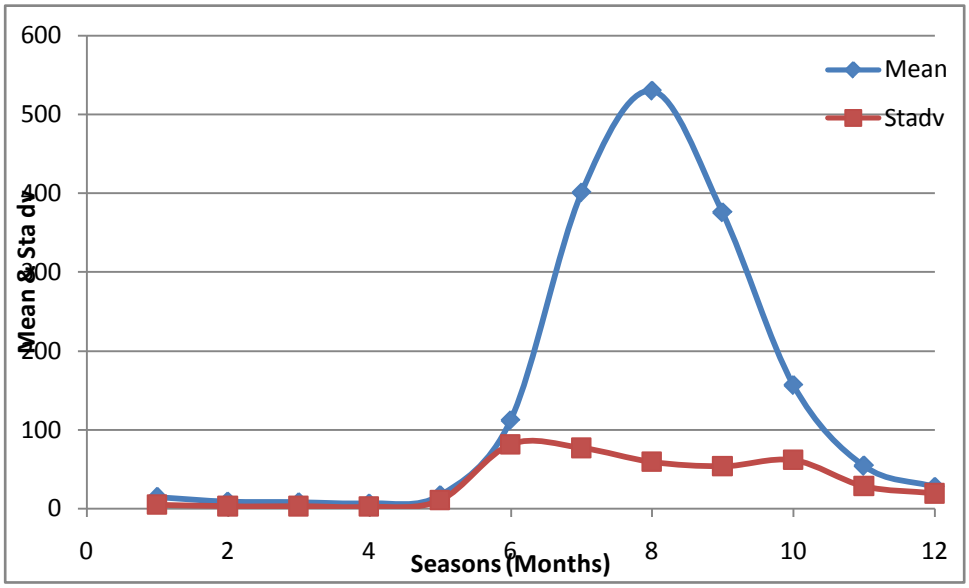


Figure 3.2: Plot of Seasonal mean and standard deviation of the original data for Gilgel Abbay near Merawi.

All stations considered here show distinctive seasonal characteristic with bimodal behavior, the peak being located in around August which is the main characteristic of the Nile basin stations.

3.2.3.3. Testing for Normality

PARMA model like other several models and approaches in hydrologic time series modeling assume that the variable under consideration is normally distributed. Therefore, it is usual practice to test the data for normality before further analysis. A widely used method for judging whether a certain data set is normally distributed is to plot the empirical frequency distribution of the data on normal probability paper. There are two commonly used plotting position formulas for plotting of the empirical frequency curve: the Cunnane plotting position, and the Weibull plotting position. The Cunnane plotting position is approximately quantile unbiased while the Weiibull plotting position has unbiased expedience probabilities for all distributions .In general the Cunnane plotting position is preferable (Salas *et al*., 2007).

In this study ; therefore, the Cunnane plotting position is used for all 13 stations normality test. In addition to graphical test for normality two normality tests are used in the study, namely the skewness test of normality (Snedecor and Cochran, 1980) and Filliben probability plot correlation test (Filliben, 1975) both applied at the 10% significance level (Table 3.2).

3.2.3.3.1 Transformation to Normal

In cases where the normality tests indicate that the observed series are not normally distributed, the data has to be transformed into normal before applying the PARMA model. To normalize the data, the following transformation Y=f(x) are commonly used.

Logarithmic	$Y=\ln(X + a)$	3.1
Gamma	$Y=\text{Gamma}(X)$	3.2
Power	$Y= (X + a)^{b}$	3.3
Box-Cox	$Y=((X+a)^{b} -1) /b ,b\neq 0$	3.4

Where, Y is the normalized series, X is the original observed, a and b are transformation coefficient. Fig (3.3) shows the plot of original and transformed data for season 1. Transformation coefficient and the normality test value of all seasons for sample stations are shown on appendix A.1.

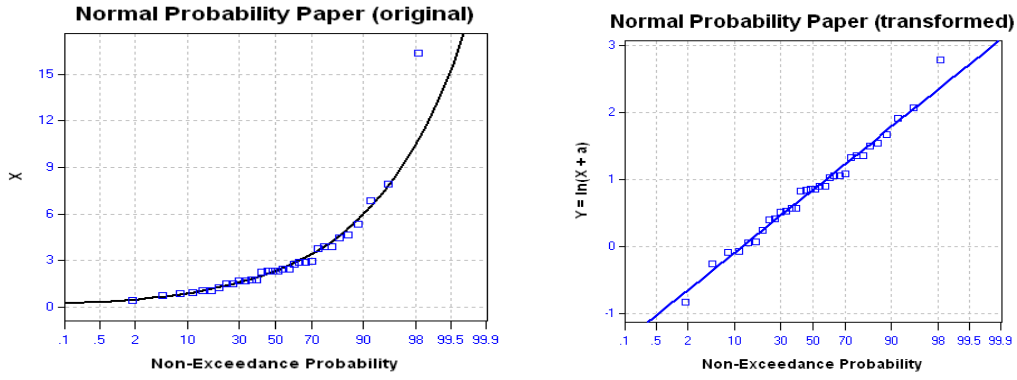


Figure 3.3: Plot of the transformed data on normal probability paper and test of normality for season one at Gilgel Abbay station.

3.2.3.4. Scaling and Standardization

Scaling of normally distributed data is required when different seasons have values that differ from each other by couple of orders of magnitude which can cause problems in parameter estimation which can happen when some of the historical time series are normally distributed and does not need to be transformed to normal while others do.

In this case some seasons do not need transformation for normality; for instance, season 8 of Muger near Chancho station where as other seasons for the same station do need to be transformed, thus in such conditions the scaling is needed. Scaling in this context is dividing all the time series which have not been transformed by Eqs. (3.1)-(3.4) by the standard deviation.

In addition, the normalized data are standardized by subtracting the mean and dividing by the standard deviation.

4. MODEL FITTING AND SIMULATION

4.1 PARMA Model Fitting

4.1.1 General

In this Chapter a detailed simulation study is presented, which was conducted to investigate the stochastic model which best fits the monthly streamflow of upper Blue Nile river basin. Here PARMA model fitting are presented first and then Thomas Fiering.

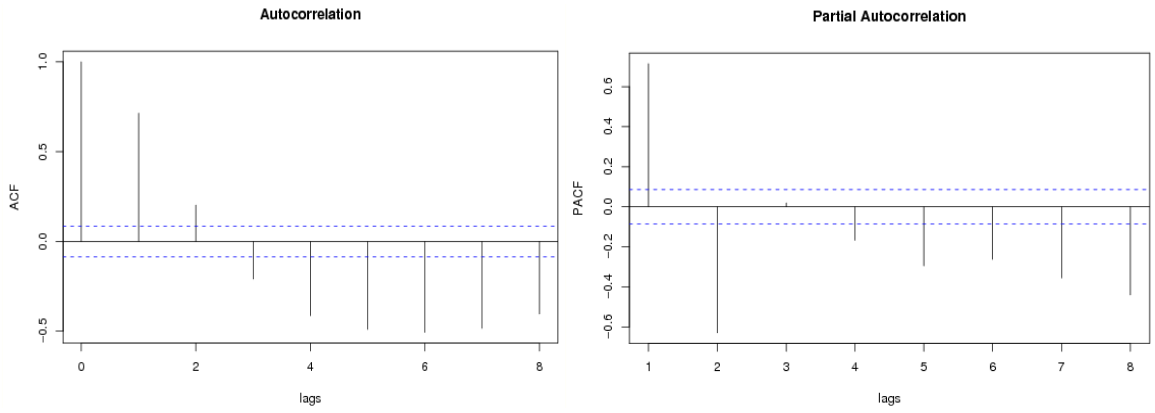
The main three phases of modeling which are used for stochastic modeling of hydrologic time series namely model identification, parameter estimation and diagnostic checking of the recommended PARMA model are presented first and finally Thomas Fiering model.

A SAMS version 2007 software was used to simulate the PARMA samples as well as to make all the necessary calculations. The series consists of average monthly streamflow (Mm^3) values for thirteen stations of Upper Blue Nile River basin as presented in the previous chapter from January to December, where season 1 corresponds to January and season 12 corresponds to December are used in this study.

4.1.2 Initial Model Identification

Initial identification methods are procedures applied to a set of data to indicate the kind of representational model that is worthy of further investigation and the specific aim is to gain some idea of the values of p and q needed in the general PARMA model and to obtain initial guesses of the parameters.

Seasonal patterns of time series can be examined via correlograms. Accordingly, ACF and PACF functions have been used for prior identification of the dependence structure in the data series which is used as identification of representative model.



(a) **(b)**

Figure 4.1: The sample ACF(fig. a) and PACF(fig. b) for the historical time series of Gilgel Abbay near Mearawi, showing the 95% confidence bounds $\pm 1.96/\sqrt{N}$.

The identification of the appropriate parametric time series model depends on the shape of the ACF and PACF. Failure of the ACF estimates other than the integer multiples of season (ω) to damp out suggests that nonseasonal differencing is needed to produce stationarity (Maidmant *et al* 1993). The sample PACF damps out at lags that are multiples of ω. This suggests the incorporation of a seasonal MA component into the model. The failure of the sample PACF to truncate at other lags may imply that a nonseasonal MA term is required (Fig. 41b). The sample ACF also reveals that the ACF truncates and is not significantly different from zero after lag 1 (Fig. 41a). Thus, a pure PARMA (p,q) the p and q value ranging from 0 to 1 are recommended in this section through ACF and PACF analysis for Gilgel Abbay near Merawi station. Similar identification is made for all stations included in this study.

The model recommended in this section as explained is simply to get an idea of the model orders p and q but the final decisions on the model specification will be made after parameter estimation and diagnostic checking.

4.1.3 Model Fitting Using Optimization

For fitting the periodic autoregressive integrated moving average (PARMA) model to the time-series estimation of model parameters is one of the three-stage procedures. The method of moment (MOM) and least square method (LS) methods are used in parameter estimation of PARMA(p,q) models (see also section 2.5).The parameters estimated for PARMA(1,0) using the method of moment Eq.(2.7-2.9) for station 113005 are shown in Table(4.1) and similar step is followed for the other stations.

Table 4.1 PAR parameters of PARMA (1, 0) model fitted to standardized data of Station 113005.

Months	Φ(1)	White_Noise_Variance:
January	0.612783	0.624497
February	0.591503	0.650124
March	0.706771	0.500474
April	0.544614	0.703395
May	0.450998	0.796601
June	0.616951	0.619371
July	0.661833	0.561977
August	0.361893	0.869034
September	0.602564	0.636917
October	0.45476	0.793194
November	0.613774	0.623281
December	0.663485	0.559788

The plot of parameters here indicates the seasonality (Periodicity) pattern of both white noise variance (WN) and autoregressive parameters (Φ) (Fig.5).

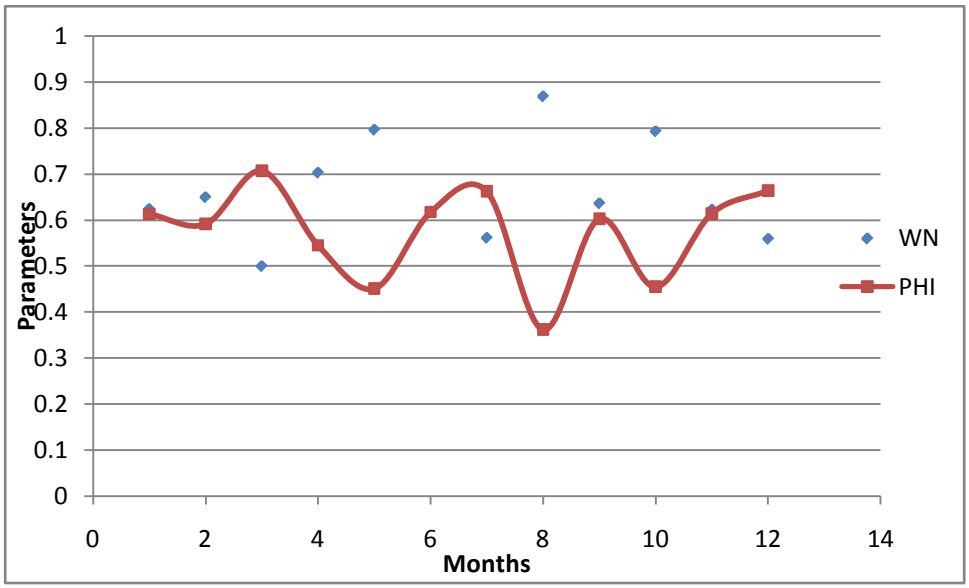


Figure 4.2: Plot of parameters of PARMA (1, 0) at station 113005.

4.1.4 Diagnostic checking

A basic tenet of model building is to keep the model as simple as possible but at the same time provide a good fit to the data being modeled (Hipel & McLeod 1994). Different automatic selection criteria are available for balancing the apparently contradictory goals of a good statistical fit and model simplicity. Among the automatic selection criteria used in model discrimination, Akaike information criterion (AICC) and Bayes information criterion (BIC) are common (Hipel & McLeod, 1994).

The diagnostic check stage determines whether residual are independent, homoscedastic and normally distributed.

Akaike Information Criterion Corrected (AICC)

In order to find the most appropriate model, the AIC is quite powerful (Burnham & Anderson 2002). The mathematical definition of AIC is given as

$$AICC = n \ln \hat{\sigma}^2(\varepsilon) + n + \frac{2(k+1)n}{n-k-2} \qquad 4.1$$

Where, n is the size of the sample used for fitting, k it the number of parameters excluding

constant terms (k = p + q) for the PARMA(p,q) model), and $\hat{\sigma}^2(\varepsilon)$ is the maximum likelihood estimate of the residual variance (biased).

The first part of the right-hand side of Eq. (4.1) explains how well the model fits and the second part accounts for the simplicity of the model. Use of the minimum AICC reinforces and complements the identification, estimation and diagnostic stages of model construction. The AICC statistic is efficient but not consistent and is good for small samples but tends to over fit for large samples and large *k* (Salas *et.el*). AICC tends to overestimate the order of autoregression. Thus, a Bayesian extension of the minimum Akaike criterion (BIC) is developed. One of the variants of BIC, Schwarz Bayesian criterion (SIC), is used here.

Schwarz Bayesian criterion (SIC)

The SIC has a structure quite similar to that of AICC.

$$SIC = n \ln \hat{\sigma}^2(\varepsilon) + n + k \ln n \qquad 4.2$$

Where n, k and $\hat{\sigma}^2(\varepsilon)$ are defined in the same way as AICC. The model that gives the minimum AICC and SIC is selected as parsimonious. Normality test and WN test of the residuals are performed through a plot of ACF of the residuals and statistical analysis.

Table 4.2 Goodness of fit of the candidate models to the monthly streamflow at Gilgel Abbay near Merawi station

Model	Autocorrelation	WN variance	AICC	SIC
PARMA(1,0)	all	0.869034	42.969	42.49
PARMA(1,1)	all	0.868979	45.266	46.294
PARMA(2,0)	all	0.868985	45.268	46.297
PARMA(2,1)	all	0.859808	47.203	49.623

The performance of the candidate models is compared using the white noise variance (WN variance), Akaike information criterion corrected (AICC) and Schwarz Bayesian criterion (SIC) (Table 4.2).The WN variance is the ratio of the sum of squares of errors (SSE) divided by the

number of observations.

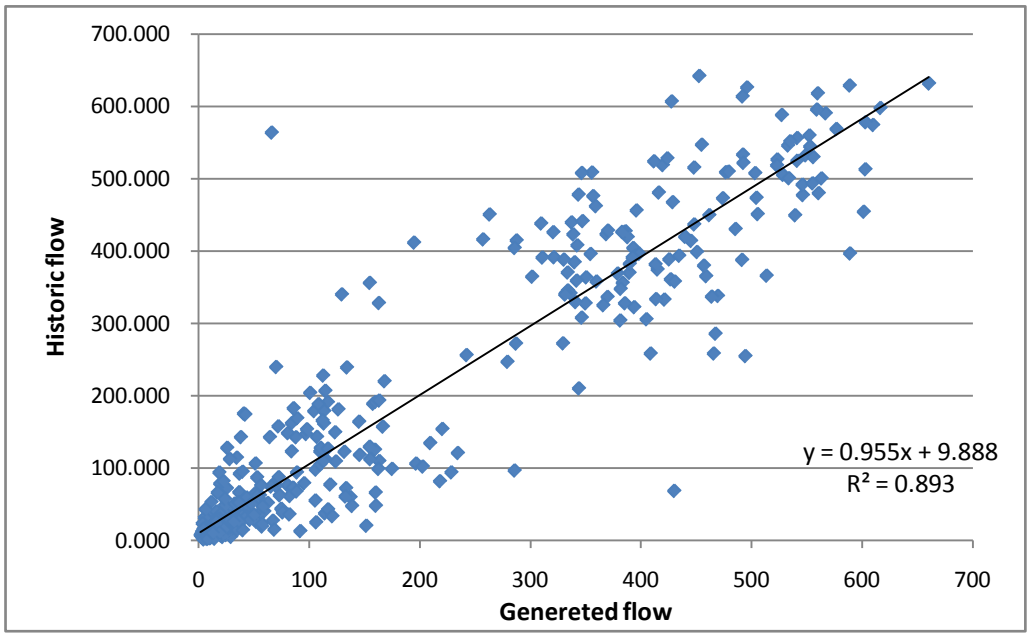


Figure 4.3: Correlation between Historic and Generated streamflow of Gilgel Abbay near Merawi station using best fitted PARMA (1, 0) model.

Figure 4.3 shows that the correlation between historic streamflow and generated streamflow is greater for low and high discharge values than for the median discharge values. From Table 4.2, it can be seen that as the order of PARMA increases the model fits poorly. Generally, as can be observed from the values of AICC, SIC and sample statistics, the model PARMA (1, 0) is the best among the candidate models, whereas the PARMA (2, 1) model performs poorly.

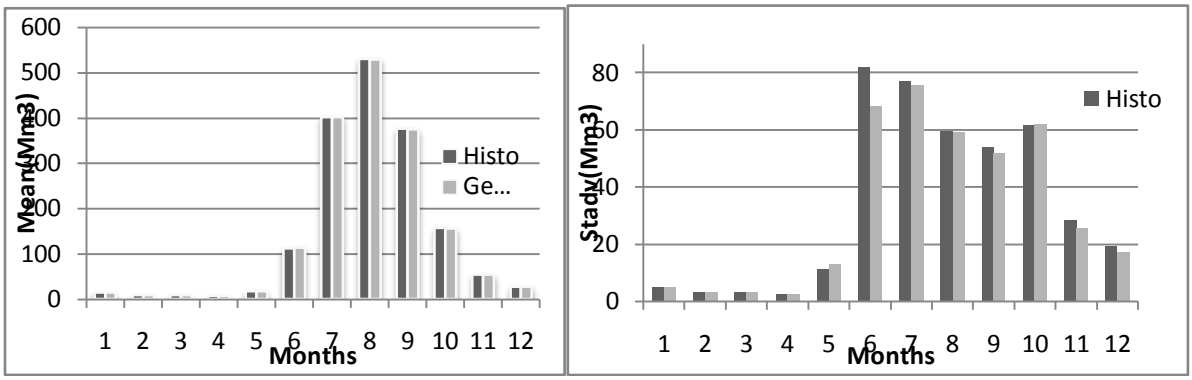


Figure 4.4: Comparison of generated and historic statistics of monthly streamflow at Gilgel Abbay near Merawi station.

The PARMA (1, 0) model fits well to the stream flow of Gilgel Abbay near Merawi station as can be seen from the plot of the simulated fit (Fig4.5) , the automatic selection criterion (Table 4.2) and the comparison of historic and generated basic statistic values (Fig4.4).
The PARMA(1,0) model can be written in form:

$$Y_{\upsilon,\tau} = \phi_{1,\tau} Y_{\upsilon,\tau-1} + \varepsilon_{\upsilon,\tau} \qquad 4.3$$

Where, $Y_{\nu,\tau}$ represents the streamflow process for year ν and season τ. For each season, τ, this process is normally distributed with mean zero and variance δ^2_τ (*Y*). The $\varepsilon_{\nu,\tau}$ is the uncorrelated noise term which for each season is normally distributed with mean zero and variance δ^2_τ (ε).

