DISSERTATION

MODELING THE UNCERTAINTY OF HYDROLOGIC PROCESSES EXHIBITING
CHANGES

Submitted By

Nidhal Saada

Department of Civil Engineering

In partial fulfillment of the requirements

for the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Summer, 1998

COLORADO STATE UNIVERSITY

May 19, 1998

WE HEREBY RECOMMEND THAT THE DISSERTATION PREPARED UNDER OUR SUPERVISION BY NIDHAL SAADA ENTITLED MODELING THE UNCERTAINTY OF HYDROLOGIC PROCESSES EXHIBITING CHANGES BE ACCEPTED AS FULFILLING IN PART REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY.

Committee on Graduate Work

Adviser

Department Head

ABSTRACT OF DISSERTATION
MODELING THE UNCERTAINTY OF HYDROLOGIC PROCESSES EXHIBITING CHANGES

The Geometric-Normal-Normal (GNN) model was analyzed and tested for the purpose of simulating hydrologic processes that exhibit changes. The general moment equations of the GNN model were derived, particularly the lag-k autocorrelation function. They can be used to estimate the model parameters based on the method of moments. Other estimation methods were also suggested. They include regression analysis, fitting the autocorrelation function (ACF), using the range properties, and using the run properties. The performance of these methods was tested by using simulation experiments. The results showed that in terms of bias and mean square error the regression and range methods are better than the other methods for estimating the model parameters. The GNN model was applied to the White Nile River flows at Malakal and the annual net basin supply (NBS) data for Lake St.Clair of the Great Lakes system. Simulation experiments were conducted to test the ability of the GNN model to preserve a number of observed statistics such as the mean, standard deviation, skewness, rescaled range, Hurst coefficient, longest drought, maximum deficit, and surplus. Results show that the GNN model, in general, performs quite well in preserving these statistics. An extended version of the GNN model was also formulated and analyzed in this study. Different methods of estimation were suggested to estimate the model parameters. However, application of this model to Malakal flows and Lake St.Clair NBS data did not show any advantage over simpler GNN.

A multivariate contemporaneous GNN (CGNN) model was also formulated and

tested. The covariance structure of the model was derived and different parameter estimation methods were suggested. The CGNN model can be used in situations were the sites to be analyzed have apparent shifts in the mean for all sites. The model was used to simulate the observed annual streamflow records at four stations in the Nile River basin and to simulate the NBS records of the Great Lakes System. In both cases, the CGNN model performed well in preserving both basic and long term statistics for these sites. In addition, a combined - contemporaneous GNN and contemporaneous CARMA models called GNN-CARMA was formulated so as to simulate apparent shifts at some sites and ARMA type of dependence at other sites. The model was applied to simulate NBS data for the Great Lakes System. The results showed the model was able to preserve the short and long term basic statistics such as the mean, standard deviation, rescaled range, and Hurst coefficient.

Nidhal Saada
Civil Engineering Department
Colorado State University
Fort Collins, CO 80523
Summer, 1998.

ACKNOWLEDGMENTS

The author would like to express his sincerest gratitude to his advisor Dr. Jose D. Salas, Professor of Civil Engineering, for his guidance, encouragement, and support during the authors's work on this research. The author would like also to express his thanks and gratitude to Dr. Vujica Yevjevich, Professor of Civil Engineering, for his valuable suggestions and comments throughout the research, Dr. Duane C. Boes, Professor of Statistics, for his assistance and guidance, and Dr. Jorge Ramirez for his encouragement and accessibility.

Thanks are also extended to Dr. Gustavo Diaz, Assistant Professor of Civil Engineering, Dr. Momcilu Markus, Chen Hua Chung and Huan Ming for their fruitful discussions and exchange of ideas and providing a friendly and cooperative research environment.

My deepest gratitude is extended to my parents who never stop providing me with their love and support throughout my life. This dissertation is dedicated to them. Thanks and gratitude are also due to my brother Said, my sisters Basima, Hala, Ghada, and Maisoon for their love and support.

My deepest and sincere appreciation and love is extended to my wife for her support and patience during the occasional tough times we had. Finally, my love is extended to my son Billal and my daughters Farah, Leena, and Sarah.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

1.1 General Remarks

Stochastic simulation of hydrologic time series is commonly applied for planning

and management of water resources systems. Stochastic simulation is usually based on

a stochastic model, i.e. a set of mathematical equations that describe an observed stochastic

process. The main applications of stochastic models are data generation (simulation) and

forecasting. Generation and forecasting are used by hydrologists for many purposes.

These include, for example, reservoir sizing, planning and management of an existing

reservoir, and reliability analysis of water supply systems (Salas et al, 1980).

The main criterion for choosing a specific stochastic model is the ability of that

model to preserve the statistical characteristics of an observed time series. Preservation of

such characteristics is meant to be in a statistical sense (Salas et al, 1980). It is difficult,

if not impossible to find a model that preserves all of the statistical characteristics of the

observed (historical) sample. The model should, though, be able to preserve the basic

statistics of the historical sample and some other statistics which are felt to be important

to preserve for the purpose at hand. For example storage and drought related statistics are

important for planning reservoir systems.

Many stochastic models have been used to model water resources systems. Some

of these models are: Autoregressive (AR), Autoregressive Moving Average (ARMA),

Fractional Gaussian Noise, Disaggregation, Broken Line, and Shifting Level models. The

1

shifting level models were introduced in water resources literature by Boes and Salas (1978) and Salas and Boes (1980). These models resemble shifting mean levels, a property that has been shown to be related to the Hurst phenomenon. With recognition that some observed hydrologic time series exhibit changes especially in the mean, shifting level models are attractive in modeling such time series. An important feature of the shifting level models is the shifts from one stage to another at random time periods. One particular type of shifting level model is known as the Geometric - Normal - Normal (GNN) model. In this study, further investigation of the GNN model will be done.

Stochastic modeling of hydrologic time series generally involve data observed at one or more sites. A stochastic model is called univariate if it uses only the observed time series at one site. However, the planning of water resources systems generally involve hydrologic time series observed at more than one site (multisite). This type of multisite system require multivariate stochastic models to take into account both the temporal and the spatial dependence. The procedures followed in multivariate modeling are basically the same as in univariate modeling. The mathematics of multivariate models is usually more difficult than for the univariate models (especially for the full vector multivariate models). However, some simple multivariate models known as "contemporaneous models" have been developed (Salas et al, 1985; Hipel and Mcleod, 1994).

1.2 Problem Definition

Hydrologists have noticed through the years that some observed hydrological data such as streamflow and precipitation had experienced significant shifts where the observed

2

data changes from say high values to lower values or vice versa. An example is the Great Lakes System. The mean of annual Net Basin Supply (NBS) series for Lake Ontario is 408,000 (cfs) for the period 1900-1970 while the mean is 508,000 for 1970-1990. In Lake Erie the mean of the annual NBS is 212,000 (cfs) and 321,000 (cfs) for the periods of 1900-1970 and 1979-1990 respectively. This change in the mean in both lakes was concurrent with an increase in the over basin precipitation for both lakes that seemed to happen around 1970. Many reasons have been proposed to explain these shifts in the annual NBS in the Great Lake System. Quinn et al (1990) reported that regulation, interbasin diversions, and land use changes are important reasons for explaining such shifts. Also, climate shift was suggested as a possible reason (Quinn et al, 1990). Kite (1990), in analyzing the Lake Superior levels at Duluth, found that a linear trend component is responsible for 20% of the total variability. Changes in climate were suggested as the reason for this trend. The prediction and forecasting on the Great Lakes System behavior are important in the planning and management of the lakes. Most of the models that are used in predicting and forecasting of net basin supplies, lake levels and outflows assume that the data is stationary. However, such an assumption may not be valid in the Great Lakes System. Therefore, other models which can simulate possible nonstationary characteristics may be useful in such cases. A more detailed review of some of the studies done on the Great Lakes System is included in section 2.4.

Another example is the sudden change that occurred in the mean annual flow volume of the Nile river at Aswan. The mean annual flow volume for the period 1870-1898 is 110×10^9 m^3 while it is 84×10^9 m^3 for the period 1899-1989. Also an upward

3

shift in the annual flows of the White Nile River at Mongalla and Malakal happened around 1961. A similar upward shift around 1961 is also apparent in the levels of Lakes Victoria, Kioga, and Albert (Salas and Boes, 1980). Another example is that of the Great Salt Lake in Utah where the water levels started to increase in the 1960's after a long period (1870-1960) of decreasing water levels (Kite, 1989). Likewise, in the Colorado River at Leez Ferry, Arizona, a change from high flows to lower flows had occurred near 1930 (Wallis, 1977). Potter (1976) showed precipitation records of six stations in the northeast region of USA that have shifts in their records and suggested that the shifts may have occurred because of natural climate shifts. However, in a later paper (Potter, 1979) he suggested that the shifts may have occurred from errors (inconsistencies) in the data. Furthermore, Hubert et al (1989) showed that the annual streamflow records of the Niger and Senegal rivers exhibit shifts. Several suggestions were proposed to explain this shifting phenomenon. Climate change, change in watershed boundary, and man made effects are the main proposed reasons (Wallis, 1977). Shifting level models may be useful in such cases as they can simulate such observed shifts (Salas and Boes, 1980). Also, shifting level models are operationally simple. Therefore, investigation of shifting level models is an important subject in hydrology and water resources that deserve further attention and investigation.