All the steps discussed above were similarly applied for the other stations of upper Blue Nile river basin included in this study.

Table: 4.3. Best fitted PARMA model for the stations in the study.

	Gauging stations	PARMA Model	
Station ID	Station Name	Model	R^2
111002	Gilgel_Abbay_Near_Merawi	PARMA(1,0)	0.893
111003	Koga_Near_Merawi	PARMA(1,0)	0.304
111005	Ribb_Near_Addis_Zemen	PARMA(1,0)	0.642
111006	Gumara_Near_Bahir_dar	PARMA(1,0)	0.632
111007	Megeche_Near_Azezo	PARMA(1,0)	0.563
112001	Abbay_Near_Kessie	PARMA(1,0)	0.61
112002	Muger_Near_Chancho	PARMA(1,1)	0.742
112003	Abbay_Near_Bahir_dar	PARMA(1,1)	0.331
113002	Fatto_Near_Guder	PARMA(1,0)	0.763
113005	Guder_At_Guder	PARMA(1,0)	0.798
113012	Gudla_Near_Dembecha	PARMA(1,0)	0.659
114001	Diddesa_Near_Arjo	PARMA(1,1)	0.628
116005	Main_Beles_Near_Metekel	PARMA(1,0)	0.57

The resulted best fit PARMA models for the stations with the Nash- Sutcliffe Coefficient (R^2) value are as in Table:(4.3). As it can be seen from the result the PARMA (p,q) model with lower values of autoregressive(p) and moving average (q) performs well for most of upper Blue Nile river basin stations considered.

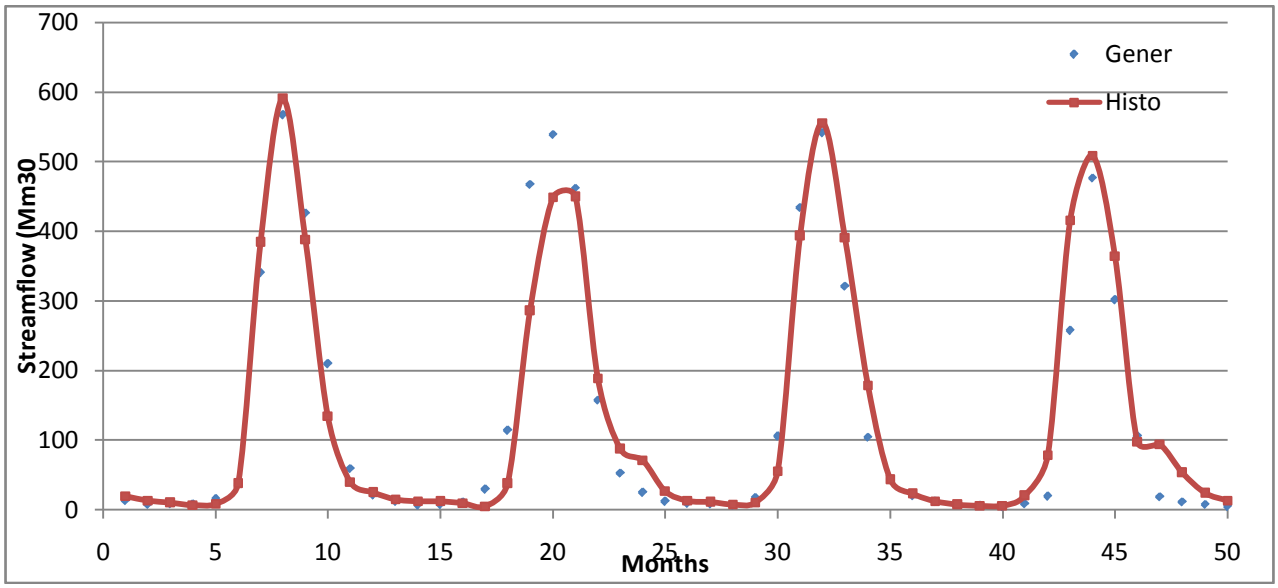


Figure 4.5: Fitted periodic autoregressive moving average (PARMA) (1, 0) model to the streamflow of Gilgel Abbay near Merawi station for some month's data.

4.1.5 Model verification

The Model verification was performed using stream flow data sets which are not used in the model calibration stage. Calibration and verification of the model is a key factor in reducing uncertainty and increasing user confidence in its predictive abilities, which makes the application of the model effective. Model verification results show simulated versus observed flow do have almost similar result as that of the model calibration. Simulated versus observed plot of Gilgel Abby near Merawi station shows this (Fig 4.4.) Verification plots for other station are included in the appendix section.

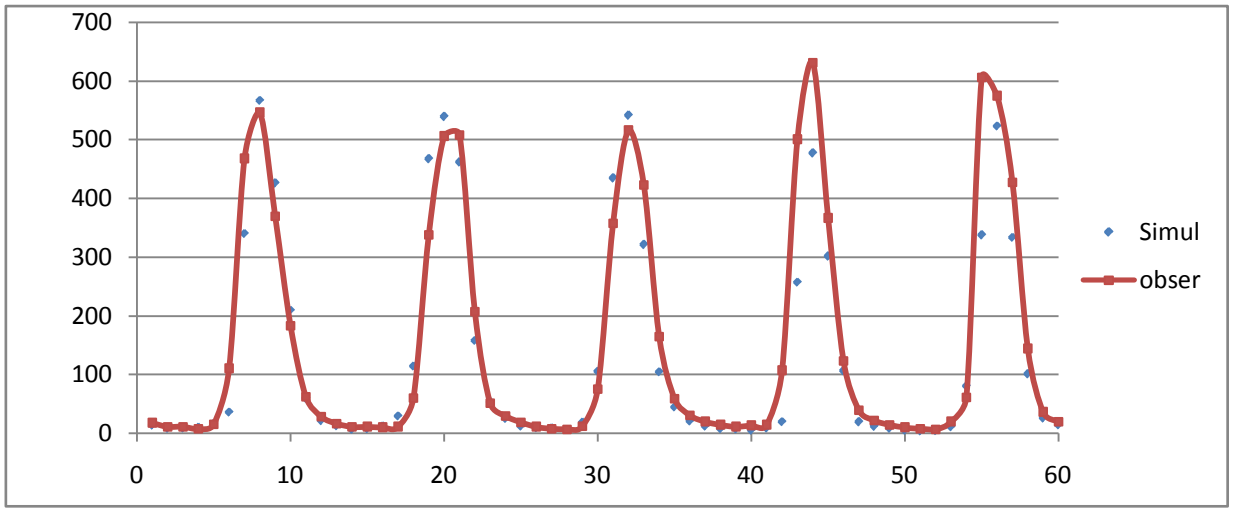


Figure 4.4: Verification plot for Gilgel_Abbay_Near_Merawi station

4.2 Thomas Fiering Model fitting

The monthly flow series are non-stationary and therefore complicated mathematical models are employed in their simulation. The first model that appeared in the hydrology literature for the generation of synthetic monthly flow sequences is that due to Thomas & Fiering (1962). Basically, this model is of a Markovian nature with periodic parameters, namely, the monthly means, standard deviations and the lag-zero crosscorrelations between successive months. In its simplest form the model consists of twelve regression equations, one for each month. The model implicitly allows for the non-stationarity observed in monthly flow data.

The Thomas–Fiering model is given by

$$X_{i,j} = \bar{X}_j + b_j (X_{i,j-1} - \bar{X}_{j-1}) + \varepsilon_t \sigma_j \sqrt{(1 - r_j^2)} \qquad 4.4$$

Where, $\bar{X}_j$ and $\bar{X}_{j-1}$ are the mean flows in seasons j and j-1 respectively. σ_j is the standard deviation of flow in the season j. r_j is the correlation between flows in the j^{th} and $(j-1)^{th}$ seasons. b_j is the regression coefficient of flow in the j^{th} season with flow in the $(j-1)^{th}$ season. ε_t is the normal random variate. $x_{i,j}$ is the flow generated in the j^{th} season of the i^{th} year, with j = 1, 2, ...,m and i = 1, 2, ... If a season is a month, then m = 12; for a weekly model, m = 52; for a daily model, m = 365. It may be noted that in the Thomas–Fiering model the flow in any season j is a sum of three terms. The first term is the mean flow in that season. The second term is the regressed component on the flow in the previous season. The third term is a random component to reflect the variance. This model accounts for the persistence only up to lag-1. Hence, the Thomas–Fiering model may be viewed as a non-stationary first-order AR model. This model can be applied directly in the form given in Equation (4.4), drawing the random variate ε_t from a normal distribution with zero mean and unit variance [N(0,1)] provided the flow in all seasons is normally distributed.

One of the prerequisites required for the generation of synthetic sequences is the convenience of the model to generate a given monthly flow sequence. Such a convenience can be checked by examining residuals of the model. If the model is suitable for the data available then the residuals calculated thereof must be independent.

Table: 4.4 Thomas-Fiering model parameters and results.

Station No.	Gauging Stations		Thomas Fiering model
	Station ID	Station Name	R^2
1	111002	Gilgel_Abbay_Near_Merawi	0.875
2	111003	Koga_Near_Merawi	0.288
3	111005	Ribb_Near_Addis_Zemen	0.672
4	111006	Gumara_Near_Bahir_dar	0.749
5	111007	Megeche_Near_Azezo	0.397
6	112001	Abbay_Near_Kessie	0.525
7	112002	Muger_Near_Chancho	0.688
8	112003	Abbay_Near_Bahir_dar	0.321
9	113002	Fatto_Near_Guder	0.765
10	113005	Guder_At_Guder	0.791
11	113012	Gudla_Near_Dembecha	0.71
12	114001	Diddesa_Near_Arjo	0.624
13	116005	Main_Beles_Near_Metekel	0.609

Here, the monthly streamflow records are simulated using Thomas Fiering model and characteristics of these simulated flow periods are derived. The mean, standard deviation and Nash - Sutcliffe Coefficient (R^2) of their values are compared with the appropriate that values for the historical critical period and the result is as shown on table 4.4.

4.3 Result and Discussions

A set of monthly streamflow records is generated and analyzed to give a set of synthetic critical periods. Some characteristics of these periods are derived, and the mean, standard deviation and Nash - Sutcliffe Coefficient (R^2) of their values are compared with the appropriate values for the historical critical period. This procedure is repeated for a number of different rivers by using both PARMA and Thomas Fiering model of generating techniques.

Table: 4.5 Comparisons of PARMA and Thomas-Fiering Models.

Gauging Stations			PARMA Model		TF model	
Station No	Station ID	Station Name	Model	R^2	Model	R^2
1	111002	Gilgel_Abbay_Near_Merawi	PARMA(1,0)	0.893	TF	0.875
2	111003	Koga_Near_Merawi	PARMA(1,0)	0.304	TF	0.288
3	111005	Ribb_Near_Addis_Zemen	PARMA(1,0)	0.642	TF	0.672
4	111006	Gumara_Near_Bahir_dar	PARMA(1,0)	0.632	TF	0.749
5	111007	Megeche_Near_Azezo	PARMA(1,0)	0.563	TF	0.397
6	112001	Abbay_Near_Kessie	PARMA(1,0)	0.61	TF	0.525
7	112002	Muger_Near_Chancho	PARMA(1,1)	0.742	TF	0.688
8	112003	Abbay_Near_Bahir_dar	PARMA(1,1)	0.331	TF	0.396
9	113002	Fatto_Near_Guder	PARMA(1,0)	0.763	TF	0.765
10	113005	Guder_At_Guder	PARMA(1,0)	0.798	TF	0.791
11	113012	Gudla_Near_Dembecha	PARMA(1,0)	0.659	TF	0.71
12	114001	Diddesa_Near_Arjo	PARMA(1,1)	0.628	TF	0.624
13	116005	Main_Beles_Near_Metekel	PARMA(1,0)	0.57	TF	0.609

Generally the best fit PARMA model performed better than that of Thomas Fiering model for the generation of monthly streamflow of Upper Blue Nile river basin but both models do not perform as expected for some stations as it can be seen from table 4.5. Askew *et al*. (1971) has shown that stochastic generation of streamflow can be influenced by the size of the reservoir and relative aridity of watershed considered .In this case both PARMA and Thomas –Fiering models do not perform at Abbay near Bahirdar station which is near the outlet of Lake Tana Reason for this possibly could be due to the significant size of the reservoir.

5. Hydrological Drought

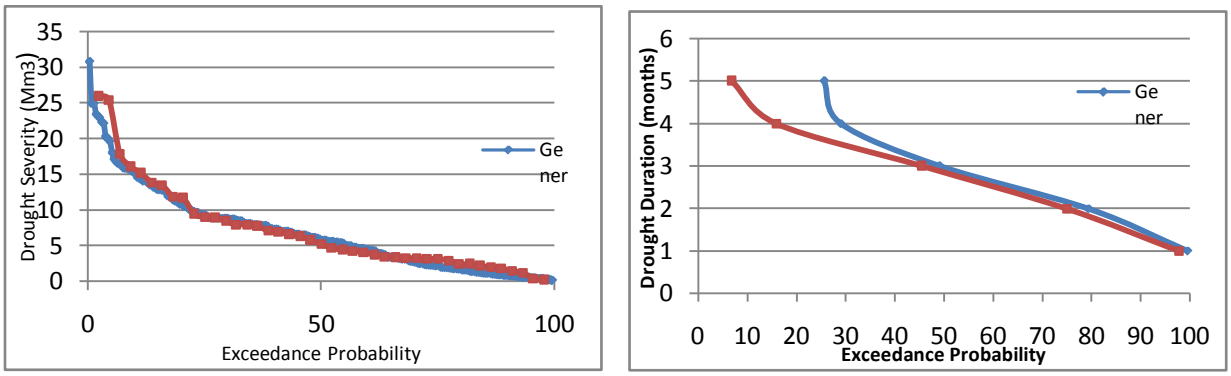
5.1 Frequency Analysis of Drought Sequence

One of the objectives of this study is to apply most of the frequency analysis to the derived drought parameters and to reach conclusions about the best-fit models based on various goodness-of fit tests. Many probability distribution functions that have been suggested by various workers for extreme-value fitting are considered.

The proposed methodology for drought simulation is illustrated by means of a case study of gauging stations with similar stochastic behavior established using hierarchical clustering as explained in the Appendix D. Thus, one at upstream portion of the basin, station 111002 (Gilgel Abbay near Merawi), and another at the downstream portion, station 113005(Guder at Guder) are used . The numbers of droughts in the historical streamflow series identified for both stations are 9 and 43 at Gilgel Abbay and 12 and 48 at Guder respectively in yearly/monthly bases.

Identification of drought parameters was done on the basis of the generated long terme series . As explained in section 4.3 several series up to a period of 500 years were generated. From the generated series there were respectively 119 and 233droughts identified Gilgel Abbay, and 104 and 236 at Guder in yearly/monthly bases when the series were subjected to drought analysis.

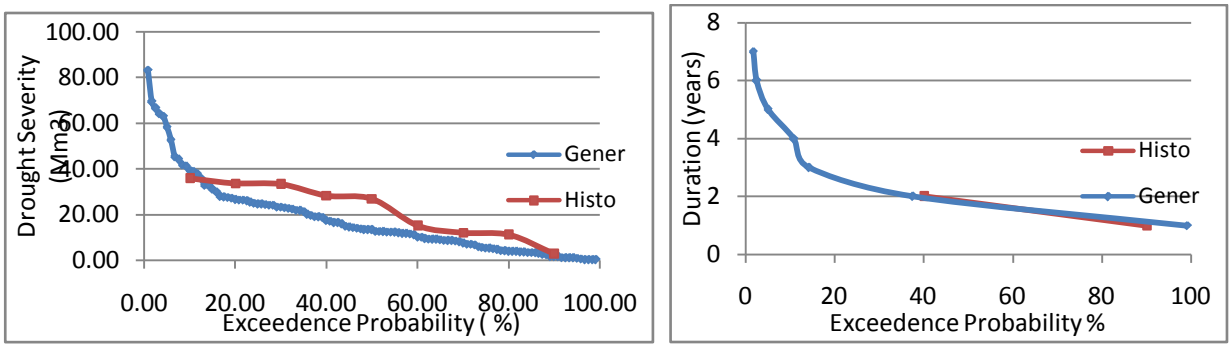
Before fitting any distribution for drought parameters, it is necessary to check whether these series are independent. To this end, autocorrelation coefficients at various lags were computed for the drought parameters as found in the generated streamflow series, and the test of significance at the 5% probability level was carried out to see whether the sample correlograms belong to the independent cases. The correlograms of drought parameters were found to be within the 95% confidence limits. Thus, it can be concluded that the drought parameters derived from the generated series, namely drought duration and drought severity, are independent random variables.



(a) (b)

Figure 5.1 Comparison of historic and simulated distribution functions :(a)monthly streamflow drought severity; (b) monthly streamflow drought duration at Gilgel Abbay.

The comparison of the probability distributions of streamflow drought parameters derived from the historic and generated flow series are shown in Figure 5.2. Similar comparisons of probability distributions of streamflow drought parameters derived from the historic and generated streamflow series are made for Guder station. It can be seen from Figure 5.2 that the probability distributions of streamflow drought parameters derived from the historic and generated series closely match each other, but the return period of more extreme droughts can only be obtained from the probability distributions of drought parameters derived from the generated series.



(a) (b)

Figure 5.2 Comparison of historic and simulated distribution functions :(a)yearly streamflow drought severity; (b) yearly streamflow drought duration at Gilgel Abbay.

However, for estimating the reliable return period of droughts events, the probability distribution of drought parameters derived from the generated series needs to be used Table 5.1 shows a comparison between the longest drought durations and greatest drought severities derived from the historic and generated streamflow series at both stations.

Table 5.1 Comparison of extreme droughts derived from the historic and generated series

Station	Gilgel Abbay		Station	Guder	
Drought parameter	Historic	Generated	Drought parameter	Historic	Generated
Longest streamflow drought duration (years)	2	7	Longest streamflow drought duration (years)	3	5
Greatest streamflow drought severity (Mm3)	35.79	83.25	Greatest streamflow drought severity (Mm3)	7.59	18.27
Longest streamflow drought duration (months)	5	5	Longest streamflow drought duration (months)	6	6
Greatest streamflow drought severity (Mm3)	25.956	30.806	Greatest streamflow drought severity (Mm3)	4.55	5.408

As expected, it was found that the generated longest drought duration is longer than that of the historic longest drought for yearly drought but in case of monthly drought the longest drought is the same for historic and generated series. In case of severity the generated greatest drought severity is larger than that of the historic greatest drought in all cases. Table 5.2 provides a comparison between the mean values of the drought magnitudes computed from historic and generated flow series.

Table 5.2 Comparison of the historic and generated drought magnitudes

Station	Gilgel Abbay		Station	Guder	
Event	Historic	Genereted	Event	Historic	Genereted
Streamflow (Mm3 $year^{-1}$)	15.32	10.36	Streamflow (Mm3 $year^{-1}$)	3.23	3.04
Streamflow (Mm3 $month^{-1}$)	2.89	2.7	Streamflow (Mm3 $month^{-1}$)	0.48	0.557

These drought magnitudes, together with the fitted distribution functions, will be a valuable tool for prediction of drought parameters.

5.2 Distribution function for drought parameters

5.2.1 Drought severity

About twenty different theoretical distribution functions which are commonly used for extreme values were tried to fit the probability distributions of streamflow drought severities derived from the generated series. The best-fit distribution functions were selected based on the Kolmogorov–Smirnov (K–S) goodness-of-fit test, chi-square test and Anderson-Darling teste. Results of the test for streamflow drought severity are indicated on Appendix D-1. Of the several theoretical distribution functions, the generalized Pareto distribution were the best fit for the streamflow severity at Gilgel Abbay both for yearly/monthly severities. The goodness of fit details for the best fit general Pareto distribution is indicated on the Table 5.4.