1.3 Research Objectives

The general objective of this research is to further develop certain types of Shifting Level processes for modelling and simulation of hydrologic processes exhibiting changes

in the mean. Specific objectives of this research are:

1. To examine and compare alternative methods for estimating the parameters of certain Shifting Level models and test their applicability by using the actual hydrologic systems and hydrologic data.

2. To formulate a multivariate Shifting Level model which can be used to model observed hydrologic records at more than one site including parameter estimation methods and testing.

3. To examine the applicability of the models to actual hydrologic data and compare the performance of such models to preserve the short term and long term characteristics of such data.

1.4 Dissertation Outline

Chapter 2 presents a review of the literature relevant to this study. The chapter begins with a brief history about the application of the stochastic theory and modeling in hydrology. Also, the chapter reviews some of the current models that are widely used in modeling water resources. Next, the Hurst phenomena is revisited and possible explanations of this phenomena that are available in the literature are reviewed.

Chapter 3 discusses the causes of the shifts that are observed in some hydrological processes. These causes will be defined and analyzed. Finally, some of the studies that have been done on observed hydrological processes that exhibit shifts are reviewed. Special attention is given to the examples of the Nile River System and the Great Lakes System.

Chapter 4 describes methods to estimate the parameters of univariate GNN model. Then simulation studies to verify and validate the GNN model are presented. Finally a new version of the GNN model is described. The multivariate shifting level models including parameter estimation methods and simulation studies to explore its applicability are discussed in Chapter 5. Then a summary of the conclusions of this research and recommendations for further research on this topic are included in Chapter 6.

CHAPTER II

LITERATURE REVIEW

2.1 General Remarks

Stochastic modeling of hydrological processes is commonly used by hydrologists and engineers to solve problems such as reservoir sizing, reservoir operations and evaluating the risk of failure of a water resources system. The application of the theories of stochastic processes has attracted the attention of hydrologists because of the random nature that is present in such processes. Hazen (1914) is believed to be the first one to introduce the problem of reservoir sizing in contexts of a random and uncertain process. In that paper, the so called probability paper, which has been used extensively in the field of statistical hydrology was introduced was introduced. Sudler (1927) published what is considered to be the first paper that introduced the idea of stochastic generation. In that paper, the observed annual runoff values were written on cards, shuffled, and then drawn one by one without replacement. Hurst (1951) in his efforts to find the required size of the Aswan Dam analyzed records of many geophysical time series. He used range analysis to find the required size. The importance of that paper from a scientific point of view was the discovery that the rescaled range of certain geophysical time series behaved differently than the usual time series available at that time. This discrepancy was called later the "Hurst Phenomena" and will be discussed thoroughly in the next section.

The formal development of stochastic modeling in hydrology occurred in the 1960's (Thomas and Fiering, 1962; Yevjevich, 1963) with the introduction of the

autoregressive (AR) models. Subsequently, extensive research has been done to improve the use of these models, introduce new models and investigate their theoretical statistical properties. Several stochastic models have been used for modeling the water resources systems (Salas et al, 1985). These are AR models, ARMA models, Fractional gaussian models, Broken line models and Shifting level models. Usually, the development and use of a stochastic model involve the following steps (Stedinger et al, 1982):

1- obtain hydrologic observed data such as streamflow data.

2- selection of models that describe the marginal probability distributions of the investigated observed data.

3- selection of an appropriate model of the spatial and temporal dependence of the data.

4- computer implementation of the suggested model and verifying that the model is working as specified.

5- validate the model by simulation.

6- use of the suggested model.

Since planning and management of water resources systems involve observed time series at many sites, multivariate stochastic analysis of such systems attracted the attention of hydrologists since the 1960's. Many multivariate models have been suggested to account for the temporal and spatial dependence of hydrologic data. Fiering (1964) proposed a two station model that preserves the historical lag-zero cross correlation and the lag-1 autocorrelation. Matalas (1967) suggested the multivariate AR(1) model that is applicable to several sites and preserves the lag-zero cross correlation and the lag-1

8

autocorrelation. O'Connell (1974) formulated a multivariate ARMA(1,1) model. Salas et al (1980) suggested the contemporaneous ARMA model which is a simplification of the multivariate model where the parameter matrices are assumed to be diagonal. The importance of the contemporaneous model is the uncoupling of the model equations such that separate models of different orders can be used for each site. The cross correlation between the sites is preserved through the noise term. Although the concept of multivariate modeling is the same as in univariate modeling, the mathematical formulation and operation of the multivariate models is more complicated when compared to univariate modeling especially for full-vector models (Salas et al, 1985). This difficulty was eased however by the development of the contemporaneous model mentioned above.

2.2 Hurst Phenomena

Hurst (1951) in his efforts to find the required size of the Aswan Dam in Egypt, analyzed many hydrological and climatological time series. What he found is , till today, one of the most controversial issues that faced hydrologists and statisticians. Hurst analyzed 690 annual geophysical time series such as streamflow, precipitation, temperature, and tree rings. For each series he plotted the logarithm of the sample mean of the observed values of the rescaled adjusted range $\bar{R}_N^{**}$ versus the logarithm of the length of the time series N divided by 2 and estimated the (Hurst) coefficient H as the slope of the straight line fit. Mathematically Hurst's fitted equation can be represented as:

$$\bar{R}_N^{**} = \left(\frac{N}{2} \right)^H \qquad (2\text{-}1)$$

9

Note that Eq. (2-1) can be generalized to

$$\bar{R}_N^{**} = a\, N^H \qquad\qquad (2\text{-}2)$$

The H values estimated by Eq.(2-1) had a mean value of 0.73 and a standard deviation of 0.092. In the same paper Hurst derived the asymptotic value of the mean adjusted range as

$$E(R_N') = a\, N^{0.5} \qquad\qquad (2\text{-}3)$$

where a = 1.2533. To verify his results he generated 30 sequences of size 100 of normally distributed independent variables and using the generated data he estimated the coefficient a close to 1.25 indicating that his derivation is probably true. This discrepancy between theoretical and practical observation is called "Hurst Phenomena".

Feller (1951) found that Eq. (2-3) does apply for any normally distributed IID variables with finite variance. Furthermore, Barnard (1956) and showed that Eq. (2-3) also applies for any short-memory time series model. Troutman (1976) showed that Eq. (2-3) is true also for any model with summable autocovariance function.

Since Hurst's early work, this phenomena (some times referred to as a puzzle) occupied the minds of hydrologists and statisticians as well. Mesa and Poveda (1993) stated:

"The Hurst effect is one of the most important unsolved problems in stochastic

10

hydrology. There is ample evidence to support this statement. Ever since Hurst's [1951] original work, there has been a proliferation of papers about it. Some of the most important awards of the hydrology community have gone to contribution towards its solution."

Several reasons have been suggested to explain this phenomena. These reasons can be generally classified into three categories. First, is that the Hurst phenomena is a pre-asymptotic or transitory behavior (Salas et al, 1979). The argument here is that the existing series are not long enough to have attained the asymptotic value of $H=0.5$. Second, the Hurst phenomena is due to the dependance structure of the time series (persistence) (Feller, 1951). The third explanation is that the Hurst phenomena is due to nonstationarity of the mean which was suggested by Hurst (1957), Klemes (1974), Klemes (1975), Potter (1975), Potter (1976), and Salas et al (1979).

Wallis and O'Connel (1973) claimed that transience and autocorrelation function are important factors in explaining the Hurst phenomena. Salas et al (1979) generated very long samples by using three models: ARMA(1,1), AR(1) and the Geometric-Normal-Normal (GNN) shifting mean model. The Hurst coefficient was shown to behave differently depending on the length of the series N. For small N the Hurst coefficient was greater than 0.5. In the pre-asymptotic transient region, the Hurst coefficient was much greater than 0.5 and in the asymptotic region (very large N) the Hurst coefficient converges to 0.5.

Feller (1951) suggested that the dependance structure of the time series is a possible explanation to Hurst phenomena. Matalas and Huzzen (1967) using Monte Carlo

11

experiments simulated 10,000 sequences of a Gaussian-Markov process for values of the lag-1 autocorrelation coefficient ranging from 0 to 0.9. Results indicated that the Hurst coefficient was found to be in the range of 0.58 and 0.87. These results are compatible with Hurst's findings for various geophysical time series. O'Connell (1974) suggested that an ARMA (1,1) may behave as a long memory model and may preserve the Hurst phenomena. To do that the AR coefficient must be close to unity. However, Klemes (1974) noted that O'Connell's model is very close to being non-stationary. Furthermore, Hipel and Mcleod (1978) showed that a properly fitted ARMA (1,1) model can preserve the Hurst phenomena. Gomide (1975) analyzed the AR (1) model with lag-1 autocorrelation coefficients in the range of 0.0 to 0.90 by using simulation experiments. His results indicated that the mean Hurst coefficient was 0.75 which is very close to Hurst's findings. Mandelbrot and Wallis (1969) and Mandelbrot and Van Ness (1968) suggested that the Hurst phenomena can't be explained by short memory models since for these models the asymptotic behavior for the range for these models follow Eq. (2-3) (i.e. $H \rightarrow 0.5$ as $N \rightarrow \infty$). As a result, they suggested that a long memory model can preserve the Hurst phenomena. The theory behind suggesting long memory models to model geophysical processes is the hypothesis that such processes are driven by a very long memory mechanism for which the effect of the distant past is not negligible. This is the main difference between long and short memory models. In the latter models, the effect of the distant past is identified but is negligible. As a result, Mandelbrot and Van Ness (1968) proposed the fractional gaussian model for modeling hydrologic processes. The important characteristic of that model is its very long memory feature. The model was

12

shown to be successful in preserving the Hurst coefficient of an observed time series. However, the model received a lot of criticism because of the lack of physical justification. Many hydrologists questioned the assumption of long memory in hydrological processes. Klemes (1974) stated that:

"By what sort of physical mechanism can the influence of, say, the mean temperature of this year at a particular geographic location be transmitted over decades and centuries?".

The nonstationarity of the mean to explain the Hurst phenomena was first suggested by Hurst himself (Hurst, 1957). He performed four experiments with playing cards that involved shifts in the mean. For each experiment, 1000 numbers were generated using these cards. Results indicated that the Hurst coefficient was calculated as 0.7 with a tendency towards 0.5 as N approached 1000. He concluded that the generated time series were very similar to the observed geophysical series he analyzed in his first paper in 1951. In this regard, Hurst is considered as being the first scientist to suggest that shifting means can explain the Hurst phenomena. Klemes (1974) analyzed the possible causes for the Hurst phenomena. He criticized the long memory hypothesis as a reason to explain the Hurst phenomena and showed that some non-stationarity models with no memory (zero memory) can explain the Hurst phenomena. Klemes, like Hurst, suggested that the non-stationarity of the mean can explain the Hurst phenomena. In that regard, he suggested possible physical reasons which support the non-stationarity of the mean hypothesis. He suggested that earthquakes, tectonic movements in the earth's crust, erosion and sedimentation, changes in vegetation cover and man made changes can have some effects

13

on the mean of many hydrological processes and thus supporting the non-stationarity of the mean hypothesis. Klemes also used simulation experiments to test weather the non-stationarity of the mean can explain the Hurst phenomena. Random samples with periodic changes in the mean and with randomly alternating means were generated and analyzed. Results showed that for such processes, the Hurst coefficient is greater than 0.5 and asymptotically approaches 0.5 as $N \rightarrow \infty$.

Potter (1975) and (1976) joined Hurst and Klemes in proposing the shifting mean hypothesis as a possible explanation for the Hurst phenomena. He showed six precipitation records that apparently exhibit shifts in their means. Like Klemes, he used simulation experiments with shifting means and reported similar results to those obtained by Klemes. Boes and Salas (1978) extended the works of Hurst, Klemes, and Potter and formulated a mixture model for shifting levels. They showed that Hurst, Klemes, and Potter models are all special cases of a generalized mixture model. One of the models that belong to the family of the shifting level models is the Geometric-Normal-Normal (GNN) model. Salas and Boes (1980) analyzed this model and suggested the method of moments to estimate the model parameters. Their estimates depend on the sample variance and the sample lag-1 and lag-2 autocorrelation coefficients. Also, Salas and Boes claimed that the apparent presence of shifts in the mean in some hydroclimatological data is attributed to phenomenological reasons rather than due to non-homogeneity or inconsistency of the data. The GNN model was also shown to have some advantages over corresponding ARMA (1,1) models in preserving the run and deficit properties of an observed time series. Obeysekera (1980) analyzed the run and range properties of the GNN model. Obeysekera

14

derived the distributions of the run and range properties of the model by using a discrete version and extending the results to the continuous case. He concluded that the GNN model is an attractive option to be used in modeling hydrological processes that exhibit shifts in their mean.

Mesa and Poveda (1993) presented several methods to determine if a time series exhibits the Hurst effect. They suggested that Pox diagrams alone are not sufficient and suggested GEOS (geophysical record diagrams), statistical tests, and empirical evidence tests should also be used. GEOS tests are visual tests for the existence of the Hurst effect in a given time series. Mesa and Poveda (1993) illustrated that these tests are more powerful than pox diagrams due to the fact that the GEOS diagrams are scaled down properly not only with respect to the adjusted range but also with respect to the variance and other moments as well. The main limitation of such tests is that the records must be long enough in order to draw any conclusions. The above tests were applied to different observed data among which is the data sets which were used originally by Husrt. According to their tests, the data do not show any existence of Hurst effect except for the mud curves. Also, they suggested that long memory does not appear to be a good explanation and that nonstationarity of the records is a more reasonable explanation. Sen (1991) illustrated that even though the annual flow series is stationary, the cumulative departures exhibit cyclic features that account for more than 95% of the variability in the Hurst coefficient. As a result, he proposed an alternative to direct modeling of time series in order to preserve the Hurst phenomena. His method is based on modeling the cumulative departures of the annual flows series from their sample mean value and

consequently simulation and data generation be done based on these departures rather than the original time series. Eltahir (1996) proposed a physically based explanation of the Hurst phenomena. The author suggested that ENSO (*El Niño* Southern Oscillations) events excites similar oscillations in the tropical climate, which are then teleconnected to the Nile flow through the rainfall producing mechanisms at the sources of the Nile (Eltahir, 1996). Eltahir developed a mathematical relationship between ENSO index and the Nile mean annual flow. He also developed a probability table for use in predicting the Nile flood. The table is based on classifying the mean annual flows as high, low, and average floods and an ENSO index as cold, normal, and average. He concluded by suggesting that the Nile flood responds to natural variability in ENSO which occurs at several time scales ranging from annual to decadal and even longer which causes significant nonstationarity in the mean of the annual flow process.

CHAPTER III

OBSERVED SHIFTS AND CAUSES OF SHIFTS IN HYDROLOGIC
PROCESSES

3.1 Observed Hydrologic Time Series That Exhibit Shifts

3.1.1 Nile River System

The Nile River is the longest river in the world. The river is 6825 Km long and

originates from lake Victoria and drains in Egypt . Figures 3-1 through 3-3 show the

historical mean annual flows at Aswan, Malakal, and Mongalla stations of the Nile River.

Figure 3-4 through 3-6 show the annual net basin supply (NBS) record for lakes Victoria,

Kyoga, and Albert. Some interesting features may be observed in these records. The

mean annual flows at Aswan for the period 1870-1989 are shown in Fig. 3-1. The lag-1

autocorrelation coefficient for the 1870 - 1989 record is 0.4 while it is 0.2 for the period

1914 - 1989. This change in the correlations is due to the shift in the flow regime

observed in the Aswan flows. The mean of the annual flows for the periods 1877-1898

and 1899-1989 are 110 milliards of cubic meters (mcm) and 84 mcm, respectively. An

upward shift in the annual NBS is noticed in Lakes Victoria, Kyoga, and Albert and in the

annual flows at Mongalla and Malakal. The upward shift was more noticeable in Mongalla

than in Malakal. The mean annual flows at Mongalla jumped from approximately 24.4

mcm for the period 1920-1960 to approximately 48.9 mcm for the period 1961-1983. In

Malakal, the mean jumped from approximately 26.6 mcm to approximately 35.3 mcm for

the same periods. The peak flow was measured in 1964 for both stations. Also Figs 3.1

and 3.2 show that another peak in the annual flows happened around 1917 and 1918 in both stations. This peak was, however, smaller than the one of 1964. This raises the question about the likelihood of experiencing other future considerable changes in the flows at both stations and the capabilities of the current models for simulating and forecasting such hydrological features. The lag-1 autocorrelation coefficient of the annual flows for Mongalla for the period 1914-1983 is 0.893. This is a very high value and is rare to find in other observed data. This is predictable , however, since high correlation coefficients can indicate that the data is experiencing some kind of grouping (periods of high flows followed by periods of low flows). The longest drought, at demand level equal to the mean annual flow is 42 years for the period 1914-1983. This value is very high and rare to experience in other stations that do not exhibit shifts. The rescaled adjusted range of the annual flows for Mongalla is 29.8 which can be categorized as a high value. The Hurst coefficient is 0.954 which is a very high value and probably is difficult to find in other stations. These statistics suggest that the shift that is apparent in Mongalla is quite significant. The main statistical characteristics for Malakal are similar to those for Mongalla. The reason for such high values for these two stations is the storage effect of the equatorial lakes.

One of the reasons proposed to explain the shifts in Lake Victoria levels and flows at Mongalla and Malakal, is that earthquakes in that region changed the hydrology of the Lake Victoria and caused this shift (Shahin, 1992). Shalash (1980) claimed that the operation policies at Owen Falls dam is the reason for this shift. In other studies, it was proposed that the Owen Falls operations were responsible for only 0.03 meters of the total

18

rise in the levels of lake Victoria (Shahin, 1992). Changes in land use have also been suggested as possible reasons (Sene et al, 1994; Salas and Boes,1980). Increases in precipitation was also suggested as a possible reason for the shift (Shahin et al, 1992). In a recent study, (Sene et al, 1994), applied annual and monthly water balance models to the lake's levels and showed that the variations in the levels are related to rainfall variations. One of the reasons suggested to explain the shift at Aswan flows that two different methods were used to measure flows in the two periods. From 1870 to 1902 the flows at Aswan were measured by a gauge calibrated by infrequent float measurements (Shahin, 1992). After 1902, the measurements were more direct and more advanced than the pre 1902 period. It was shown that the flows of the 1870-1902 period were overestimated by 8% (Shahin, 1992). Still, the flows of the 1870-1902 period are higher than those of 1902-1989 even if we subtract the 8% from the 1870-1902 flows (Shahin, 1992). *El Niño* was also suggested as a possible reason for the shift (Shahin, 1992). However, this theory was not proven conclusive since *El Niño* years were not always correlated (Shahin, 1992).