Table 5.3 Goodness of fit details of General Pareto distribution for drought severity.

Kolmogorov-Smirnov					
Sample Size					233
Statistic					0.04427
P-Value					0.73385
Rank					2
∝	0.2	0.1	0.05	0.02	0.01
Critical Value	0.07029	0.08012	0.08897	0.09945	0.10672
Reject?	No	No	No	No	No
Anderson-Darling					
Sample Size					233
Statistic					0.62286
Rank					1
∝	0.2	0.1	0.05	0.02	0.01
Critical Value	1.3749	1.9286	2.5018	3.2892	3.9074
Reject?	No	No	No	No	No
Chi-Squared					
Deg. of freedom					7
Statistic					10.9
P-Value					0.14306
Rank					1
∝	0.2	0.1	0.05	0.02	0.01
Critical Value	9.8032	12.017	14.067	16.622	18.475
Reject?	Yes	No	No	No	No

The fitted distribution function to the yearly streamflow drought severity is given by

$$F(S) = 1 - \left[1 - 0.095\,\frac{(s-0.0495)}{19.499}\right]^{1/0.095} \qquad 5.1$$

Where, S is the drought severity. The distribution function fitted to the monthly drought severity is given by

$$F(S) = 1 - \left[1 - 0.2466 \frac{(s+0.05168)}{8.6379}\right]^{1/0.2.466} \qquad 5.2$$

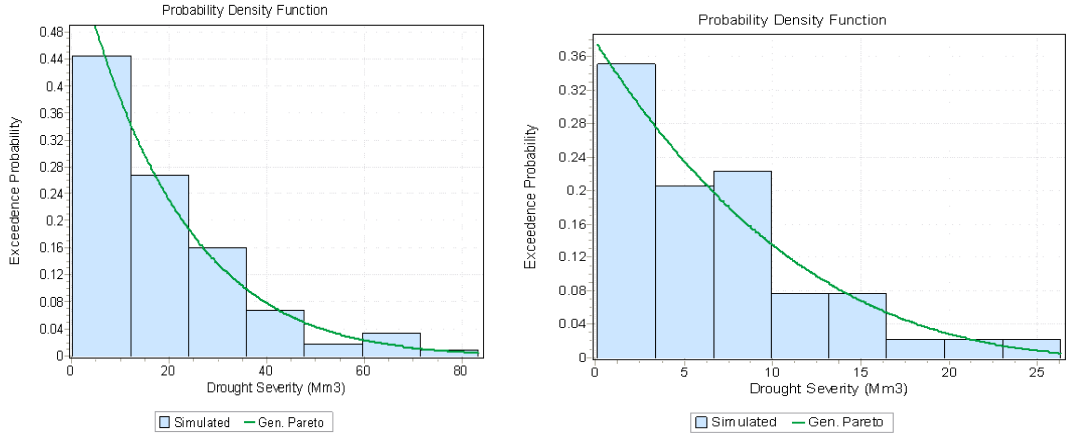


Figure: 5.3. Comparison of stimulated and theoretical probability distribution :(left) yearly streamflow drought severity; (right) monthly drought severity.

The comparisons between the theoretical probability distributions and probability distributions derived from the generated streamflow series for the drought severity are shown in Figure 5.3, which depicts the closeness between the simulated and theoretical probability distributions.

5.2.2 Drought duration

In a similar way to that of drought severity, different selected theoretical distribution functions were tried to fit the probability distributions of streamflow drought durations derived from the generated streamflow series. The best-fit distribution function was selected based on the K–S , chi-square test and Anderson-Darling tests at the 5% significance level. Of the several distributions, it was found that a log-logistic distribution fits well to the probability distribution of yearly streamflow drought duration and logistic distribution fits well to the monthly drought duration derived from the generated series. The cumulative function of the fitted log logistic distribution to yearly drought duration is given by

$$F(D) = \frac{1}{1+(\frac{d}{2.776})^{-1.428}} \qquad 5.3$$

Where, D is the drought duration series and d is an element of the drought duration series. It was found that the logistic distribution fits well to the probability distribution of monthly drought duration derived from the generated series, and the resulting distribution function is given by

$$F(D) = \frac{1}{1+e^{-(d-2.55)/0.64}} \qquad 5.4$$

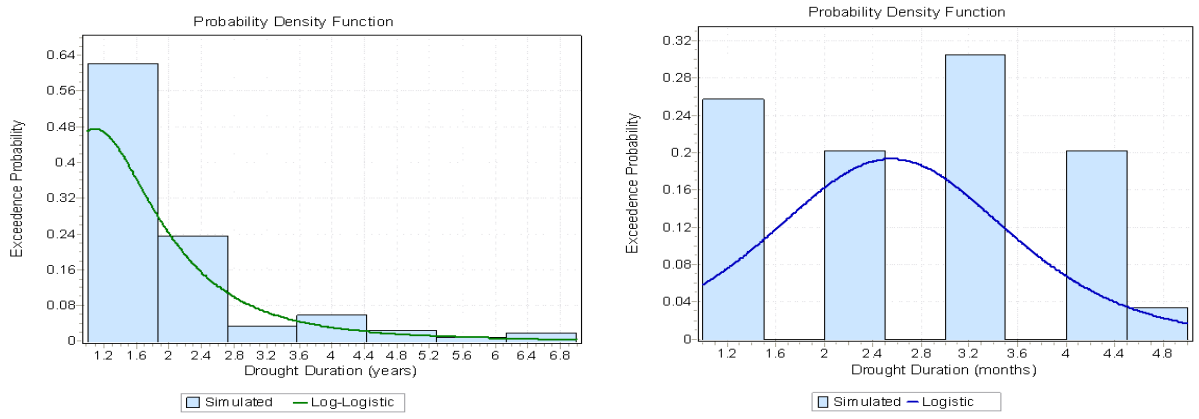


Figure: 5.4. Comparison of stimulated and theoretical probability distribution :(left) yearly streamflow drought duration; (right) monthly streamflow drought duration at Gilgel Abby near Merawi station

Figure 5.4 shows the comparison between the fitted probability distributions and probability distributions obtained from the generated streamflow series, from which it can be seen that the fitted probability distribution fits not well as in the severity case to the simulated probability distribution.

5.3 Drought Duration-Severity Relationship

Cross-correlation and regression analysis were carried out to find the relationship between drought parameters, namely drought duration and drought severity, derived from the generated series. The correlation coefficient between yearly streamflow drought severity and its duration is 0.6, whereas the cross-correlation between monthly drought severity and its duration is 0.62. The relationship between streamflow drought severity and streamflow drought duration is linear for both cases and the regression equation is S=16.67D-10.82, and S=6.17D-8.88 respectively for yearly/monthly. Where, S is the drought severity and D is the drought duration. For

example,Figure: 5.5 shows the correlation between drought parameters from which it can be seen that thought there is no such a strong cross-correlation between the parameters the correlation is persistent for yearly and monthly parameters for the Gilgel Abbay near Merawi station.

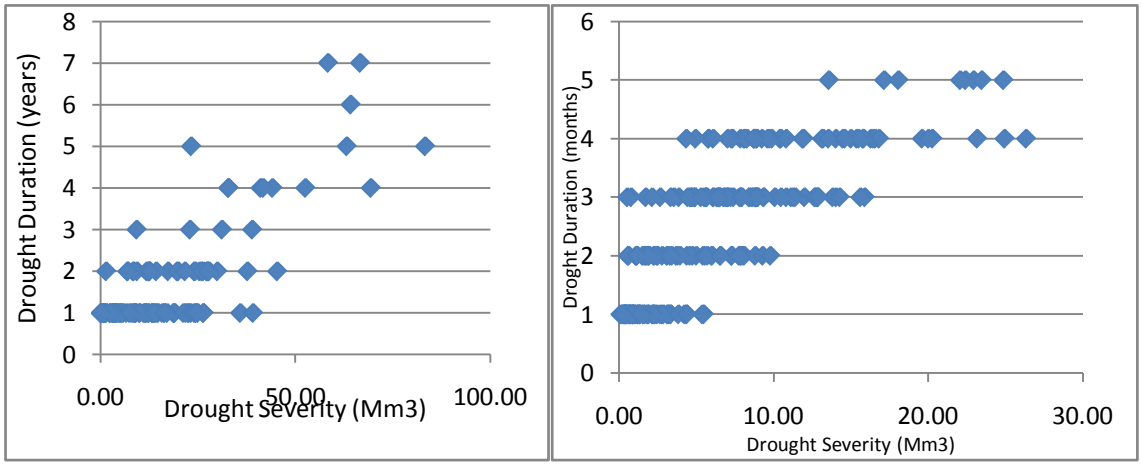


Figure: 5.5. Correlation between drought parameters: (left) yearly streamflow drought severity and its duration; (right) monthly drought severity and its duration at Gilgel Abbay station.

5.4 Joint Distribution Function of Drought Parameters

Since drought is a two-dimensional process, the joint distribution of both severity and duration is very important. Figure 5.8 shows that the correlation between drought severity and drought duration is greater for less severe yearly droughts than for extreme droughts. To substantiate this, ranking the drought severity of the 20 greatest droughts and the corresponding rank numbers of drought duration have been worked out, from which it was found that the rank numbers of drought severity for the 20 greatest droughts are not in conformity with the rank numbers of drought durations. Thus, drought duration and drought severity for yearly extreme droughts can be logically assumed to be independent of each other. Hence, the joint distribution function of yearly streamflow drought severity and its duration can be worked out and is given by

$$F(S,D)=\exp\left\{-e^{-\left(\frac{S-39.8}{12.16}\right)}\right\}\left(1-\left[1-\frac{-0.77}{4.89}(d-0.94)\right]^{\frac{1}{-0.77}}\right) \qquad 5.5$$

Where, S is the drought severity (Mm^3) and D is the drought duration (years)

It is to be noted here that these joint distribution functions are applicable only to extreme yearly droughts.

5.5 Application of the Results

Obtaining drought properties is important for planning and management of water resources system. For example, the design of water supply capacity of a given city may be based on meeting water demands during a critical drought that may occur in a specified planning horizon. Moreover, the estimation of return periods associated to severe droughts can provide useful information in order to improve water systems management under drought condition.

The following two case examples may illustrate the applications of the results obtained from the drought analysis of this study.

Case 1

Determine (a) the longest streamflow drought duration and its severity and (b) the largest streamflow drought severity and its duration expected within a return period of 5 years (60 months) in the Gilgel Abbay catchment.

Solution: As the return period (60 months) is included in both the historic flow record and the generated streamflow series length, the behavior of the drought parameters established either from the historic flow series or from the generated flow series can be used to determine the required estimates. If both are used, then a comparison between the estimates resulting from two types of flow series can be made. The steps to be followed are explained below.

- ✓ Determination of exceedance probability. The periodicity of the hydrologic cycle is 12 months and the expected number of hydrologic cycles within a period of 60 months is five. The exceedance probability corresponding to this return period p = 1/5 =0.2.

- ✓ Longest drought duration and its severity. From Figure5.4, corresponding to a exceedance probability 0.2, the longest drought duration resulting from the historic flow series is 3.5 months. The longest duration from the generated flow series is also 3.5 months. From Table 5.3, the expected drought magnitude resulting from historic flow series is 2.89 Mm^3 $month^{-1}$. The expected drought magnitude resulting from the generated flow series is equal to 2.7 Mm^3 $month^{-1}$. Thus, the expected drought severity during the longest drought duration from historic series is 3.5x2.89=10.115 Mm^3. The

expected drought severity during the longest drought duration from generated series is 3.5x2.7=9.45 Mm^3. Compared with 10.115 Mm^3, the value of 9.45 Mm^3 must be more reliable as this estimate resulted from a large population of drought events.

- ✓ Greatest drought severity and its duration. From Figure 5.4, corresponding to a exceedance probability of 0.2, the greatest drought severity resulting from historic flow series is 11.67 Mm^3. The greatest drought severity resulting from generated flow series is 10.82 Mm^3. Therefore, the expected drought duration corresponding to a severity of 11.67 Mm^3 resulting from historic flow series is 11.67/2.89, which is equal to 4.03 months. Similarly, the expected drought duration corresponding to a severity of 10.82 Mm^3 from generated flow series is 10.82/2.7, which is equal to 4.322 months. The latter result may be considered more reliable, as it results from a large population of drought events.

Case 2

Determine the probable longest streamflow drought duration and the greatest streamflow drought severity within a period of 100 years.

Solution: These estimates can be made using the drought simulation results derived in this chapter of the present study. As a first step, it is necessary to determine the exceedance probability corresponding to the given return period. The required estimates can be made by making use of the simulated probabilistic relationships or the fitted theoretical probability distribution function for the drought parameters of interest derived from generated flow series. As the return period is 1200 months, which exceeds the historical flow records (540 months), the required estimates can be obtained by using the simulated results.

- ✓ Determination of exceedance probability. The expected number of yearly cycles is 100. The exceedance probability is 1/100 = 0.01.
- ✓ Determination of longest drought duration. From Equation (5.4), corresponding to a exceedance probability of 0.01, the probable longest drought duration is approximately 1.27 months.
- ✓ Determination of the greatest drought severity. The anticipated greatest severity corresponding to a exceedance probability 0.01 is obtained for Equation (5.2) and is given by 7.568 Mm3.

6. CONCLUSION AND RECOMMENDATIONS

Generation of synthetic river flow data is important in planning, design and operation of water resources systems. River flow series usually have periodical stationarity; that is, their mean and covariance functions are periodic with respect to time. The common procedure in modeling such periodic river flow series is first to standardize or filter the series and then fit an appropriate stationary stochastic model to the reduced series. However, standardizing or filtering most river flow series may not yield stationary residuals due to periodic autocorrelations.

In the first part of this study, performance of two stochastic models including PARMA and Thomas-Fiering approaches was focused on. These models were applied to thirteen station's monthly streamflow sequences from upper Blue Nile basin.

Periodic autoregressive moving average (PARMA) models provide a powerful tool for the modeling of periodic hydrologic series in general and river flow series in particular. PARMA models are extensions of commonly used ARMA models that allow parameters to depend on season. The PARMA modeling procedure involves iterative steps of model identification, parameter estimation, model diagnosis and fitting the residuals (noise) with a probability distribution function (pdf).

Thomas-Fiering model is basically, of a Markovian nature model with periodic parameters, namely, the monthly means, standard deviations and the lag-zero crosscorrelations between successive months. In its simplest form the model consists of twelve regression equations, one for each month. The model implicitly allows for the non-stationarity observed in monthly flow data.

In the first part of this thesis, after the stochastic simulation of thirteen gauging station monthly streamflow data of upper Blue Nile river basin with the two periodic models the following conclusions were arrived at.

- The best fit PARMA (p,q) model gives better result than the Thomas-Fiering model for the stochastic simulation of monthly stremflow for the upper Blue Nile basin data considered in this study.
- The low order PARMA (p,q) models gave good results for most of the gauging station's data as expected.

- Both models did not give good result for some of gauging stations.

The main concern of the second part of this thesis is on the hydrologic drought analysis. simulation of hydrological drought is generally needed for reservoir sizing, to determine the risk of failure of water supply for irrigation systems, to determine the risk of failure of capacities of hydroelectric systems, to plan studies of future reservoir operations, to plan capacity expansion of water supply systems and similar applications. For the simulation study here first, all the gauging stations considered in the study are grouped hierarchically into groups of similar stochastic behavior and then from the formed fusible groups one at the upstream (station 111002) and one ate downstream (station113005) were considered for the drought analysis. The best fit periodic stochastic models selected in the first part of the thesis are used to generate long series of monthly streamflow data and drought analysis are carried on both stations in monthly and yearly basis.

Based on the simulation study on droughts, the following conclusions are arrived at

- The best fit PARMA (p,q) model for generation of streamflow series was found to be efficient in preserving the drought characteristics and preserving the statistics of the historical series.
- A reliable distribution of drought parameters can be obtained from samples of synthetically generated data.
- The drought parameters, namely drought duration and drought severity, are independent random variables, but they do have certain dependence on each other.
- There drought components at both stations are more or less the same but, the downstream portion (station 113005) is a little bit higher than the upstream portion (station 111002).
- A General Pareto distribution fits the best, both to the distribution functions of simulated yearly streamflow drought severity and monthly streamflow drought severity series at Gilgel Abbay near Merawi station.
- Log Logistic distribution fits well to the distribution function of simulated yearly streamflow drought duration series, whereas Logistic distribution fits well to the distribution function of simulated monthly streamflow drought duration series at Gilgel Abbay.

The derived distributions of drought parameters are useful in predicting the longest drought duration with its severity and largest severity with its drought duration for the given period of analysis. Also the derived distributions of drought parameters are useful in predicting the longest drought duration with its severity and largest severity with its drought duration for the given period of analysis.

The stochastic simulation of streamflow with periodic stochastic models and drought simulation methodology used in this thesis is found to be an efficient tool for streamflow time series generation and in identifying the drought parameters thus, could easily be adapted to any other similar case studies. The results presented in this thesis need to be modified for catchments in other regions or catchments in other climatic conditions. However, the general methodology used here in this thesis can be easily applied to other catchments, both for hydrological time series generation and hydrological drought analysis.

REFERENCE

Askew AJ, Yeh WWG, Hall WA. 1971. A comparative study of critical drought simulation. *Water Resources Research* **7**: 52–62.

Barnes, F.B. (1954), Storage required for a city water supply. *J. Inst. Eng., Australia, 26*, 198–203.

Conway., D. (2000). "The Climate and Hydrology of the Upper Blue Nile River." The Geographical Journal 166(1): 49-62.

Conway D. and Hulme (1993) "Recent fluctuations in precipitation and runoff over the Nile sub basins and their impact on main Nile discharge" Climate change 25: 125-151.

Conway D., (1997). " A water balance model of the Upper Blue Nile in Ethiopia" Hydrological sciences Journal 42(2)

Chung, C. and J.D. Salas (2000), Drought occurrence probabilities and risks of dependent hydrologic processes. *Journal of Hydrologic Engineering, 5* (3), 259–268.

Dracup, J.A, K.S. Lee and E.G. Paulson (1980a), On the statistical characteristics of drought events. *Water Resources Research, 16* (2), 289–296.

Filliben, J.J., 1975. The probability plot correlation coefficient test for normality. Technometrics, 17(1):111–117.

French Engineering Consultants and ISL Consulting Engineers (1999). Abbay River Basin Integrated Development Master Plan Project, Main report Volume 1.

Glady˘sev, E.G. (1961), Periodically correlated random sequences, *Sov. Math., 2* (2),385– 388.

Hazen, A. (1914), Storage to be provided in impounding reservoirs for manucipal water supply. *Trans. Amer. Soc. Civil Eng., 77*, 1539–1669.

Hipel, K. and McLeod, A.I. (1994) Time Series Modelling of Water Resources and Environmental Systems. Elsevier Science B.V., Amsterdam, the Netherlands.

Hurst, H. E., Long-term storage capacity of reservoirs, *Trans. Am. Soc. Civ. Eng.*, *116*, 770–799, 1951.

Loucks, D.P., J.R. Stedinger, and D.A. Haith (1981), *Water Resources Systems Planning and Analysis*. Prentice-Hall, Englewood Cliffs, N.J.

Maidment, D.R. (1993), *Handbook of Hydrology* . New York: McGraw-Hill.

Pagano, M. (1978), On periodic and multiple autoregressions. *Annals of Statistics, 6* (6), 1310–1317.

Panu, U.S. and Sharma, T.C. (2002) Challenges in Drought Research: Some Perspectives and Future Directions. Hydrological Sci. J., 47, 19–30.

Salas, J.D, J.T.B Obeysekera and R.A. Smith (1981), Identification of streamflow stochastic models. *ASCE Journal of the Hydraulics Division, 107* (7), 853–866.