In a recent study, Eltahir proposed that ENSO is indeed related to the natural variability of the Nile river flows. Eltahir used the annual flow record of the Nile River and SST anomaly averaged over the regions $6^0 - 2^0$ N, $170^0 - 90^0$ W; 2^0 N - 6^0 S, $180^0 - 90^0$ W; and $6^0 -10^0$ S, $150^0 - 110^0$ W (ENSO index) for the period 1872 - 1972. The relationship between the Nile mean annual flows and the ENSO index is:

mean annual flow of the Nile = 88.5 - 8.7 (ENSO index)

The coefficient of correlation for the above relation is -0.9 (Eltahir, 1996). He also

suggested that the natural variability in the annual flows of the Nile can be decomposed into two components, namely the mean which is strongly related to ENSO and a random fluctuation that occurs around the varying mean due to climatic factors other than ENSO. The author claimed that the nonstationary mean and the random fluctuation explain 25% and 75% of the observed natural variability, respectively (Eltahir, 1996). He also used the Nile mean annual flows and classified them as low, average, and high floods and developed a probability table that relates such floods with SST. He suggested that mean of the Nile flood oscillates in time according to the oscillations of SST which exhibits, among others, annual, decadal, and longer oscillations. Finally, he concluded by suggesting that this relationship between the Nile floods and ENSO is a potential candidate for explaining the Hurst phenomena.

3.1.2 Great Lakes System

Another significant shift is the one observed in the Great Lakes system. Figures 3-7 through 3-11 show the annual NBS record for lakes Superior, Erie, St.Clair, Michigan-Huron, and Ontario. Low lake levels during 1930's and 1960's are observed in lakes Erie and Ontario. In contrast, high levels are observed in the 1950's, 1970's and mid 1980's. In 1986, a record high water level was set in all lakes except lake Ontario which caused major economic losses (Quinn et al, 1990). This record level was followed by a significant decline in the levels in 1987 and 1988 (Quinn et al, 1990). Many studies have been done in an attempt to find the possible reasons for these shifts in the lakes system. Quinn et al (1990) compared the record of 1986 with the 1838 lake levels. It was

20

found out that the 1838 record for lake Michigan - Huron is 50 cm higher than the 1986 record. Several reasons were suggested to explain this difference in the records between 1838 and 1986. These reasons are regulation, interbasin diversions and land use changes primarily deforestation and artificial drainage (Quinn et al, 1990). All these effects were found to cause some changes in the conditions between 1838 and 1986 but could not explain the 50 cm difference in the lake levels. Therefore, a possibility of climate shift that resulted in high precipitation was proposed to explain the high records in 1838 (Quinn et al, 1990). The recurrence of such climatic conditions could result in major flooding and huge economic loss. Quinn (1978) studied the effect of regulations on Lake superior. A hydrologic response model was used to evaluate these effects. Results showed that regulations have caused an increase in the lake Superior levels since the construction of the International Railroad Bridge and the Chandler-Dunbar power canal in 1901-1921. Bishop (1990) studied the lakes Erie and Michigan-Huron levels using the measured records, archeological evidence and historical documents. Maximum possible water levels were deduced from archaeological information of early European settlements as well as prehistoric Indian settlements. Also, water level data was also deduced from stratigraphic evidence and hydrologic simulations. By comparing the deduced data with the measured data since 1819, it was concluded that the water levels fluctuations for the last 1800 years are not different from the recorded data since 1819. Kite (1990) used spectral analysis to analyze some of the observed time series data. A linear additive model consisting of a linear trend, periodicity, and autoregressive components was assumed. The above model was applied to lake superior levels at Duluth, Minnesota. It was found that 97% of the

21

total variance can be explained by this model. The trend component was responsible for 20% of the total variance. It was suggested in that study that the existence of the linear trend is a result of "changes in climate" rather than a "green house" climate change.

Prediction and forecasting on the Great Lakes system is an important subject to water resources planners and decision makers. Stochastic analysis can be used in such cases to find these predictions and forecasts. Most stochastic models in operational hydrology rely on the assumption of stationarity of the data. If the data is not stationary, the usual practice in applied hydrology is to change such nonstationary data into stationary. This is done by removing all parameters which cause the nonstationarity like trend and seasonality. Privalsky (1992) examined the mean annual water levels of the period 1860-1988 of lake Erie at Cleveland and found that their mean, variance, and spectrum are non stationary and indeed are time dependant. The mean annual data was still nonstationary even after the linear trend was removed. Therefore, the data was best presented as a mixture of a linear trend and a non-stationary product random process with the stationary part represented by an AR process. It was found that study that with this structure the statistical predictability of the data was proven to be low. Irvine et al (1992) used multiplicative seasonal ARIMA models to the monthly mean level data for the period 1900-1986 of lakes Erie and Ontario. One, two, three and six month step ahead forecasts were obtained for the period 1984-1986 and were compared with the observed data of that period. Results show that the general trend of the data was reproduced for all the forecasting time steps. However, a three month was recommended as the maximum ahead steps in order to obtain good results.

22

3.1.3 Other Observed Data

An apparent shift can be observed in the annual flows of the Colorado River at Leez Ferry, Arizona. The observed annual flows at this station are shown in Fig. 3-12. The mean for the period 1914-1933 was 17 million acre-ft (maf) while the mean for the period 1946-1965 was 13.3 maf (Salas and Boes, 1980). Stockton (1977) used paleoenvironmental indicators to reconstruct flows for the past 450 years. He concluded that the apparent shift is a result of climate change. In contrast to Stockton claim, change in watershed boundary conditions was also proposed to explain the shift in the flows (Salas and Boes, 1980). Figure 3-13 shows the mean annual levels for the Great Salt Lake in Utah. The water levels started to increase in the 1960's after a long period (1870-1960) of decreasing water levels (Kite, 1989). The water levels rose to a record of 1283.77 m in 1986. This increase in the water levels resulted in an increase in surface area of the lake from 2500 Km^2 to 6500 Km^2. Consequently, the damages were estimated as $300 million. The rise in the water levels was caused probably because of the increase in precipitation within the basin (Kite,1989). The three years preceding 1986 were the wettest on record and the annual inflows from adjacent rivers to the lake were way above normal for these years. Karl and Young (1986) analyzed the Salt Lake basin from a meteorological point of view and claimed that such wet period has a return period of around 120 years and concluded that the occurrence of such periods is a result of climate fluctuations rather than climatic change (Kite, 1989). Potter (1976) showed precipitation records of six stations (Fig. 3-14) in the North Eastern part of USA that apparently have shifts in their mean and argued that most of the shifts in these records are significant. Hubert et al (1989) analyzed

the annual stream flow records of the Niger and Senegal rivers from 1900 to 1984 (Figs. 3-15 and 3-16 respectively). They suggested that the record can be classified into 5 different phases. The first , third and fifth are characterized by low flow periods whereas the other two periods are dominated by high flows. By analyzing the precipitation records in the basins of both rivers, Hubert et al (1989) concluded that climatological conditions which existed are responsible for the apparent shifts.

The United States Geological Survey (USGS) published a report called Hydro-climatic Data Network (HCDN) which is a collection of selected runoff stations in the USA for the period 1874-1988. These stations were selected on the basis to study the effects of climate variability on the observed runoff. The runoff data of these stations are available on a CD-ROM prepared by USGS. The selection criteria of the runoff stations can be summarized as follows: (1) Record length. Record length of at least 20 water years. All records of less than 20 years were not selected unless the record is uniquely located in an under represented geographic area. (2) Record accuracy. Stations are included if the mean daily discharge is rated at least good. The term good was defined as that 95% of daily mean discharge values are assessed to be within 10% of the true value. This rating will reflect the judgement of the office that obtained and prepared the discharge records. (3)Basin conditions. All records that have no flow diversions, augmentation, regulation of streamflow and reduction in base flow due to ground water pumping. The only exception is that when the above conditions remain unchanged throughout the record length. If a record is presently affected by the above activities and an earlier part of the record is not affected, then only the early part will be selected if it meets the other criteria.

24

(4) <u>Measured discharge values</u>. All discharge data measured by the standard practices followed by USGS. If due to any practical problem which prevented measurement, such as ice conditions, then the discharge will be estimated. If a large number of estimated values existed, then the station will not qualify for HCDN. Also, no station is included in the HCDN if the data at that station is constructed using information from other sources. Also no filling of missing data or extension of records was carried out in the HCDN. Based on these condition, 1659 stations in USA were selected and are available in the HCDN CD-ROM. In analyzing some of the records of the 1659 stations in HCDN, the following stations were obtained where they show an apparent shifts in their records.