Salas, J.D, G.Q. Tabios III and P. Bartolini (1985), Approaches to multivariate modeling of water resources time series. *Water Resources Bulletin, 21*, 683–708.

Salas,J.D.,1993:Analysis and Modeling of Hydrologic Time Series. In Handbook of Hydrology, .R. Maidment Editor ,McGraw Hill Inc, New York.

Sveinsson, O. G. B. J. D. Salas, W. L. Lane, and D. K. Frevert ,(2007): Stochastic Analysis, Modeling, and Simulation (SAMS) Version 2007. User manual Technical manual.

Tallaksen, L.M. and Hisdal, H. (1999) Methods for regional classification of streamflow drought series: the EOF method and L-moments. Technical Report to the ARIDE Project No. 2. University of Oslo, Norway.

Tao, P.C. and J.W. Delleur (1976),Seasonal and nonseasonal ARMA models. *ASCE Journal of the Hydraulic Division,102* (HY10), 1541–1559.

Tiao, G.C. and M.R. Grupe (1980), Hidden periodic autoregressive moving average models in time series data. *Biometrika 67*, 365-373.

Vecchia , A.V. (1985a), Periodic autoregressive-moving average (PARMA) modeling with applications to water resources. *Water Resources Bulletin 21,* 721–730.

Yevjevich, V. (1967), An objective approach to definitions and investigations of continental droughts. Hydrology paper 23, Colorado State University, Fort Collins, Colorado.

Yevjevich, V. (1983) Methods for determining statistical properties of droughts. In:

Yevjevich, V., da Cunha, L., Vlachos, E. eds. *Coping with droughts*. Colorado, Water Resources Publications. 22-43.
Zelenhastic, E., and A. Salvai (1987), A method of streamflow drought analysis. *Water Resources Research, 23* (1), 156–168.

Appendix A

Table A.1 Test of normality and transformation coefficients for sample stations.

Skewness Test of Normality: Two Sided Test Statistic at 10% Significance Level = 0.6443
Filliben Test of Normality: One Sided Test Statistic at 10% Significance Level = 0.9718

(a) Test of normality and transformation coefficients for sample stations Gilgel Abbay near Merawi

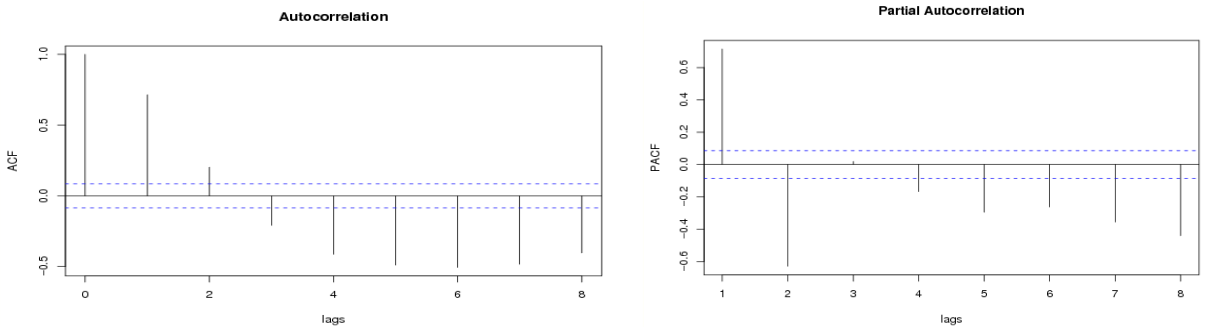
Seas	Type	Coeff		Skewness Test		Filliben Test	
Num	Trans	a	b	Result		Result	
1	Log	1.8266	1.8266	-0.0771	accept	0.997	accept
2	None	3.1318	3.1318	0.1078	accept	0.9921	accept
3	None	3.1176	3.1176	0.4003	accept	0.9929	accept
4	None	2.5518	2.5518	0.5781	REJECT	0.9795	accept
5	Log	1	1	-0.3447	accept	0.9923	accept
6	Log	0.01	0.01	0.4703	accept	0.9786	accept
7	None	77.0048	77.0048	0.0357	accept	0.9873	accept
8	None	59.261	59.261	-0.0391	accept	0.9934	accept
9	Log	1	1	0.309	accept	0.9863	accept
10	Log	1	1	-0.0329	accept	0.9952	accept
11	Log	1	1	0.4248	accept	0.9877	accept
12	Log	-9.5	-9.5	-0.2063	accept	0.9686	REJECT

(b) Test of normality and transformation coefficients for sample stations Guder at Guder.

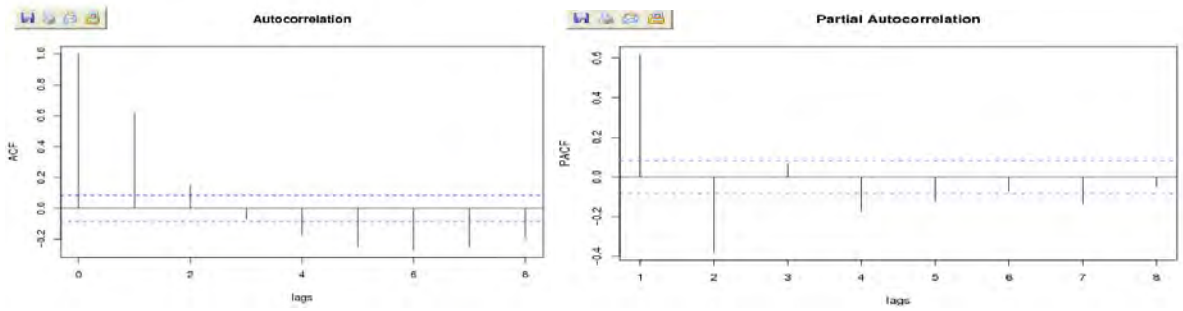
Seas	Type	Coeff		Skewness Test		Filliben Test	
Num	Trans	a	b	Result		Result	
1	None	0	0	0.0687	accept	0.9906	accept
2	Log	0	0	0.366	accept	0.9706	REJECT
3	Log	-0.1999	-0.1999	0.3862	accept	0.9535	REJECT
4	Log	0.5496	0.5496	0.0166	accept	0.9901	accept
5	Log	-0.8091	-0.8091	0.1332	accept	0.9905	accept
6	Log	0	0	0.0116	accept	0.9902	accept
7	Log	-1	-1	-0.6164	REJECT	0.9808	accept
8	Log	-80.4384	-80.4384	-0.1919	accept	0.9906	accept
9	None	0	0	0.0503	accept	0.9944	accept
10	Log	0	0	-0.1671	accept	0.9919	accept
11	Log	-1.1589	-1.1589	-0.4936	accept	0.9489	REJECT
12	Log	-0.9771	-0.9771	0.2827	accept	0.9866	accept

Appendix B

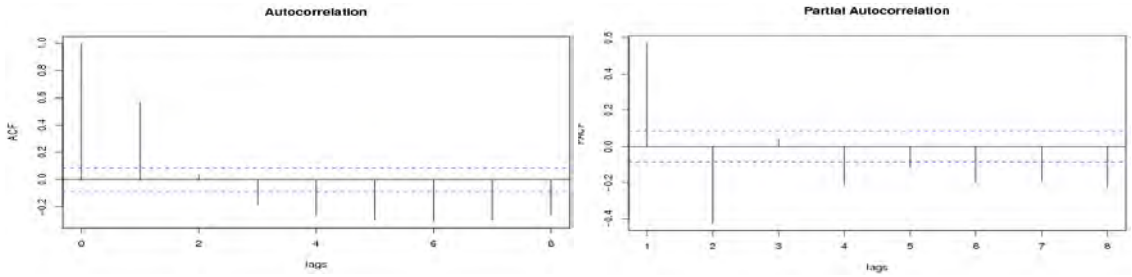
The sample ACF(left) and PACF(right) for the historical time series.



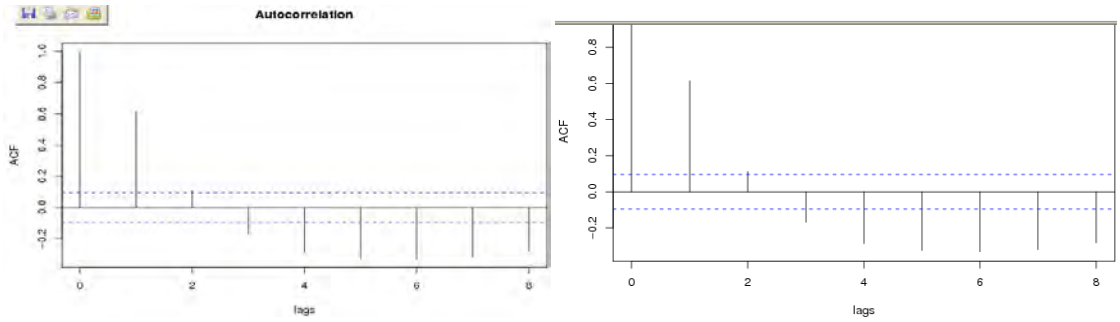
Gilgel_Abbay_Near_Merawi



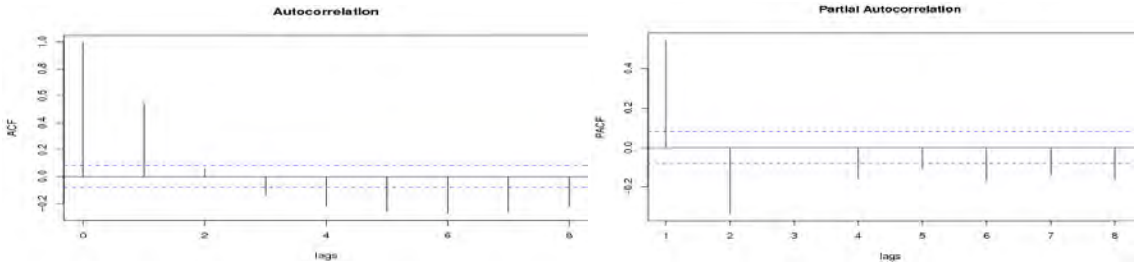
Koga_Near_Merawi



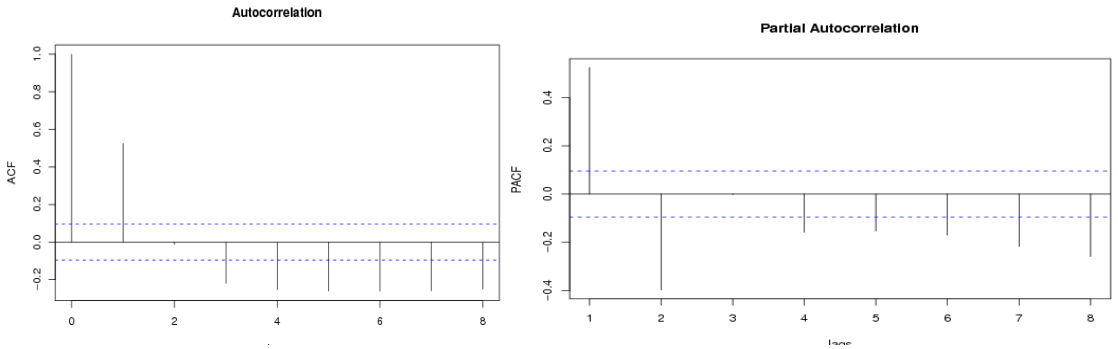
Ribb_Near_Addis_Zemen



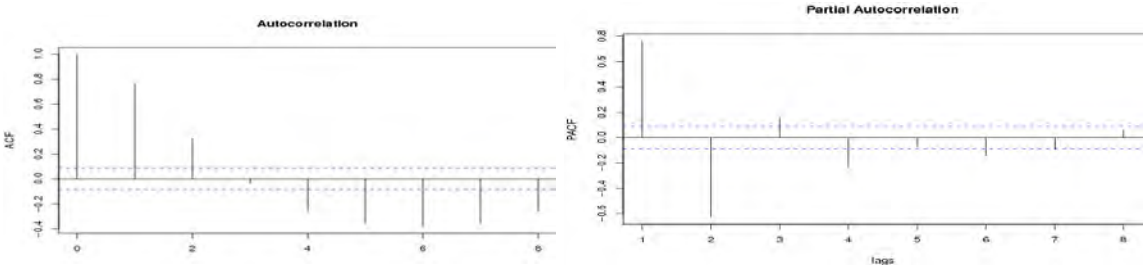
Gumara_Near_Bahir_dar



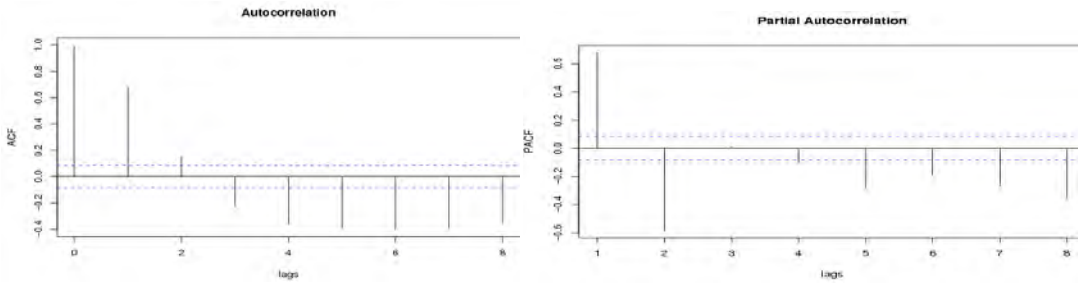
Megeche_Near_Azezo



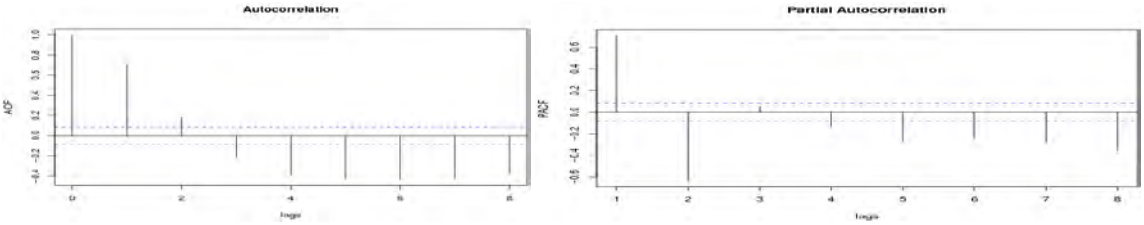
Abbay_Near_Kessie



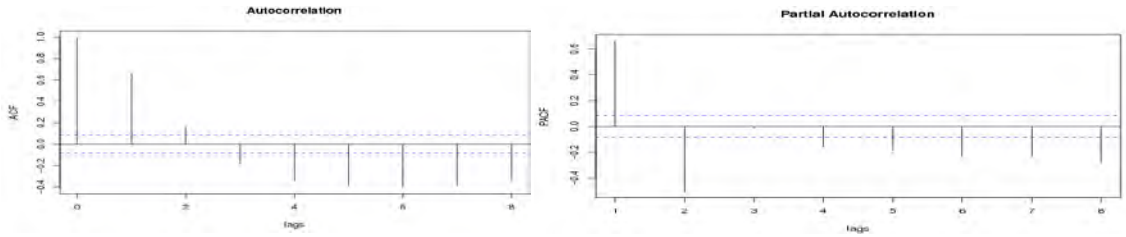
Muger_Near_Chancho



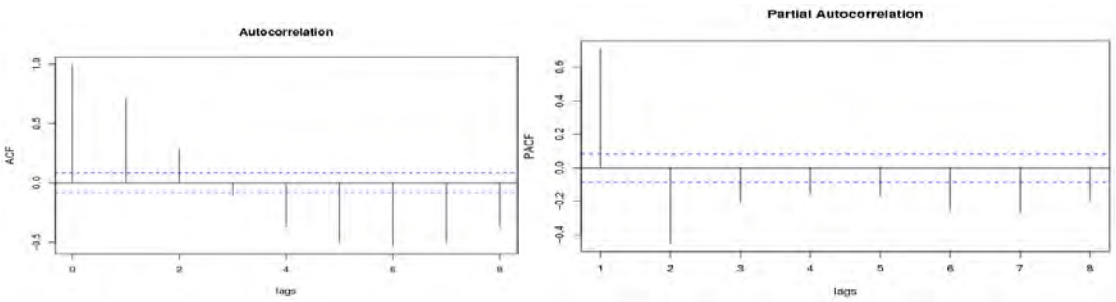
Abbay_Near_Bahir_dar



Guder_At_Guder



Gudla_Near_Dembecha



Main_Beles_Near_Metekel

Appendix C. Goodness of fit of the candidate PARMA models to the monthly streamflow.

C.1- Gilgel Abbay near Merawi station

Model	Autocorrelation	WN variance	AICC	SIC
PARMA(1,0)	all	0.869034	42.969	42.49
PARMA(1,1)	all	0.868979	45.266	46.294
PARMA(2,0)	all	0.868985	45.266	46.297
PARMA(2,1)	all	0.859808	47.203	49.623

C-2 Koga_Near_Merawi

Model	Autocorrelation	WN variance	AICC	SIC
PARMA(1,0)	all	0.746181	36.11	35.631
PARMA(1,1)	all	0.718323	36.698	37.726
PARMA(2,0)	all	0.7125346	37.254	38.568
PARMA(2,1)	all	0.71291	38.772	41.192

C-3 Ribb_Near_Addis_Zemen

Model	Autocorrelation	WN variance	AICC	SIC
PARMA(1,0)	all	0.717331	33.015	32.476
PARMA(1,1)	all	0.701589	34.376	35.283
PARMA(2,0)	all	0.704939	34.581	35.488
PARMA(2,1)	all	0.693304	36.302	38.533

C-4 Gumara_Near_Bahir_dar

Model	Autocorrelation	WN variance	AICC	SIC
PARMA(1,0)	all	0.792196	31.222	30.402
PARMA(1,1)	all	0.791739	33.601	33.937
PARMA(2,0)	all	0.791748	33.601	33.938
PARMA(2,1)	all	0.791677	36.157	37.49

C-5 Megeche_Near_Azezo

Model	Autocorrelation	WN variance	AICC	SIC
PARMA(1,0)	all	0.803342	38.658	38.149
PARMA(1,1)	all	0.798568	40.703	41.671
PARMA(2,0)	all	0.79977	40.769	40.769
PARMA(2,1)	all	0.759062	40.896	43.223

C-6 Abbay_Near_Kessie

Model	Autocorrelation	WN variance	AICC	SIC
PARMA(1,0)	all	0.539197	22.995	22.626
PARMA(1,1)	all	0.539047	25.254	26.504
PARMA(2,0)	all	0.539047	25.254	26.504
PARMA(2,1)	all	0.527215	26.542	29.308
PARMA(3,1)	all	0.537339	27.474	30.24

C-7 Muger_Near_Chancho

Model	Autocorrelation	WN variance	AICC	SIC
PARMA(1,0)	all	0.996763	39.262	38.442
PARMA(1,1)	all	0.870198	36.908	37.244
PARMA(2,0)	all	0.869021	37.905	38.0254
PARMA(2,1)	all	0.838273	38.159	39.492

C-8 Abbay_Near_Bahir_dar

Model	Autocorrelation	WN variance	AICC	SIC
PARMA(1,0)	all	0.749066	35.58	35.071
PARMA(1,1)	all	0.640352	30.988	31.956
PARMA(2,0)	all	0.660941	32.38	33.348
PARMA(2,1)	all	0.660926	34.805	37.132

C-9 Fatto_Near_Guder

Model	Autocorrelation	WN variance	AICC	SIC
PARMA(1,0)	all	0.939307	46.468	45.989
PARMA(1,1)	all	0.937003	48.657	49.685
PARMA(2,0)	all	0.937358	48.674	49.702
PARMA(2,1)	all	0.916662	50.084	52.504