1. Kalhi stream near Honolulu, Hawaii.

2. Green River at Green River, Utah.

3. Beaver River near Beaver, Utah.

4. Salmon river at White Bird, Idaho.

It should be pointed out that the author did not analyze all the 1659 stations available in HCDN. The author's purpose in analyzing some of the stations in HCDN was to show some examples of records that exhibit shifts. The purpose was not to find all the stations that exhibit shifts among the 1659 stations. The annual flow records of the above stations are available in Figs 3-17 through 3-20.

Turner (1996) studied the reliability of storage schemes with dendrohydrology and the Hurst phenomena. He suggested that the water supply planners should investigate whether a yield study period of the water supply storage schemes provides a representative sample of long term streamflow characteristics of the watershed. The author suggested

that the most recorded severe drought (1928-1934) based on observed streamflow records was also the worst drought in 421 years of reconstructed record using dendrohydrology in the Sacramento River. On the other hand, the most sever drought (1945-1951) in the Santa Ynez River is actually the ninth worst drought in 443 years of reconstructed record. Figure 3-21 shows the mean annual flows of these rivers which shows the variability of the mean of annual flows during consecutive periods of about 60 years. This suggests that the area had experienced different climatological conditions for the past 400 years which could explain the apparent shifts in these records.

Ropelewski and Halpert (1996) studied the relationship between precipitation amounts and ENSO in different regions of the world. Specifically, the authors attempted to quantify the relationship based on shifts in the statistical distribution of precipitation amounts with emphasis on shifts in the median, which are associated with the warm and cold phases of ENSO (Ropelewski and Halpert, 1996). The data used was the monthly precipitation totals for about 1500 stations in different parts of the world for the period 1901 through 1992. Figures 3-22 through 3-25 show some of their results. Figure 3-22 shows the precipitation distribution for northeastern South America. The authors suggested that for that region, a subtle shift in the precipitation amounts is associated with warm phase conditions of ENSO. In North America region, the authors suggested that the range of precipitation amounts between the 10th and 90th percentiles remain near the base period range of about 200 mm for the Gulf region (Fig.3-23) for both warm and cold phases (Ropelewski and Halpert, 1996). The Great Basin region of North America showed a shift toward wetter conditions in association with warm episodes (Ropelewski and

26

Halpert, 1996). The central Pacific, south-central Pacific, and Indonesia regions (Figs. 3-24 and 3-25) experienced dramatic ENSO related shifts in precipitations (Ropelewski and Halpert, 1996). The central Pacific median precipitation amount of 2650 mm was associated with the warm phase of ENSO. This amount is almost twice the base period median and more than three times the median amount associated with the cold phase as shown in Fig. 3-25 (Ropelewski and Halpert, 1996).

3.2 Causes For Shifts in Hydrologic Processes

Many geophysical time series have been shown to exhibit shifts in their records. Several reasons were suggested to explain these shifts. These reasons can be summarized as follows:

1- Climate

Nowadays, it is not possible not to hear about the effects of climate on our daily lives. The most recent example is the current increase in coffee prices all over the world which has been connected to climate conditions in South America. The Boston Globe (February, 1997) reported that this increase in coffee prices is the most increase since the 1970's and expected that the prices will continue to rise. Climate is one of the important elements which effects the geophysical time series such as precipitation and runoff records. There is a definite relationship between climate conditions and precipitation amounts which in turn will effect the runoff process. Much research have dealt with analyzing, exploring, and identifying relationships between climate and precipitation. Of importance to us are

27

the studies which dealt with possible climate conditions which can cause shifts in the statistical properties (such as the mean or median) of observed geophysical processes. Several studies had suggested that there is a consistent relationship between precipitation and *El Niño* events.   *El Niño* is a climate event associated with anomalous warmings of the western coasts of the South American continent especially Peru, Ecuador, and Chile. Quinn (1987) provided a detailed definition of *El Niño* as follows: "Originally, El Nino referred to the warm current that sets southward each year along the coast of southern Ecuador and northern Peru during the southern hemisphere summer when the southeast trade winds are weakest. It was named El Nino ("the child") by devout inhabitants of this region in reference to the "Christ Child" since it ordinarily sets in shortly after Christmas."

El Niño has become one of the hot topics that even the layman is aware of because of its climatic and economic impacts around the world. Barber (1988) illustrated the economic impact of the 1972-1973 *El Niño* on world market. That event was responsible for decreases in Anchovy Catch and increases in the soy consumption. This resulted in higher prices for soy around the world which resulted in permanent clearings of Amazon forest and North American wetlands for agriculture. As a result, investigation of *El Niño* which includes forecasting and predicting of *El Niño* events and the possible impact of such events on precipitation and agriculture is one of the important topics that not only interest scientists but also the average layman.

Ropelewski et al (1986) analyzed several temperature and precipitation records in different regions in USA, Canada, and Mexico to investigate the possible association

between temperature and precipitation patterns with the *El Niño* / Southern Oscillation (ENSO) events. The study showed that the above normal precipitation amounts in the regions of southern US and northern Mexico are consistently correlated with ENSO years (Fig. 3-26). Of the 22 ENSO events, 18 are associated with above norma precipitation in this region as shown in fig. 3-26. Similar findings were reported for the Great Basin area of the western USA (Fig. 3-27). However, in the high plains area, there was no evidence of association between ENSO and precipitation amounts in that region (Fig. 3-28). Temperature patterns were also found to be consistent with ENSO events in the areas of Alaska and western Canada. These regions experimented positive temperature anomalies during ENSO years. On the other hand, parts of southern USA experienced negative temperature anomalies in ENSO years.

In a recent study, Ting et al (1996) illustrated that there is an apparent consistency between north pacific sea temperatures (SST) and the precipitation over the Great Plains region. Specifically, the authors used a singular value decomposition between North Pacific SST and precipitation in USA for the period 1946-1988. They suggested that a wet year over the great plains is accompanied by above normal SST over eastern north pacific and below normal SST over the central north pacific. O'Brien et al (1995) suggested that the USA precipitation is largely associated with *El Niño* and El Viejo. However, Ting et al (1996) in studying the relationship between winter time telconnection patterns during extreme phases of the zonal mean circulation and the *El Niño* / ENSO events for the period of 1947-1994 suggested that the large mid-latitude natural variability may also contributes toward the anomalous U.S. precipitation.

Eltahir (1996) explored the hypothesis that the natural variability in flows of the Nile river is related to ENSO events by testing the relationship between the annual flow record at Aswan and the SST anomaly for the period 1872 through 1972. The author provided a mathematical relationship between ENSO and the annual flow record at Aswan. The author concluded by proposing the hypothesis that the mean of the annual flow process in the Nile River varies in time following ENSO, resulting in a non-stationary process and causing the Hurst phenomena.

The controversy over the responsibility of *El Niño* / ENSO to cause upnormal events has not yet been solved. Some scientists believe that ENSO itself may be changing due to long term rise in sea surface temperatures over the entire equatorial pacific (Science Bit, 1996). To put this issue in prospective, NCAR climatologist Gerald Meehl states (Science Bit, 1996):

"There is the possibility that the mean SST regime in the pacific has warmed up due to some kind of longer term fluctuation. In that case, you could have oscillations occurring that are superimposed on that warmer mean. If you compare these recent oscillations to the older mean, it's all going to look relatively warm, like a perpetual *El Niño* . If the average state of the climate system is undergoing longer term fluctuations, then what you define as *El Niño* or *La Niña* La Nina depends on what you take as the mean itself"

The long term changes in the climate can also supported by the argument that some regions which were covered with ice and snow millions or thousands years ago are now farm lands or in some cases changed to deserts. Turner (1996) showed that the reconstructed (using

tree rings) mean annual flows in the Sacramento and the Santa Ynez Rivers varies significantly for the past 400 years. Long periods of time (60 years) were shown to be wet when compared to other periods which can be considered as dry.

Several investigations of the possible effects of climate change on the hydrologic processes have been done. Flascka et al (1987) applied a water balance model to some selected watersheds in Nevada and Utah. Their results indicate that with an increase in the average annual temperature by 2^0 C coupled with a 10 percent decrease in precipitation would result in reducing runoff from 17 to 28 % of the present mean (Flaschka et al, 1987). Lettenmaier et al (1990) used a rainfall snow melt - runoff models to simulate runoff in four catchments in California. Climate scenarios produced from different GCM's with CO_2 doubling assumption were used to drive the rainfall snowmelt - runoff models. Results show that runoff will be shifted from spring to winter and also deliveries from the reservoirs will be lowered under the produced climatic scenarios. More recently, Gutowski et al (1993) used a physical model for severe hurricanes to study the effect of global warming on hurricane - induced floods in southeastern Florida based on CO_2 doubling climate change. Results show that under this assumption a 40% increase in precipitation for severe hurricanes would result in the studied area. As a result, the 100-year flood which is used for many water resources management in that area would increase by 10% and the flooded area would increase by 22%.