C-10 Guder_At_Guder

Model	Autocorrelation	WN variance	AICC	SIC
PARMA(1,0)	all	0.841105	41.499	41.02
PARMA(1,1)	all	0.764225	44.85	45.878
PARMA(2,0)	all	0.768733	39.75	40.778
PARMA(2,1)	all	0.738383	40.352	42.772

C-11 Gudla_Near_Dembecha

Model	Autocorrelation	WN variance	AICC	SIC
PARMA(1,0)	all	0.861145	41.715	41.207
PARMA(1,1)	all	0.85235	43.571	44.539
PARMA(2,0)	all	0.852357	43.571	44.539
PARMA(2,1)	all	0.802154	43.326	45.653

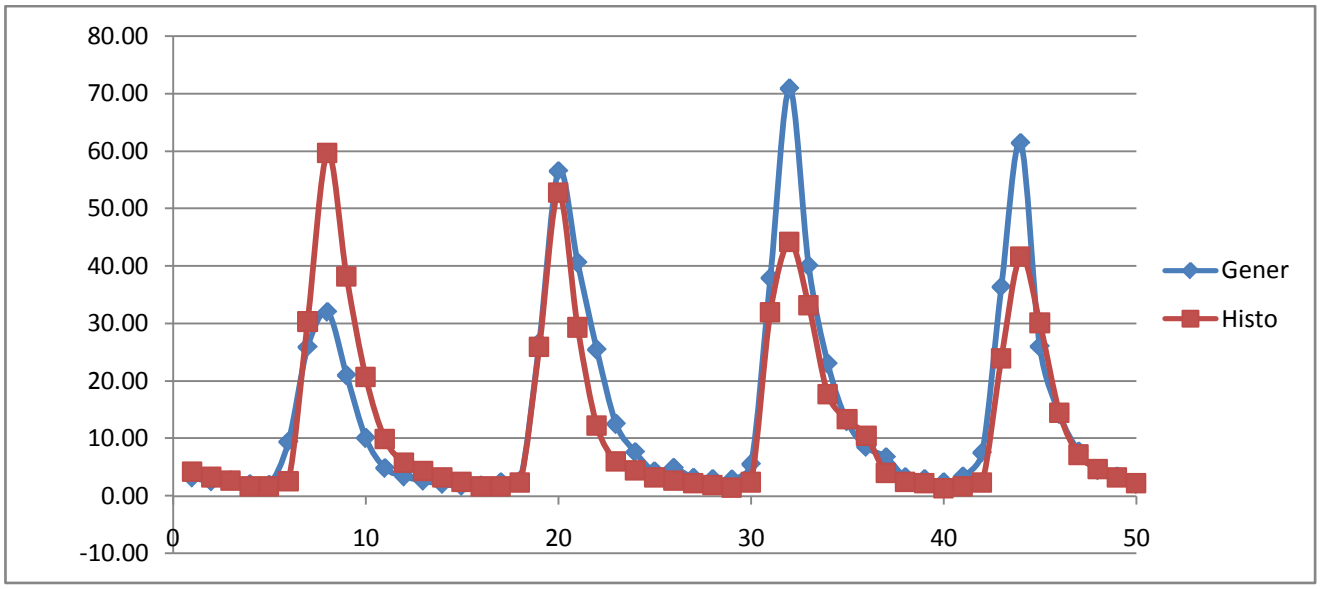
C-12 Diddesa_Near_Arjo

Model	Autocorrelation	WN variance	AICC	SIC
PARMA(1,0)	all	0.751878	36.453	35.973
PARMA(1,1)	all	0.696899	35.335	36.363
PARMA(2,0)	all	0.697938	35.402	36.43
PARMA(2,1)	all	0.769455	37.534	39.954

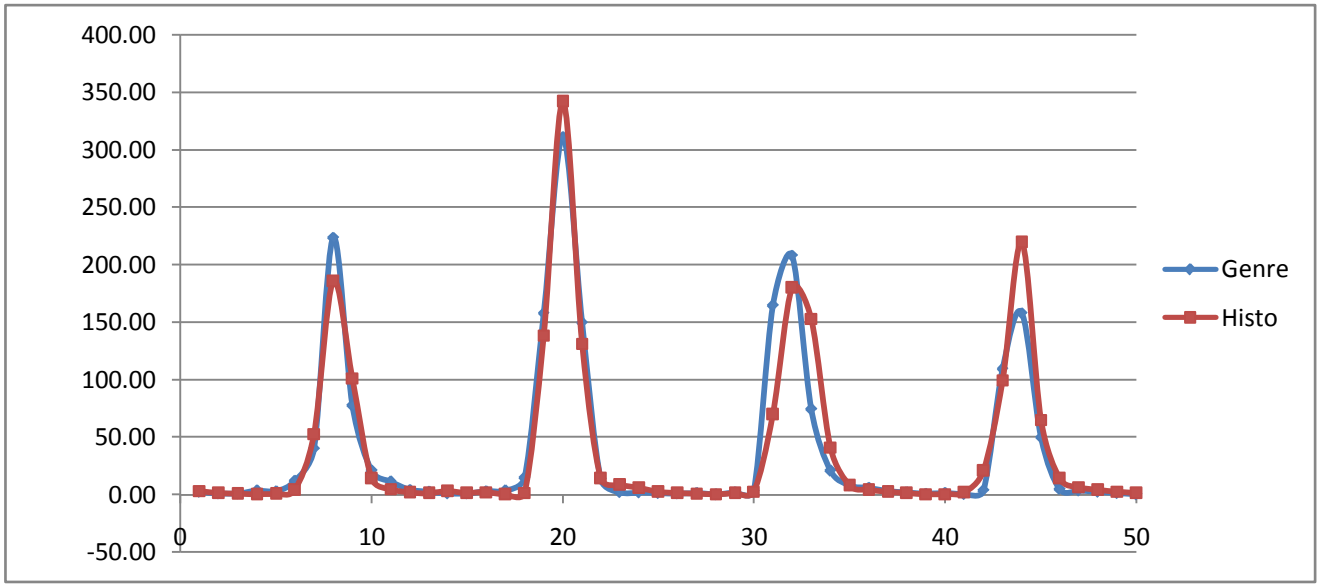
C-13 Main_Beles_Near_Metekel

Model	Autocorrelation	WN variance	AICC	SIC
PARMA(1,0)	all	0.818312	37.886	37.316
PARMA(1,1)	all	0.810911	39.828	40.672
PARMA(2,0)	all	0.811128	39.84	40.684
PARMA(2,1)	all	0.809374	42.198	44.33

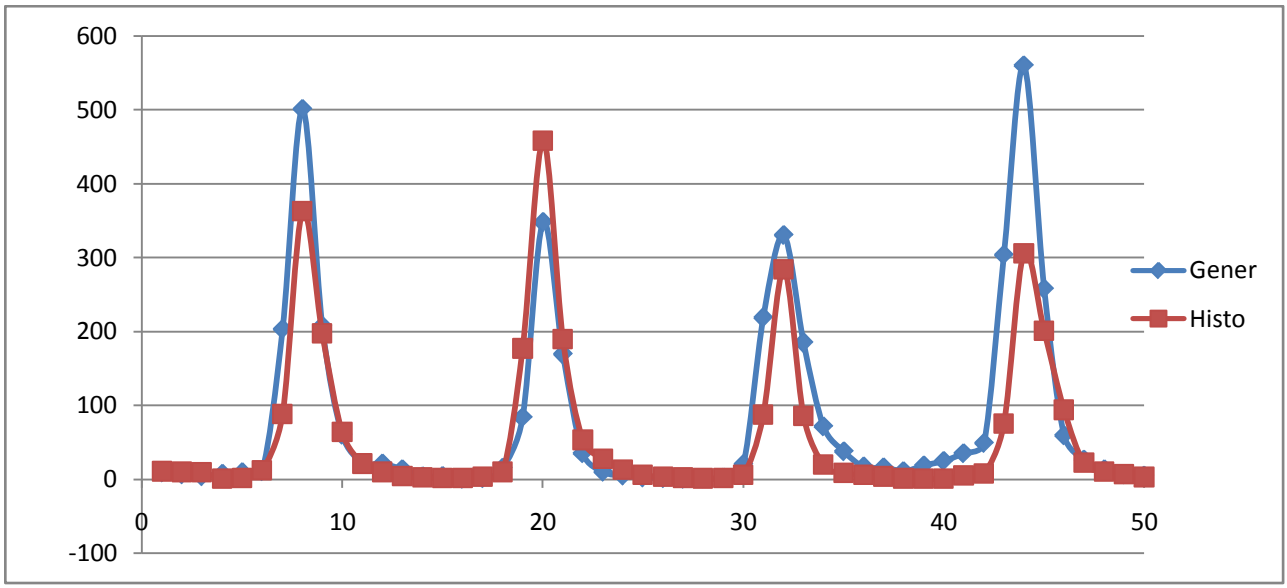
Appendix D: Fitted periodic autoregressive moving average PARMA (p, q) models.