2-Other Factors

There are many other factors that can effect the geophysical time series. Watershed

31

conditions is an important factor which can have a great impact on the geophysical processes. There are many factors which will effect the response of a watershed. Geology, topography, soil type and condition, vegetation, degree of urbanization.....etc will effect the response of a certain watershed to an input such as precipitation. The watershed conditions always change with time and the degree of change depends on the intensity of the forcing factors and the unique characteristics of a certain watershed. Natural events such as earthquakes, land slides, fires ....etc may have a great impact on the response of a watershed and thus changing the hydrologic balance of this watershed. Scott D.F (1993) studied the effects of man-made fires in some catchments in south Africa. It was found in that study that, as a result of these fires, the hydrographs of post fire periods were higher and steeper. The duration, however, was little changed. The most important reason for the changes in the hydrological response of the studied catchments is the reduced infiltration which is caused by the water repellency of the soils of the burned catchments. Other factors that also contribute to the response of a catchment to fires are soil properties, degree of soil heating and vegetation.

Human activities when dealing with water resources systems have a major impact on these systems. Many activities will cause changes in some parameters of the watershed conditions and as a result some changes in the observed water resources time series might occur. The major activities by humans that can have significant effect on the water resources of any area are:

1-Reservoir or dam construction.

2-urbanization.

3-water use by agriculture or industry.

There are many purposes for building dams and reservoirs. Some of these purposes are flood control, water storage, power generation ...etc. These dams or reservoirs will have a significant effect on the distribution of the down stream flow. Flows which are released from them will be regulated. Seasonal variability of the original flows will be altered and changed because the dam or the reservoir will decrease the outflow in the flood season and increase it in the dry or low season. As a result, the temporal distribution of the outflow will be less variable and more uniform. Williams and Wolman (1984) reported an average decrease of 40% of the annual flows below 29 dams in the central and western united states.

Another effect of dams is that they trap sediment behind them and prevent the downstream movement of this sediment. This process will have significant effects on the size and shape of downstream channels and thus effecting the flows downstream (Hirsch et al, 1990). To study the effect of dams and reservoirs, the Platte River in Wyoming, Colorado and Nebraska is a good example. Major storage structures were built in 1909, 1938, 1939, 1941 and 1957 along the river. This major regulation had substantially decreased the annual instantaneous peak discharge at North Platte from about 511 m³/s before 1909 to about 72 m³/s after 1957. Also, decreases in channel width occurred at many cross sections along the river. In contrast, the mean discharges of the Green river, which is a major tributary of the Colorado river, were not affected by the construction of the Gorge reservoir in 1962 even though the sediment discharge was decreased by 21% (Hirsch et al, 1990). The construction of large reservoirs may also have significant

impact on the local climate conditions. With the existence of large reservoirs, the evaporation rates will be affected because of the existence of a large source of moisture for the atmosphere (Bates et al, 1994). This will possibly result in a local climate change and thus resulting in significant effects on the water resources of the area. In a study to evaluate the effect of the Great Lakes on the local climate of the lakes area, Bates (1993) used a local atmospheric model coupled with a global GCM to find that the existence of the lakes is responsible for 50-70 % of the precipitation amounts over the lakes area.

Urbanization has a significant effect on the response of a watershed and thus on the observed output (i.e. runoff) from that watershed. With urbanization, parts of the watershed will be changed into impervious areas. This change will result in decreases in the amounts of water that are usually lost due to the infiltration process and thus increases in the surface runoff will result (Hirsch et al, 1990). Also, as a result of urbanization, the runoff velocity will increase and the time of concentration, defined as the time between the center of mass of precipitation to the center of mass of runoff, of the watershed will decrease. As a result, the peak discharge will increase even if the total amount of runoff is constant (Hirsch et al, 1990). The effect of urbanization on smaller floods is greater than the effect on large floods (Hirsch et al, 1990). This is because for large floods, the soil will be usually saturated and thus infiltration losses will decrease and as a result the soil will behave as an impervious material.

Water use by industry or agriculture is another factor that will cause changes in the water resources of an area. This is because the use of water by these industries is increasing every year due to population increases and the higher standard of living of

people. Therefore, many water resources projects were developed to supply water needed by these industries and as a result, changes of the water resources of that area are expected. Also, because of human activities to cultivate land, significant changes in the watershed conditions could happen and as a result changes in the response of the watershed are also expected.

Besides the above three activities, there are many other human activities that could have significant effects on the water resources. Examples are interbasin transfers, land drainage, mining, man induced vegetation and channel alterations.
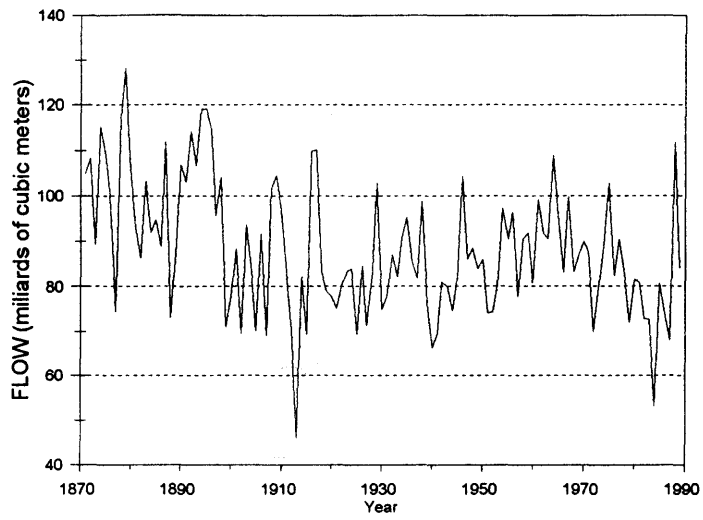
Fig 3-1   Historical annual flows for the Nile River at Aswan
            (1871 - 1989)
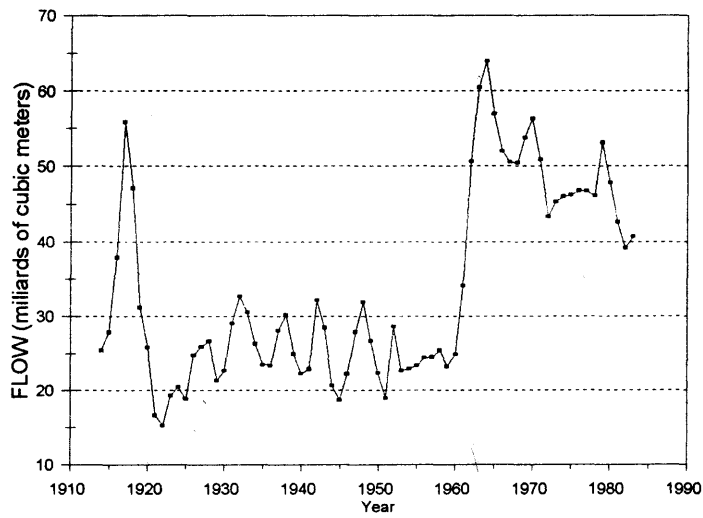
Fig 3-2   Historical annual flows for the White Nile River
            at Mongalla (1914 - 1983)

36
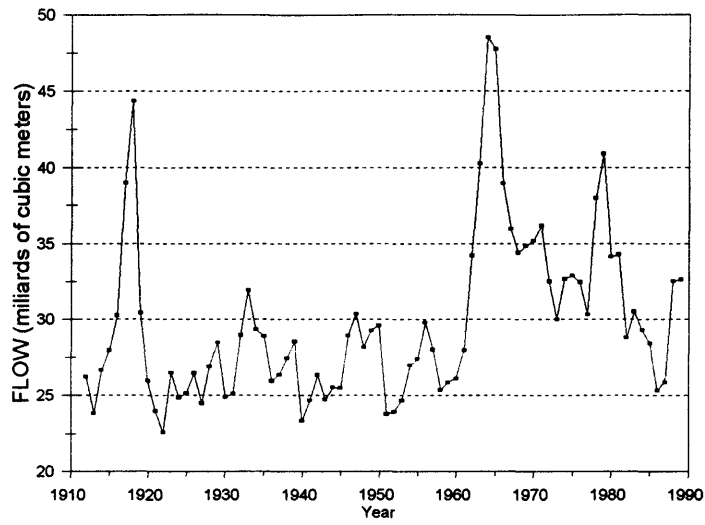
Fig 3-3   Historical annual flows for the White Nile River
           at Malakal (1912 - 1989)
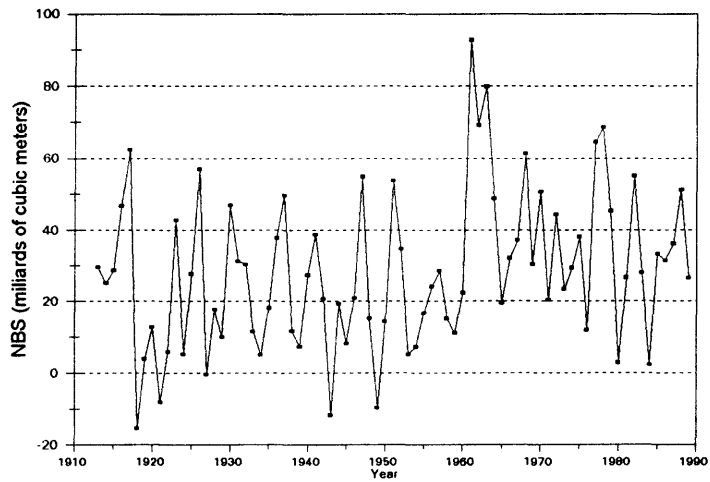
Fig 3-4   Historical annual NBS for Lake Victoria (1913 - 1989)