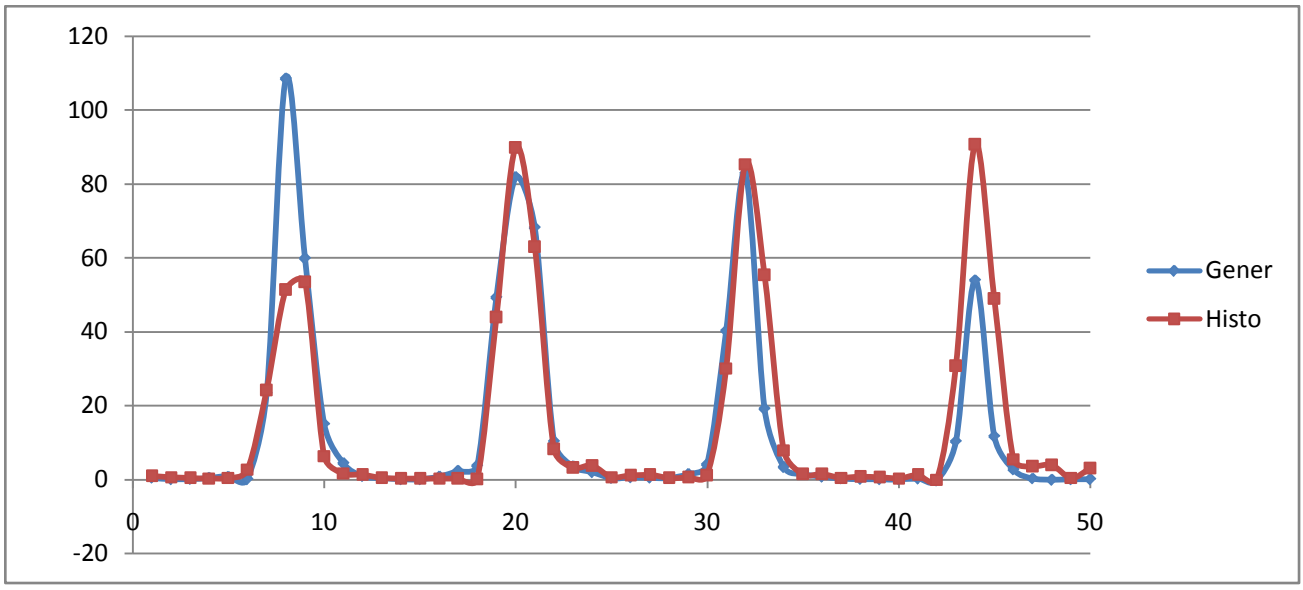
Koga_Near_Merawi



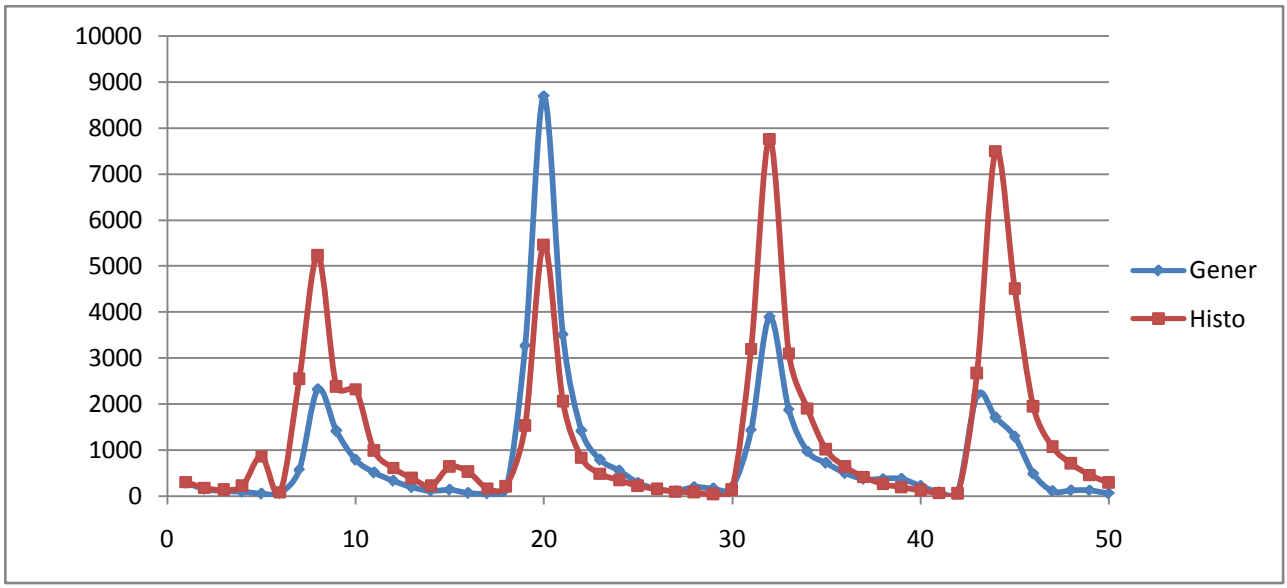
Ribb_Near_Addis_Zemen



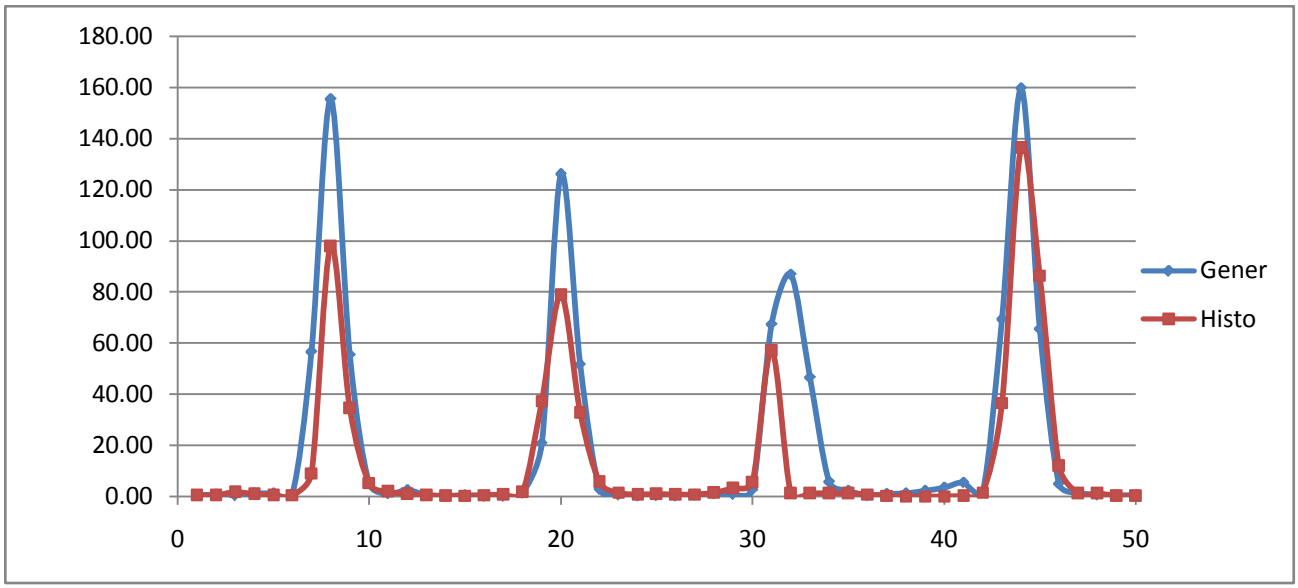
Gumara_Near_Bahir_dar



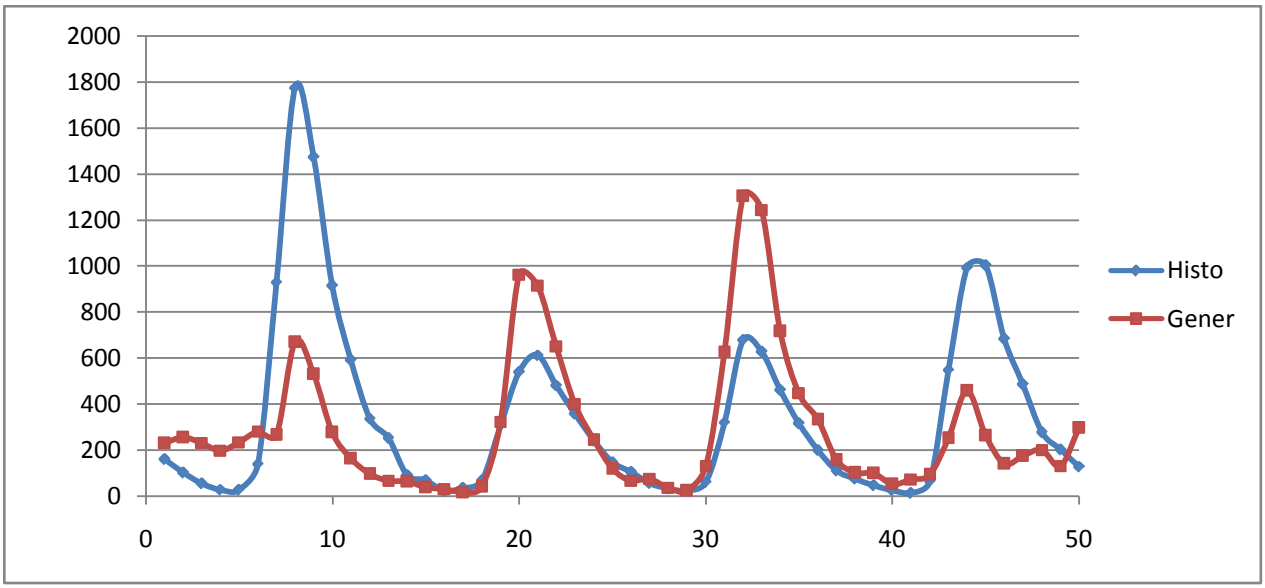
Megeche_Near_Azezo



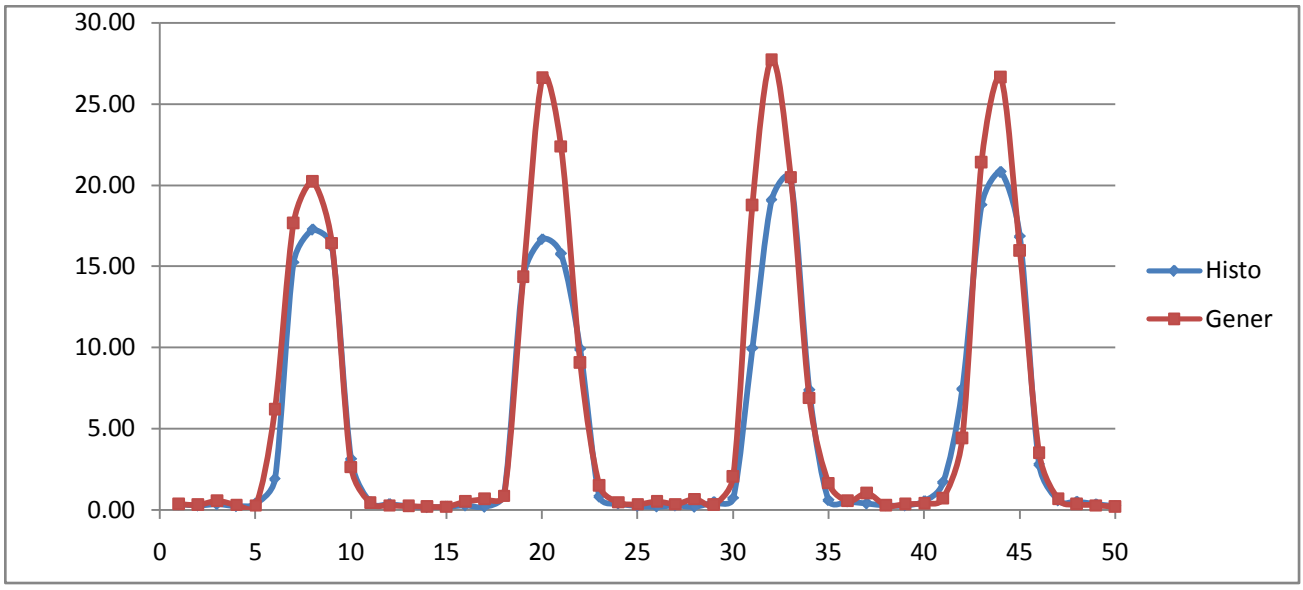
Abbay_Near_Kessie



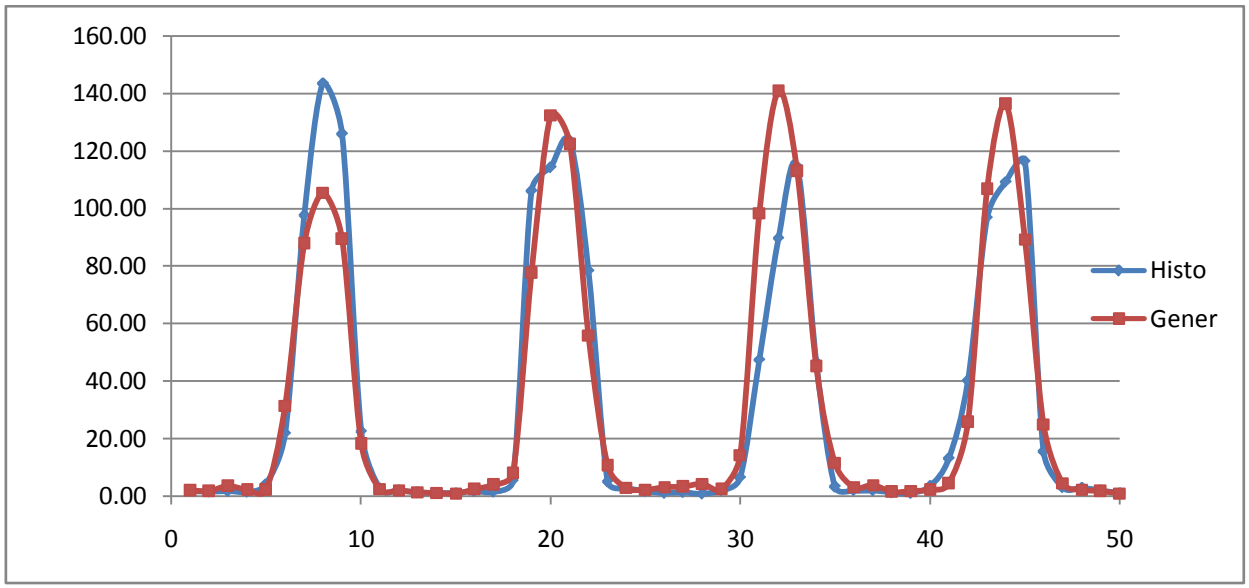
Muger_Near_Chancho



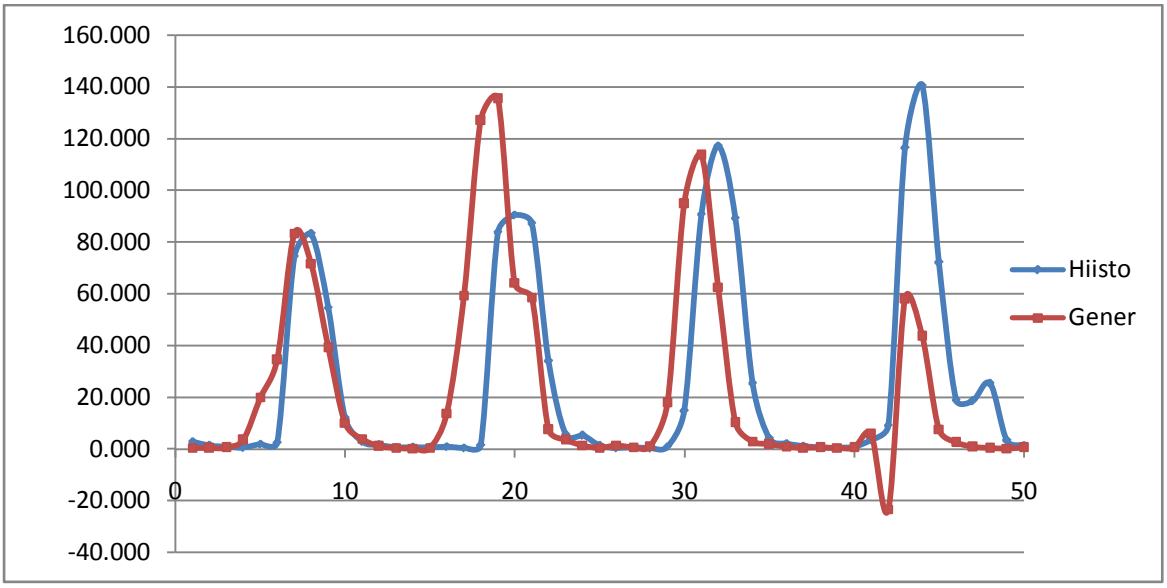
Abbay_Near_Bahir_dar



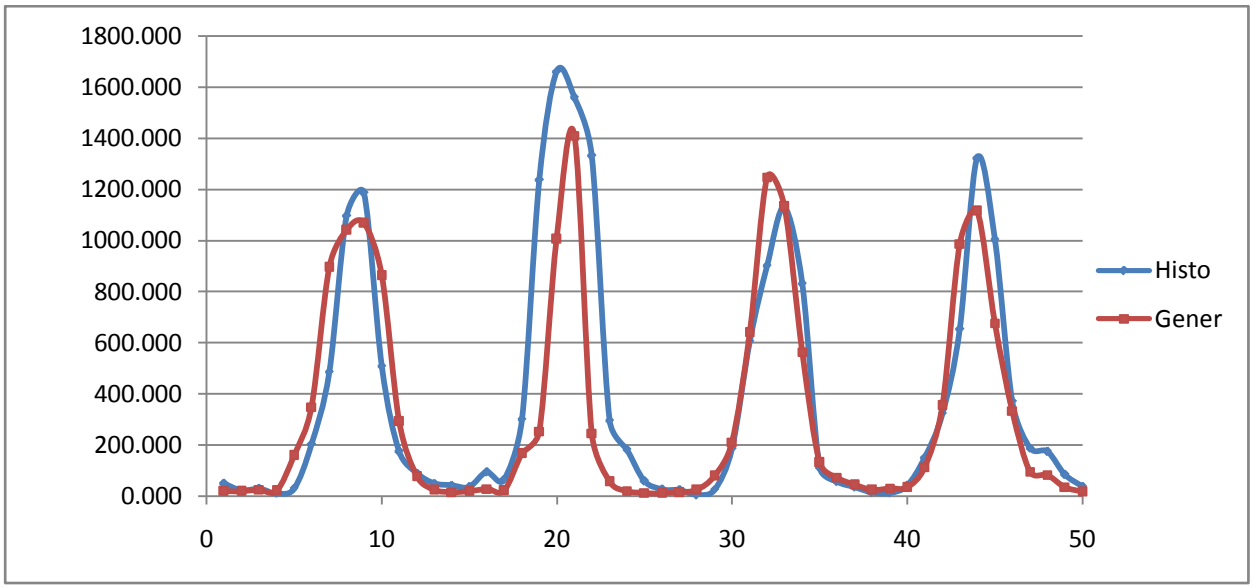
Fatto_Near_Guder



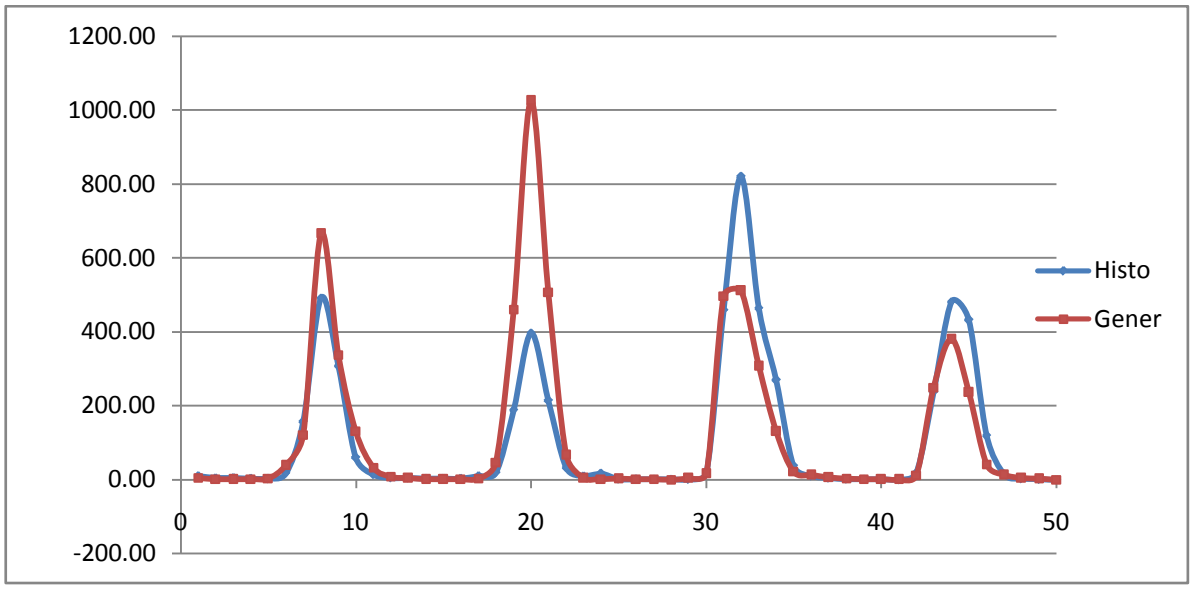
Guder_At_Guder



Gudla_Near_Dembecha

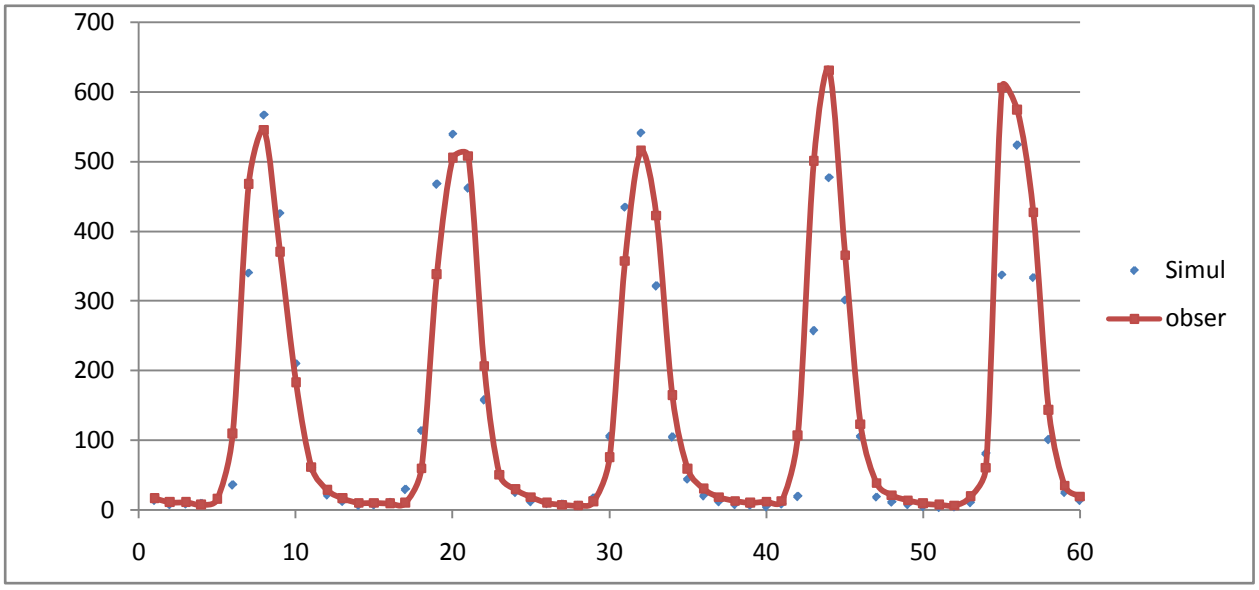


Diddesa_Near_Arjo

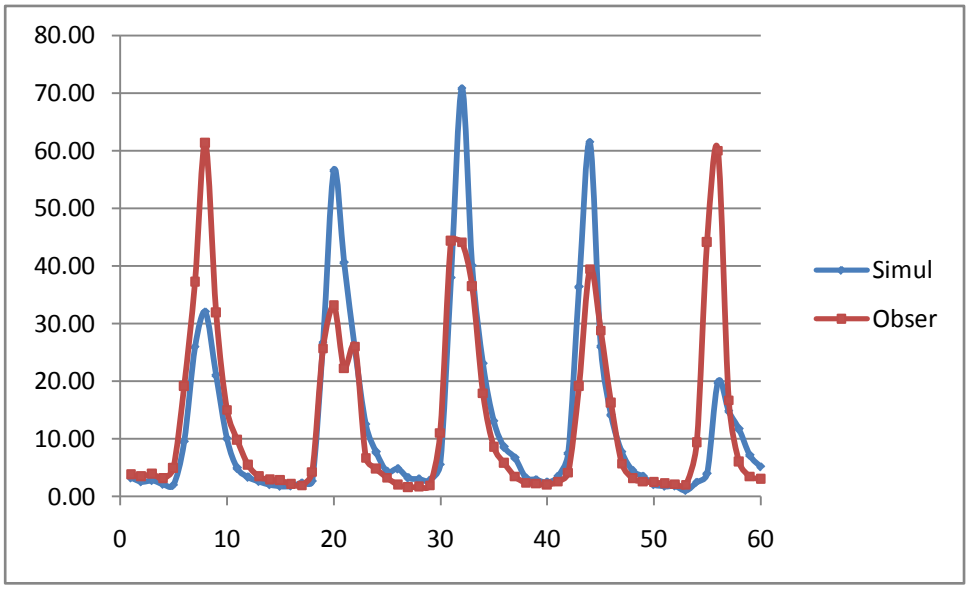


Main_Beles_Near_Metekel

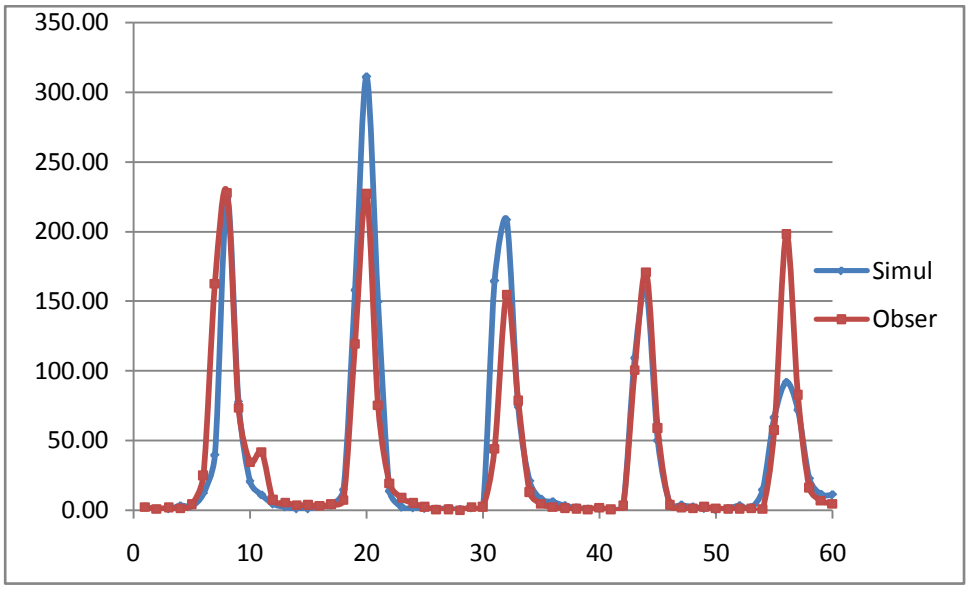
Appendix E: Plot of the verification results



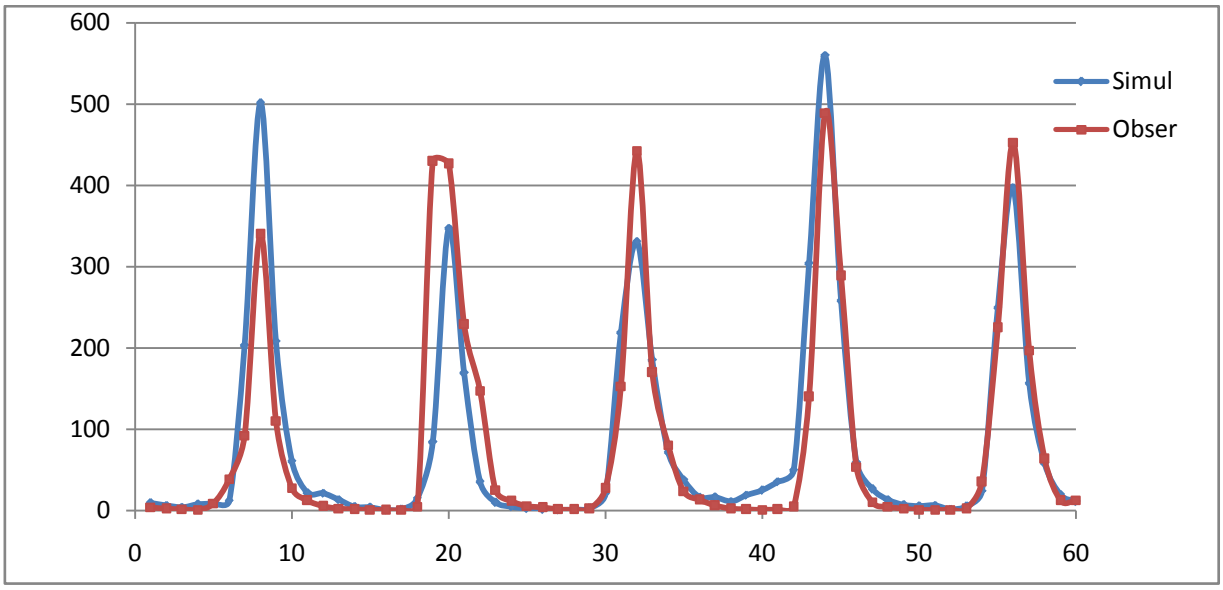
Gilgel_Abbay_Near_Merawi



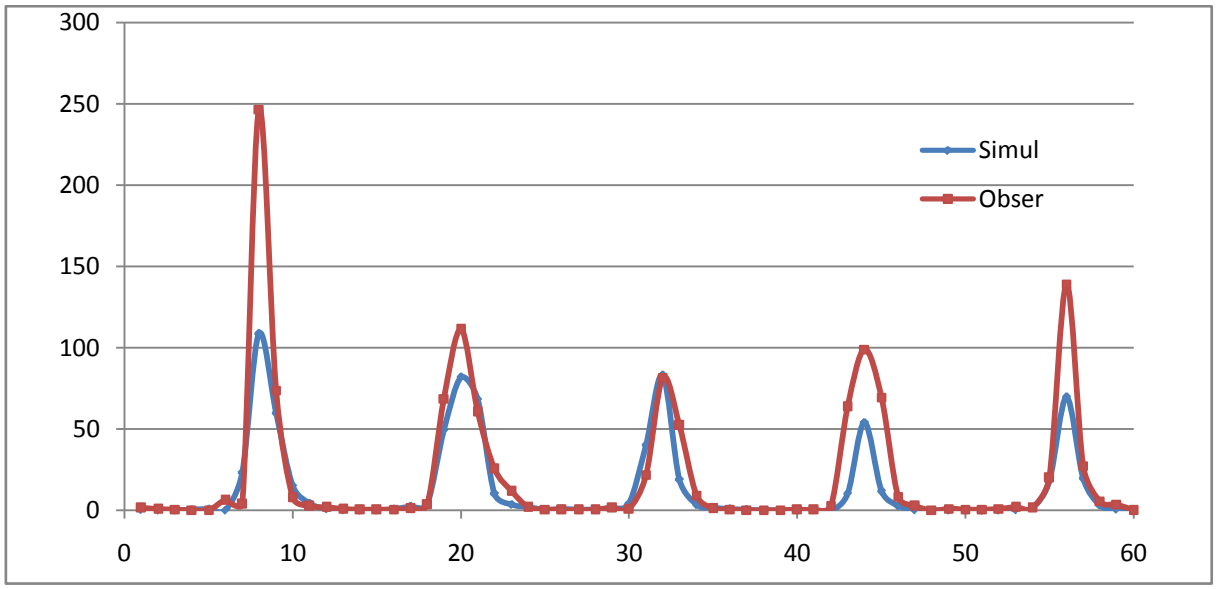
Koga_Near_Merawi



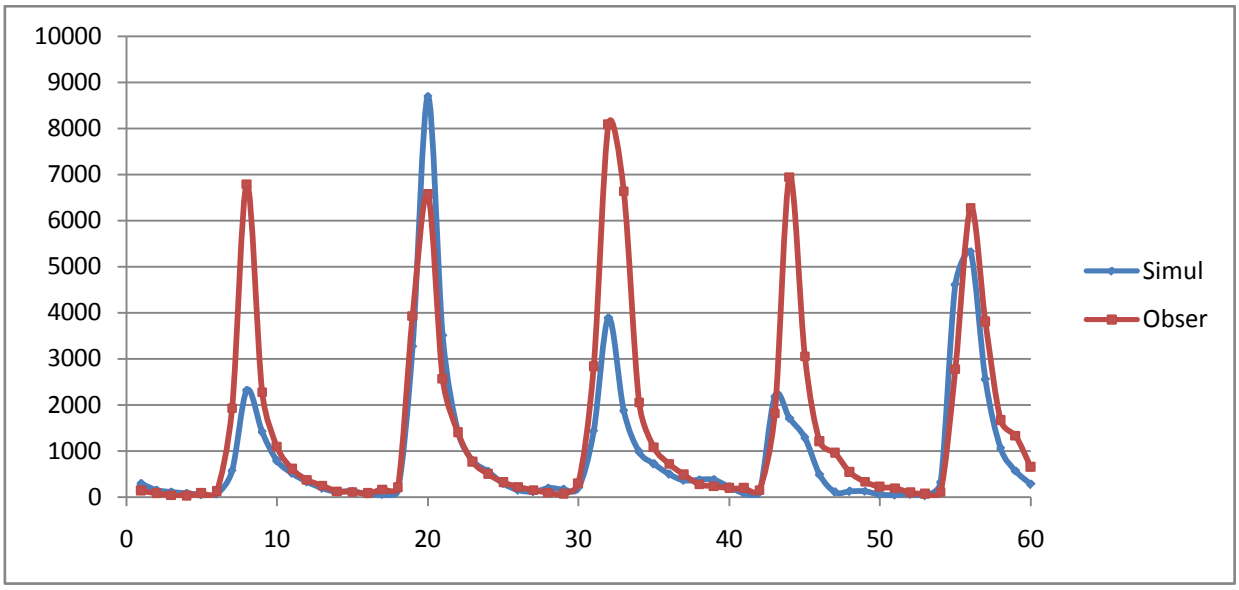
Ribb_Near_Addis_Zemen



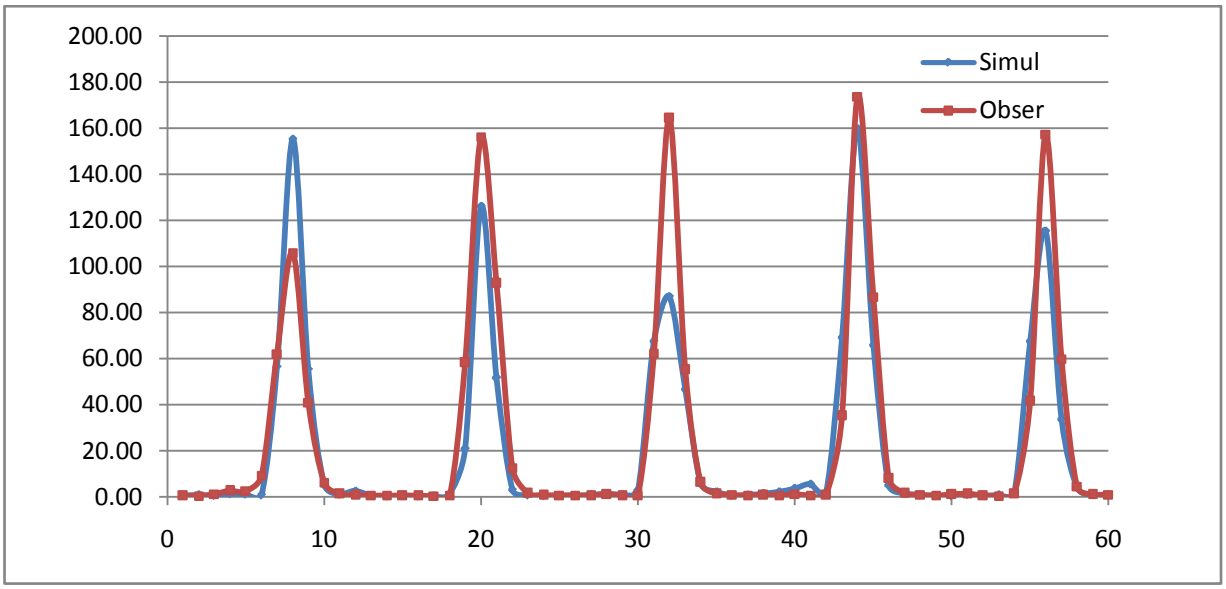
Gumara_Near_Bahir_dar



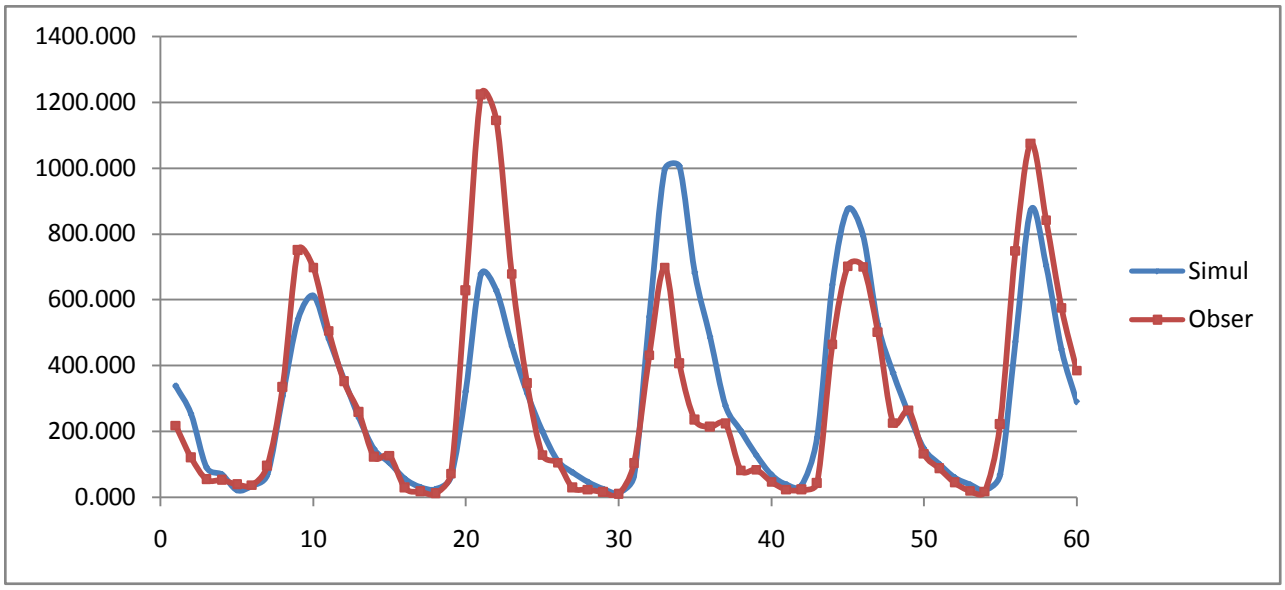
Megeche_Near_Azezo



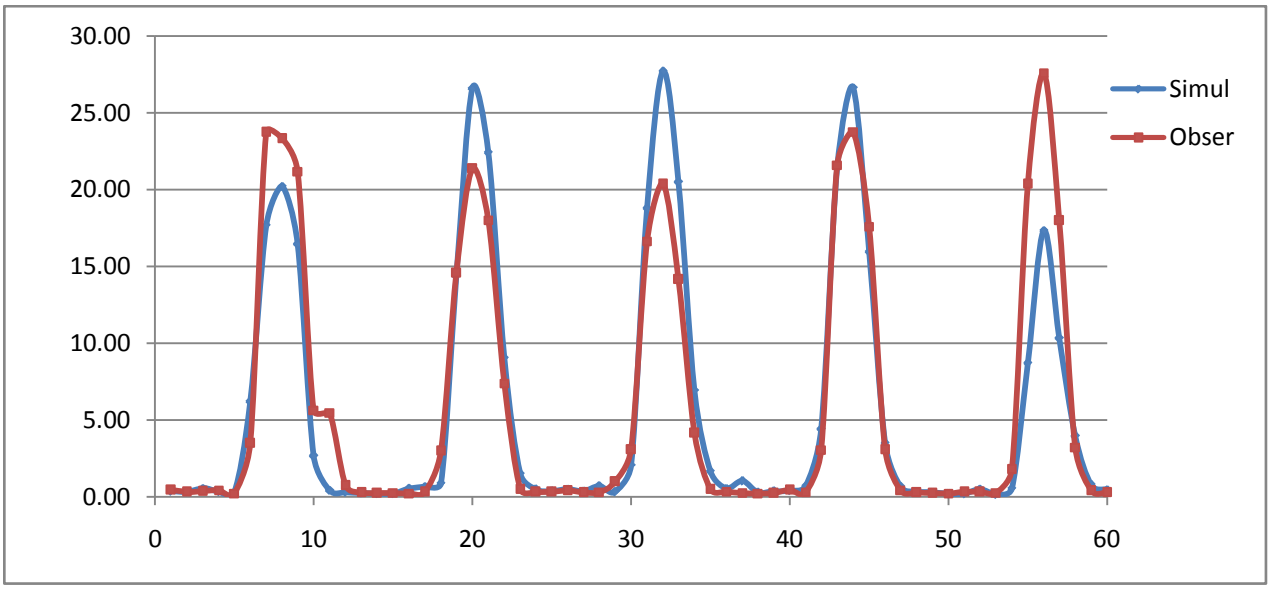
Abbay_Near_Kessie



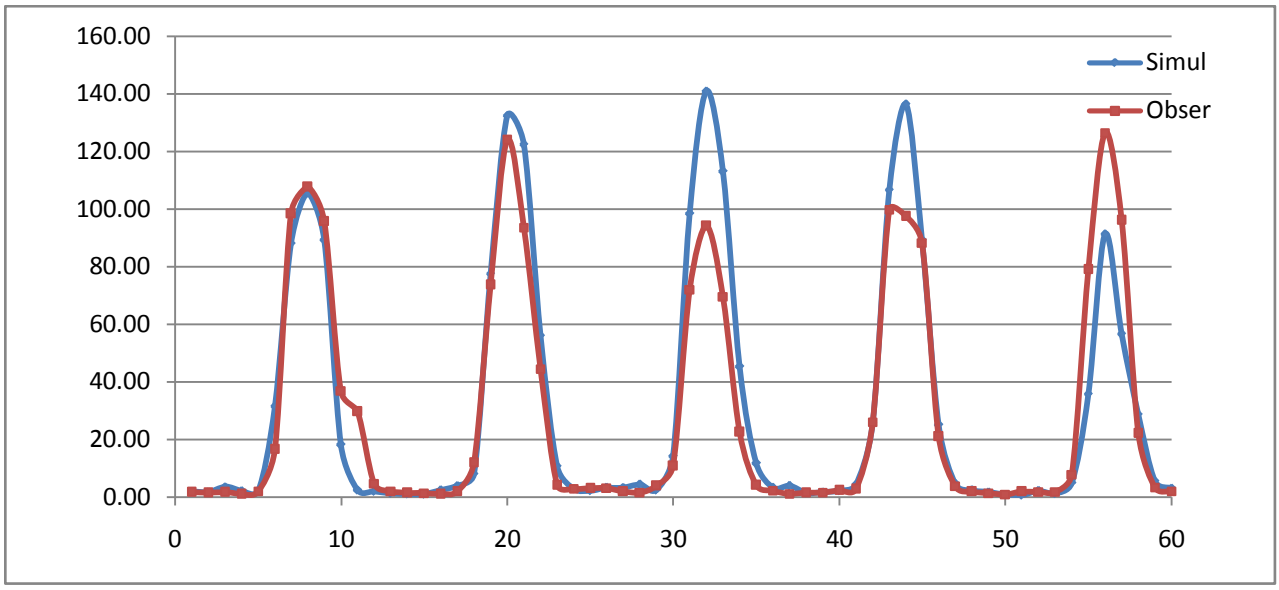
Muger_Near_Chancho



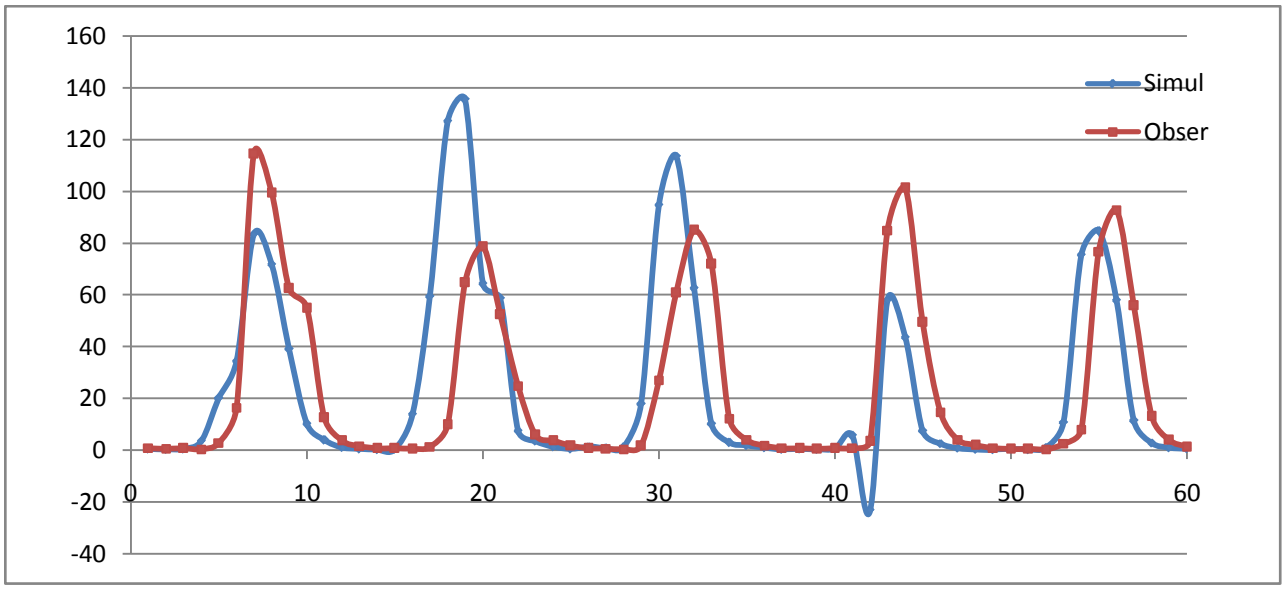
Abbay_Near_Bahir_dar



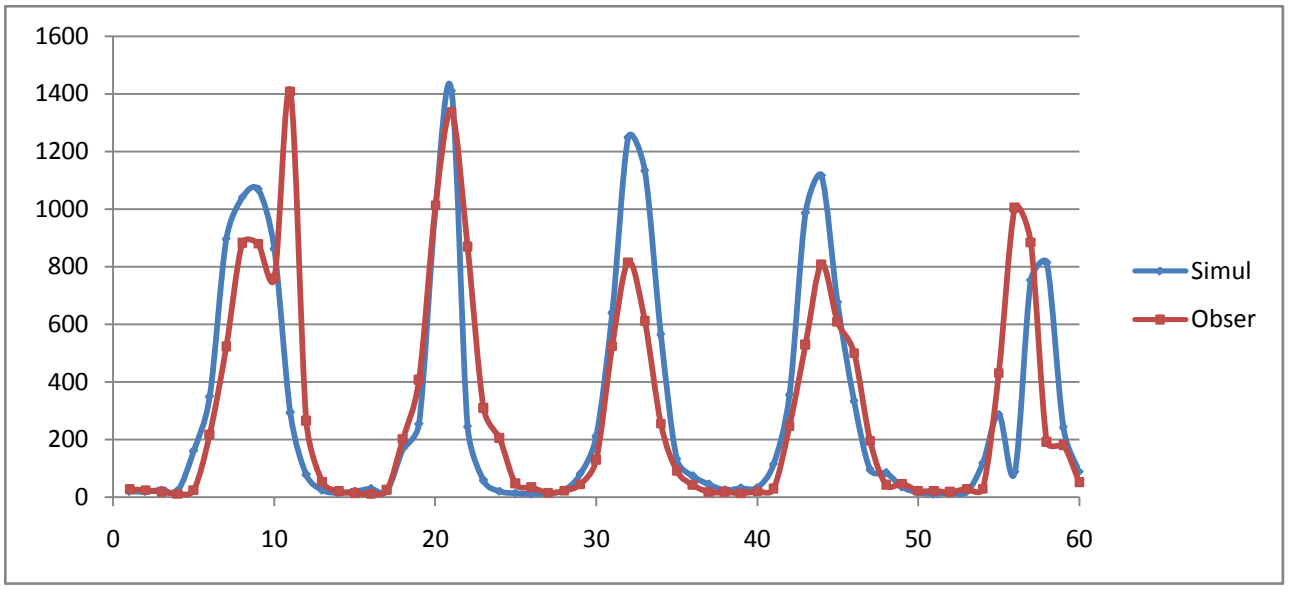
Fatto_Near_Guder



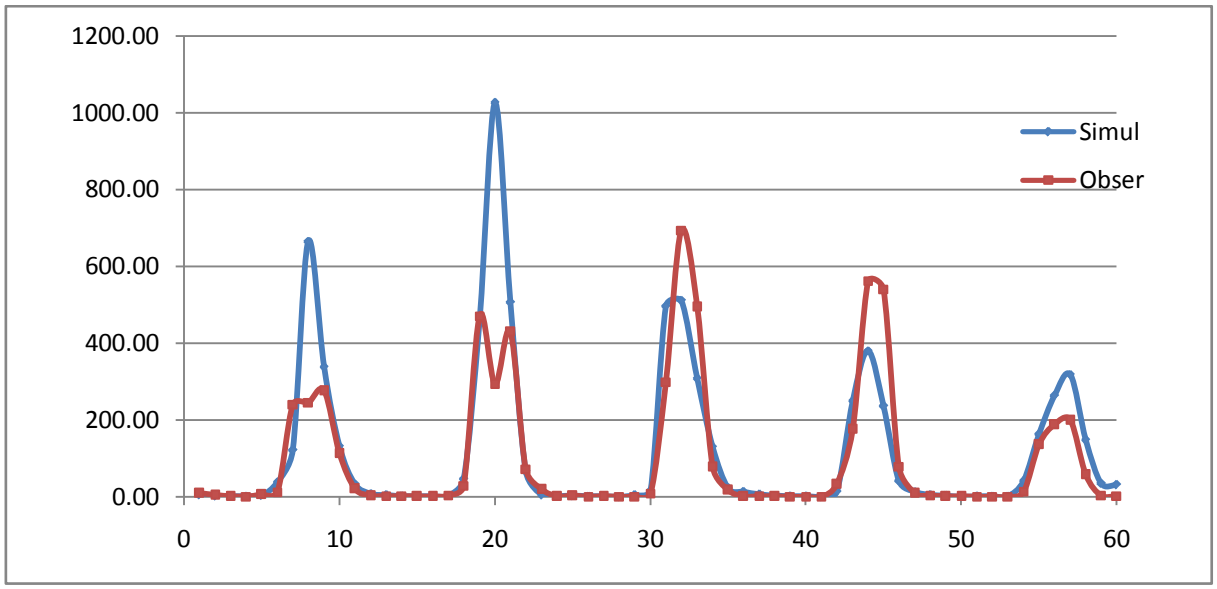
Guder_At_Guder



Gudla_Near_Dembecha



Diddesa_Near_Arjo



Main_Beles_Near_Metekel

Appendix F

F.1 Cluster Analysis

The use of cluster analysis for regionalization involves grouping of various observations and variables into clusters, so that each cluster is composed of observations or variables with similar characteristics such as geographical, physical, statistical or stochastic behavior.
In the cluster analysis each gauging site is associated with one or more catchment characteristics then a measure of similarity between two sub-catchments is measured by a scaled Eucludean distance between them in catchment characteristic space. A number of variables are considered in the cluster analysis.
The variables considered in this study for the identification of gauging stations which have similar stochastic behavior are;

- The autocorrelation coefficients between the average discharge
- Mean $_i$ is the mean value of month $_i$ of the year;
- Sd $_i$ is the standard deviation of month i of the year;
- Correlation coefficients with upstream and downstream discharge gauging stations,
- Drought duration
- Drought severity

The reason for choosing these variables are as follows: 10 autocorrelation coefficients are used because serial dependence is of most importance when fitting univariate model to a time series; monthly mean and standard deviation are basic statistics of flow series; the correlation coefficients with upstream and downstream discharge gauging stations are used to see weather gauging stations reflect the cross-correlation relationship between streamflow at upstream and downstream stations.