37
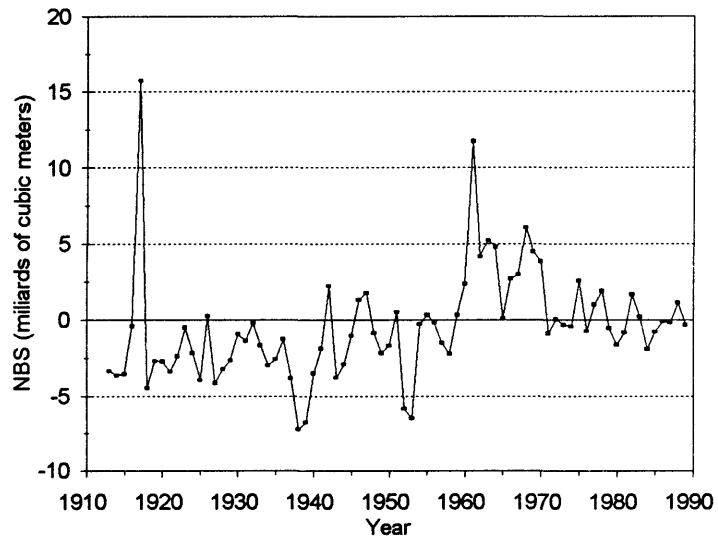
Fig 3-5   Historical annual NBS for Lake Kyoga (1913 - 1989)
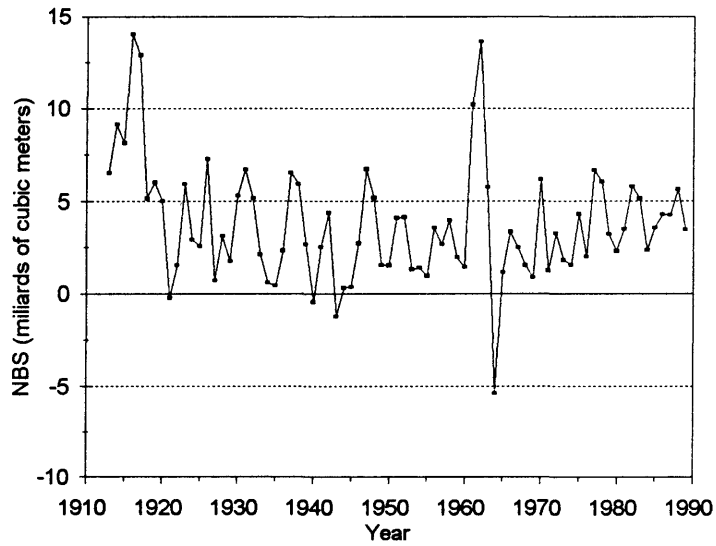
Fig 3-6   Historical annual NBS for Lake Albert (1913 - 1989)

38
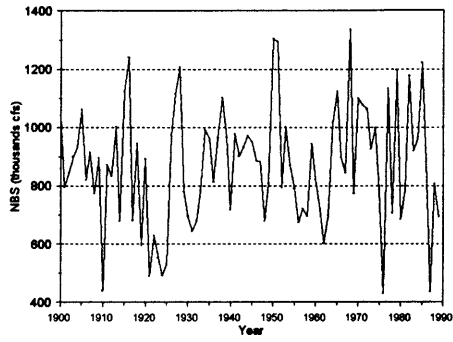
Fig 3-7   Historical annual NBS for lake Superior (1900 - 1989)
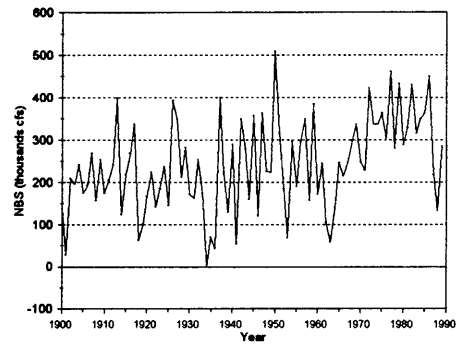
Fig 3-8 . Historical annual NBS for Lake Erie (1900 - 1989)
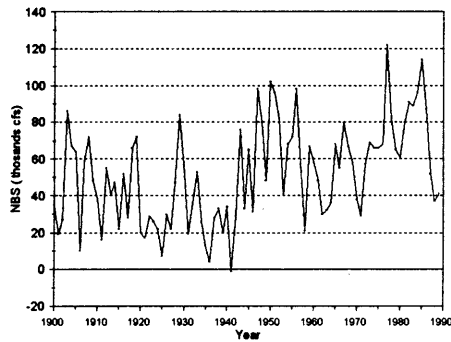
Fig 3-9   Historical annual NBS for Lake St.Clair (1900 - 1989)
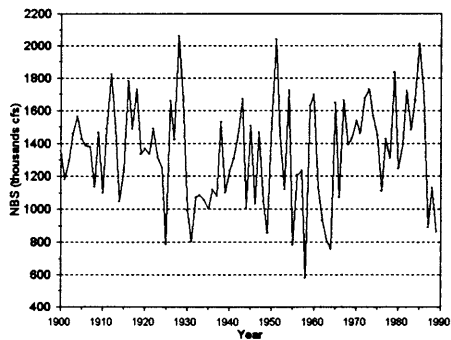
Fig 3-10 Historical annual NBS for Lake Michigan-Huron
          (1900-1989)
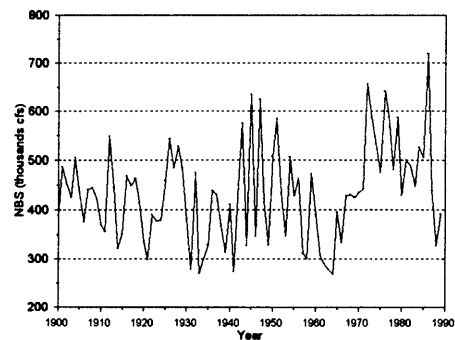
Fig 3-11   Historical annual NBS for Lake Ontario (1900 - 1989)
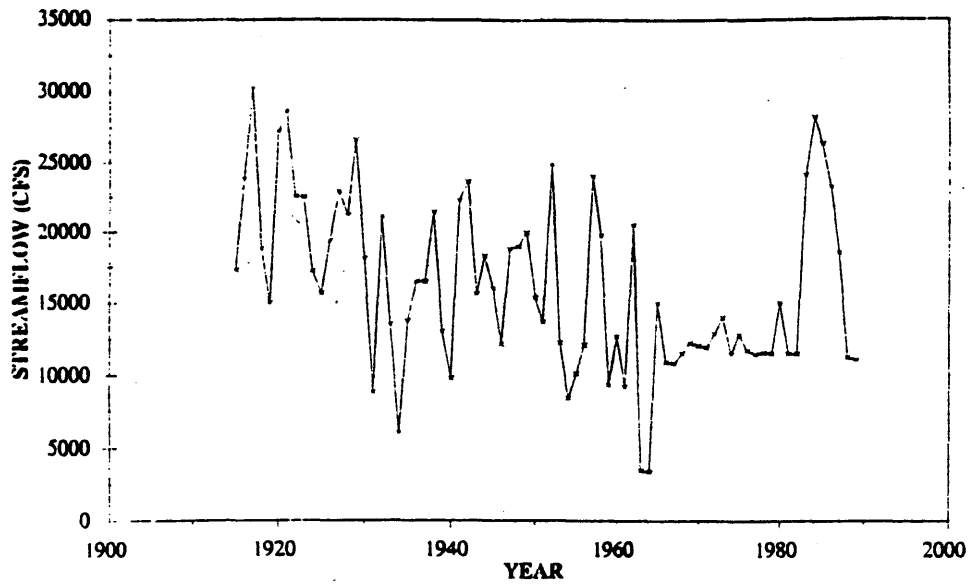
Fig 3-12 Historical annual flows of the Colorado River at Leez Ferry, Arizona for the period 1914 through 1920.
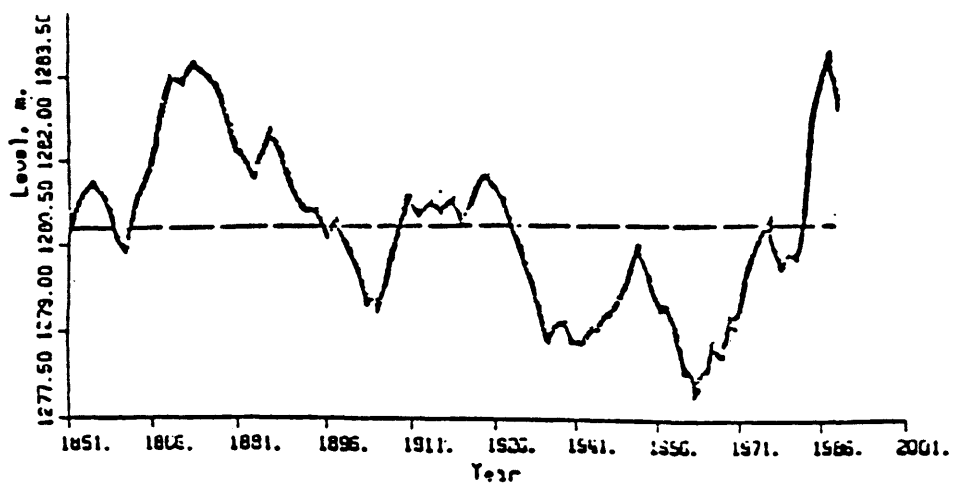
Fig 3-13 Historical mean annual levels for the Great Salt Lake, Utah. After Kite (1989).