In this analysis hierarchical clustering using Ward's method is used. This procedure attempts to identify cases (or variables) based on selected characteristics , using an algorithm that starts with each case (or variable) in a separate cluster and combines clusters until only one is left

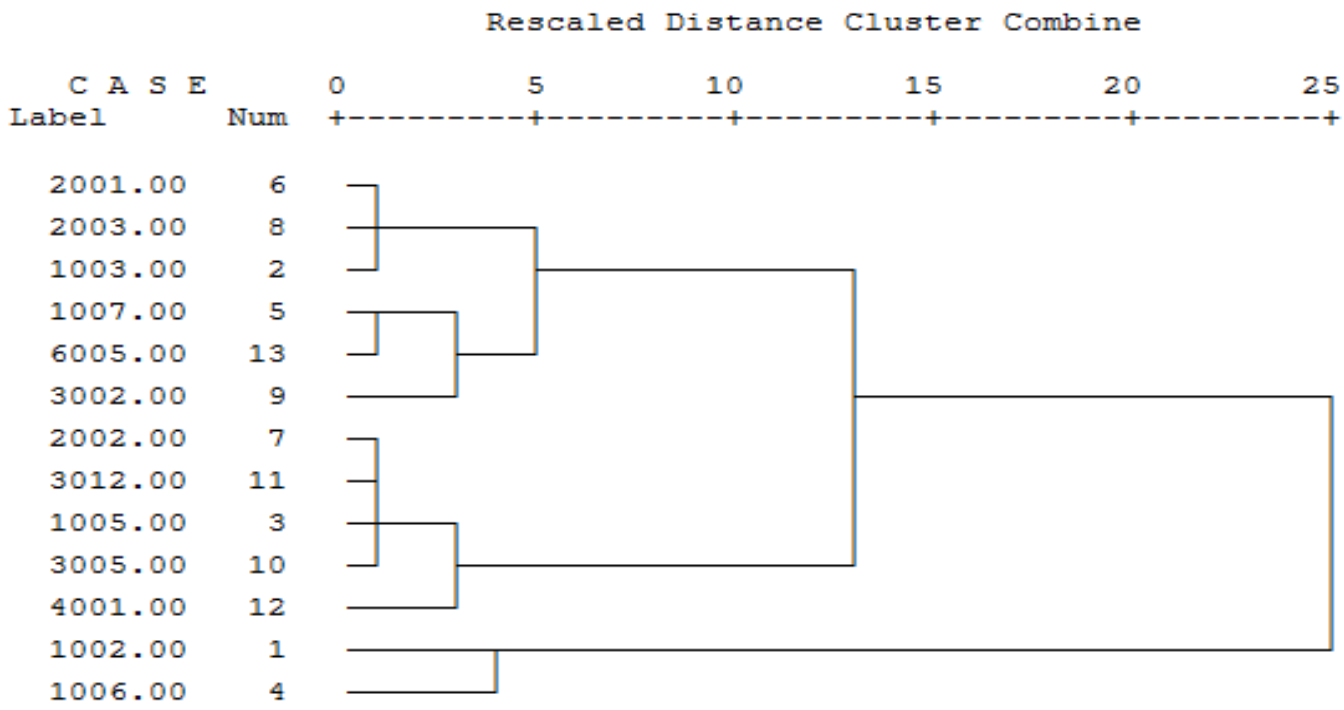


Figure:(D.1) Dendrogram using Ward's Method

In the dendrogram (fig:D.1) it can be noticed that if only two clusters are required all the stations except two stand as one cluster and station no. 1002 at Gilgel abbay and station no.1006 at Megech stands out as a separate cluster. If three clusters are required the bigger cluster splits to two.

D.1.1 Discriminant Analysis

After homoginenous stations have been identified, a test was carried out to find out whether the proposed pool was reasonably homogenous. For that, we have to look at the agglomeration schedule in fig: 5.1. In the column labeled Coefficients, we can see the value of the distance (or similarity) statistic used to form the cluster. From these numbers, we get an idea of how unlike the clusters being combined are. If we are using dissimilarity measures, small coefficients tell us that fairly homogenous clusters are being attached to each other as it is in our case. Large coefficients tell us that we are combining dissimilar clusters. If we're using similarity measures, the opposite is true: large values are good, while small values are bad.

The actual value shown depends on the clustering method and the distance measure we're using. we can use these coefficients to help you decide how many clusters we need to represent the

data. We may stop cluster formation when the increase (for distance measures) or decrease (for similarity measures) in the Coefficients column between two adjacent steps is large. The proposed pool was further sub-divided if the value of coefficient was found to be more than 0.3. This value was considered a reasonable upper limit to considered homogeneity.

In our case, we may want to stop at the two cluster solution, after stage 8. Here, as we can confirm from fig: 5.1, station no.1002 and station no. 1006 are in cluster 1; all the remaining stations are in cluster 1. If we go on to combine these two clusters in stage 9, the distance coefficient across the last combination jumps from 0.219 to 0.603.

Table 5.1 The Agglomeration schedule

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	2	8	0.002	0	0	3
2	7	3	0.006	0	0	5
3	2	6	0.012	1	0	7
4	5	4	0.021	0	0	7
5	3	7	0.035	0	2	6
6	3	6	0.079	5	0	9
7	2	5	0.219	3	4	8
8	2	9	0.296	7	0	8
9	1	2	0.603	6	0	0

Gauging stations with similar stochastic behavior are established using hierarchical clustering as explained above. Thus, two prominent groups are developed. Among the groups of stations identified, one at upstream portion of the basin, station 111002, and another at the downstream portion, station 113005, was selected for streamflow drought analysis.

Appendix G

Table G-1 List of distributions function used for drought parameters analysis.

#	Distribution	Kolmogorov Smirnov		Anderson Darling		Chi-Squared	
		Statistic	Rank	Statistic	Rank	Statistic	Rank
1	Exponential	0.08903	17	1.9287	11	13.877	3
2	Exponential (2P)	0.08759	16	1.7533	10	16.808	10
3	Gamma	0.07381	9	2.3611	14	21.308	18
4	Gamma (3P)	0.07394	10	1.2608	5	15.606	7
5	Gen. Extreme Value	0.07017	6	1.7274	9	14.699	4
6	Gen. Gamma	0.07093	7	1.4271	8	17.99	13
7	Gen. Gamma (4P)	0.0432	1	4.451	18	N/A	
8	Gen. Pareto	0.04427	2	0.62286	1	10.9	1
9	Gumbel Max	0.07595	12	1.9341	12	11.157	2
10	Gumbel Min	0.15133	23	22.032	25	42.565	22
11	Inv. Gaussian	0.15632	24	30.165	26	60.623	23
12	Inv. Gaussian (3P)	0.08612	15	2.3888	15	16.744	8
13	Log-Logistic	0.13038	22	4.8756	20	32.515	21
14	Log-Logistic (3P)	0.07793	13	3.1577	17	20.582	17
15	Log-Pearson 3	0.07121	8	1.2437	4	18.882	15
16	Logistic	0.10625	19	5.7985	21	24.59	19
17	Lognormal	0.12269	21	4.5585	19	26.353	20
18	Lognormal (3P)	0.08375	14	2.4714	16	16.789	9
19	Normal	0.11735	20	5.8347	22	20.051	16
20	Pareto	0.33749	27	54.978	27	155.48	26
21	Pareto 2	0.09521	18	2.2191	13	18.413	14
22	Pearson 5	0.1944	25	13.264	23	84.07	24
23	Pearson 5 (3P)	0.20011	26	16.816	24	95.498	25
24	Pearson 6	0.06708	5	1.2697	6	15.171	5
25	Pearson 6 (4P)	0.0748	11	1.3182	7	16.856	11
26	Weibull	0.06217	3	0.93296	2	17.179	12
27	Weibull (3P)	0.06331	4	1.1179	3	15.442	6
28	Log-Gamma	No fit					

Appendix H

Figure H-1. Comparison of historic and simulated distribution functions :(a)yearly streamflow drought severity; (b) yearly streamflow drought duration at Guder at Guder station.

(a) (b)

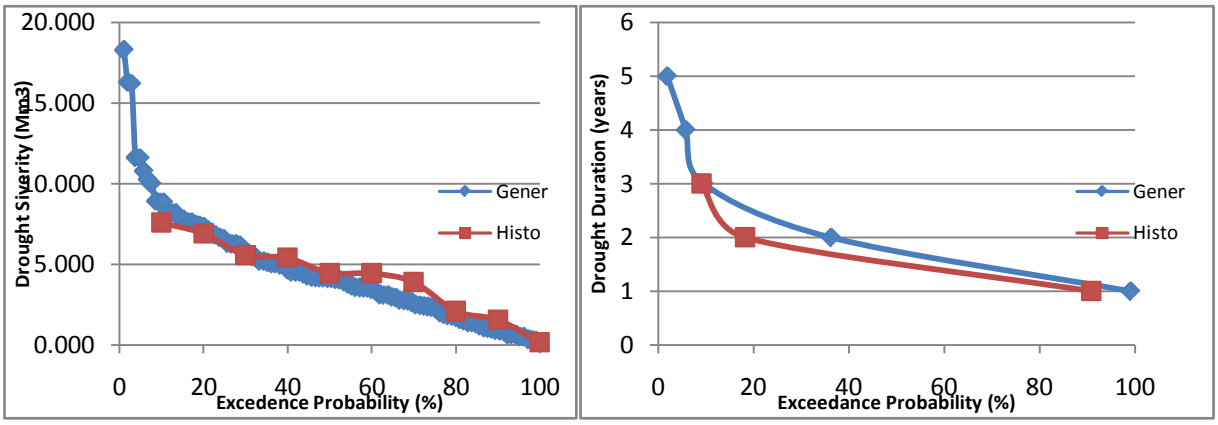


Figure H-2. Comparison of historic and simulated distribution functions :(a)monthly streamflow drought severity; (b) monthly streamflow drought duration at Guder station.

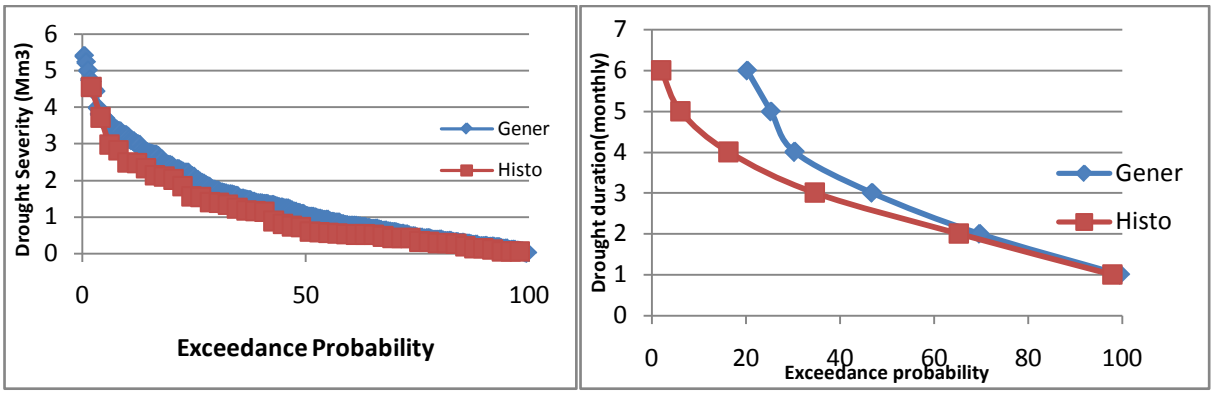


Figure: H-3. Comparison of stimulated and theoretical probability distribution :(left) yearly streamflow drought severity; (right) monthly drought severity at Guder station.

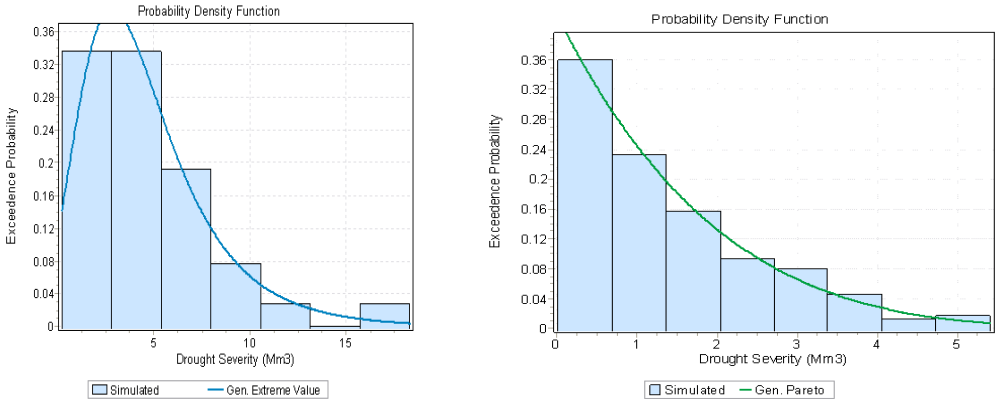
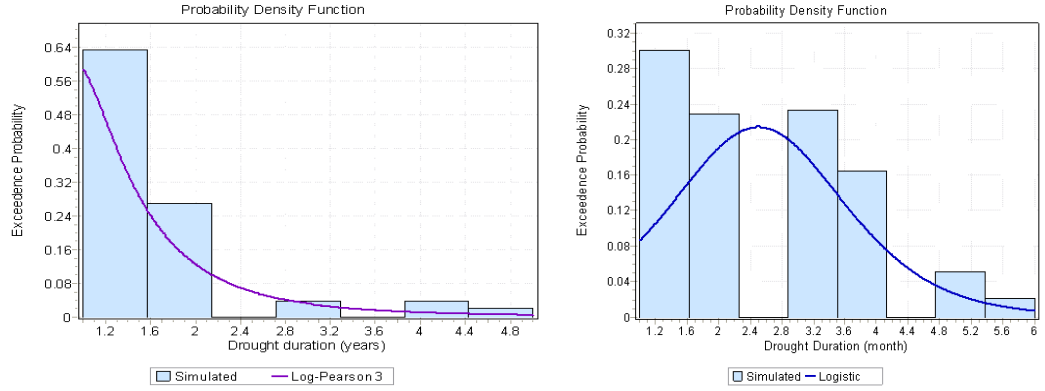


Figure: H-4. Comparison of stimulated and theoretical probability distribution :(left) yearly streamflow drought duration; (right) monthly streamflow drought duration at Guder at Guder station.



Appendix I: Fitted Comulative distribution function to the drought parameters

Station	Fitted Cdf to Drought Severity (S)	Fitted Cdf to Drought Duration (D)
111002	$F(S) = -\left[-0.095\frac{(s-0.0495)}{19.499}\right]^{1/0.095}$	$F(D) = \frac{1}{1+(\frac{d}{2.776})^{-1.428}}$
111003		
111005	$F(S) = 1-\left[-0.271\frac{(s-0.059)}{9.35}\right]^{1/0.271}$	$F(D) = exp\left\{-e^{-\frac{d-0.96}{0.471}}\right\}$
111006	$F(S) = 1-\left[-0.22\frac{(s-0.32)}{23.526}\right]^{1/0.22}$	$F(D) = exp\left\{-e^{-\frac{d-0.74}{0.561}}\right\}$
111007	$F(S) = 1-\left[-0.36\frac{(s-0.092)}{5.308}\right]^{1/0.36}$	$F(D) = exp\left\{-e^{-\frac{d-0.825}{0.471}}\right\}$
112001	$F(S) = exp\left\{-\left[1-0.105(\frac{s-2.35}{2.25})\right]^{1/0.105}\right\}$	$F(D) = \frac{1}{1+(\frac{d}{1.35})^{-0.25}}$
112002	$F(S) = 1-\left[-0.294\frac{(s-0.105)}{4.811}\right]^{1/0.294}$	$F(D) = exp\left\{-e^{-\frac{d-0.943}{0.309}}\right\}$
112003		
113002	$F(S) = 1-\left[-0.179\frac{(s-0.136)}{1.139}\right]^{1/0.179}$	$F(D) = \frac{1}{1+(\frac{d}{1.68})^{-0.35}}$
113005	$F(S) = -\left[-0.0896\frac{(s-0.0485)}{20.135}\right]^{1/0.089}$	$F(D) = \frac{1}{1+(\frac{d}{2.776})^{-1.428}}$
113012	$F(S) = \frac{1}{\sqrt{2\pi}} * \frac{1}{0.94S} exp\left\{-\frac{1}{2}(\frac{logS}{0.94}-1.33)^2\right\}$	$F(D) = exp\left\{-e^{-\frac{d-0.903}{0.539}}\right\}$

114001	$F(S) = \frac{1}{\sqrt{2\pi}} * \frac{1}{0.86S} exp\left\{-\frac{1}{2}(\frac{logS}{0.86} - 1.39)^2\right\}$	$F(D) = \frac{1}{1 + (\frac{d}{2.608})^{-1.023}}$
116005	$F(S) = -\left[-0.238\frac{(s - 1.004)}{27.607}\right]^{1/0.238}$	$F(D) = \frac{1}{1 + (\frac{d}{1.253})^{-1.541}}$

Declaration and Copyright

Here I declare that this thesis is my own original work, and has not been presented for a degree in any university and that all sources of material used for the thesis have been dually acknowledged.

Yohannes Hagos Subagadis