40
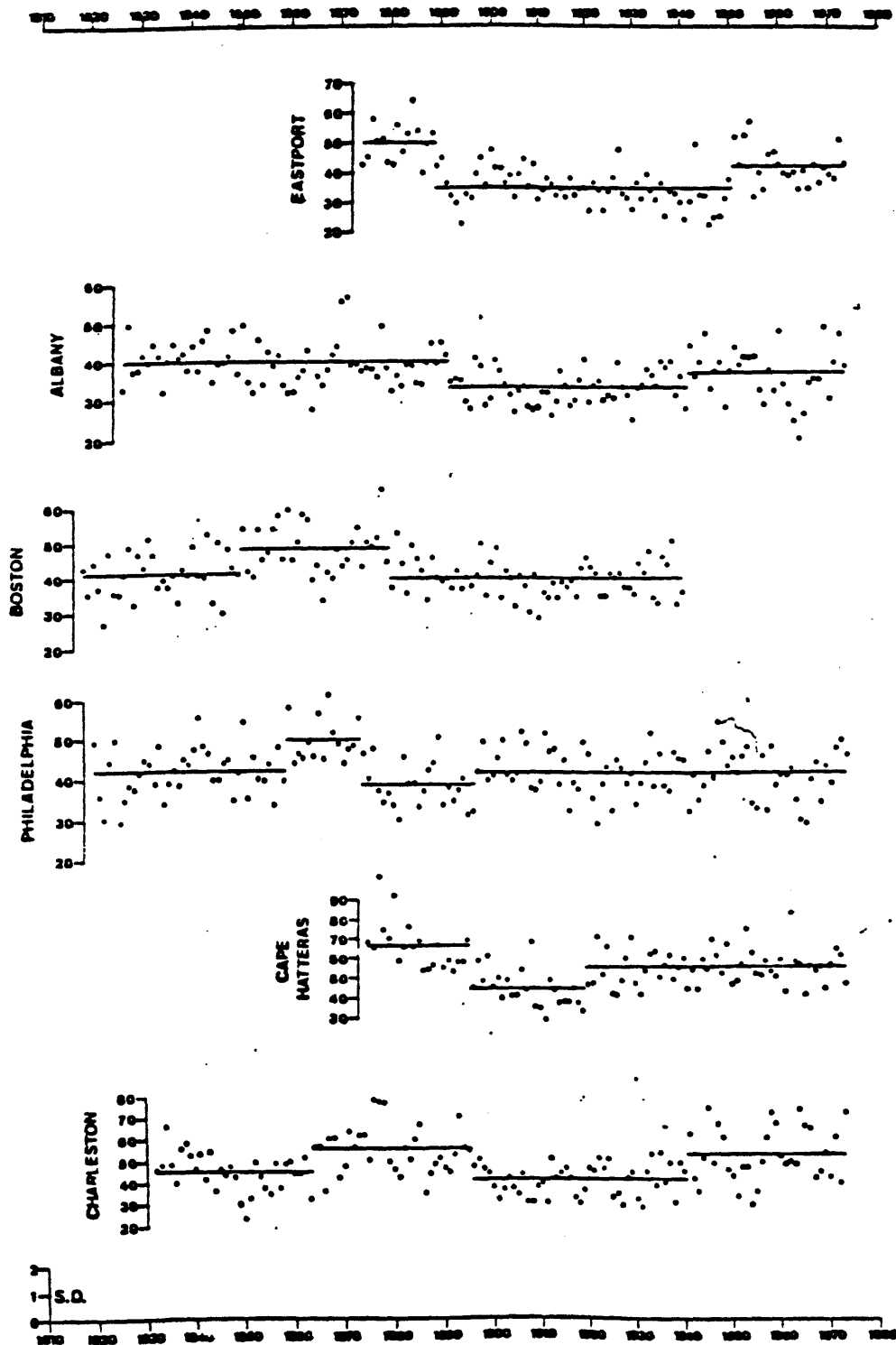
Fig 3-14 Precipitation Records (in inches) for six stations in the Northeast, USA.
After Potter (1976).   41
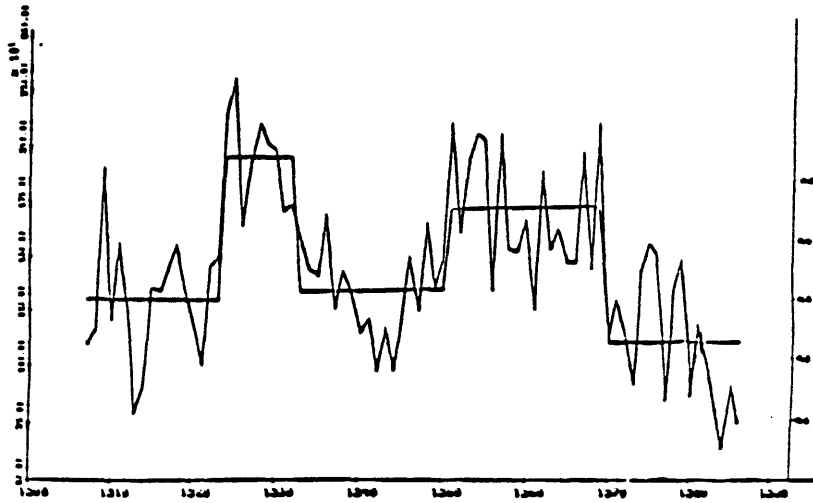
Fig 3-15 Historical annual flows (cfs) for Niger River at Koulikoro. After Hubert et al (1989).
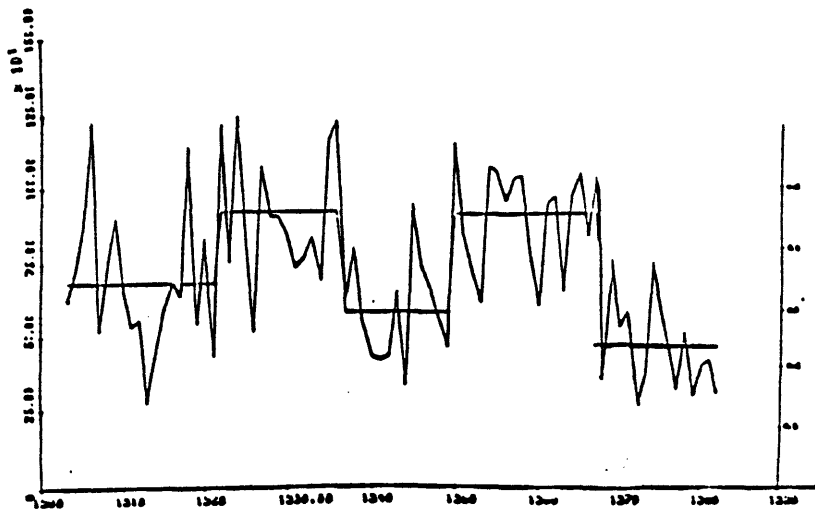
Fig 3-16 Historical annual flows (cfs) for Senegal River at Bakal. After Hubert et al (1989).
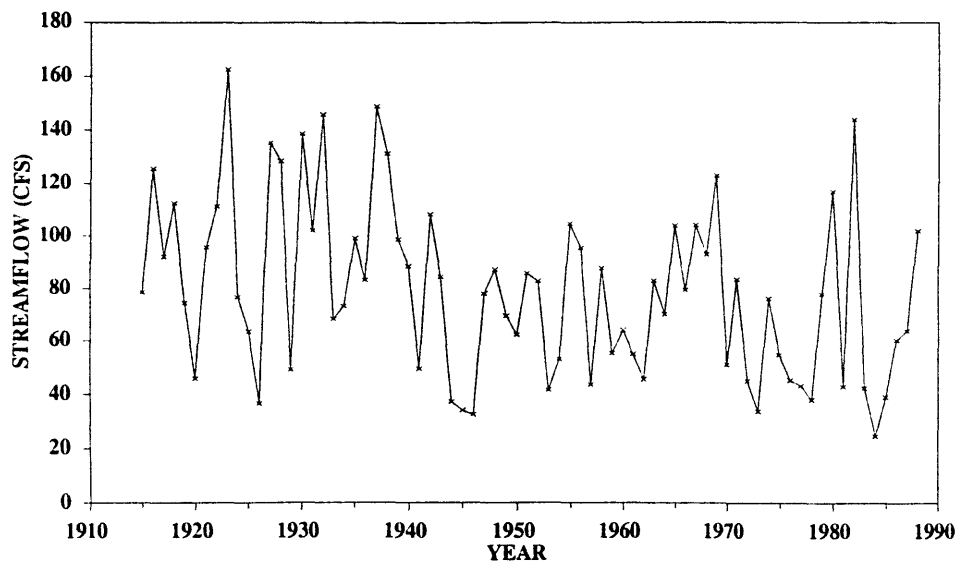
Fig 3-17 Historical annual flows for Kalhi stream near Honolulu, Hawaii.
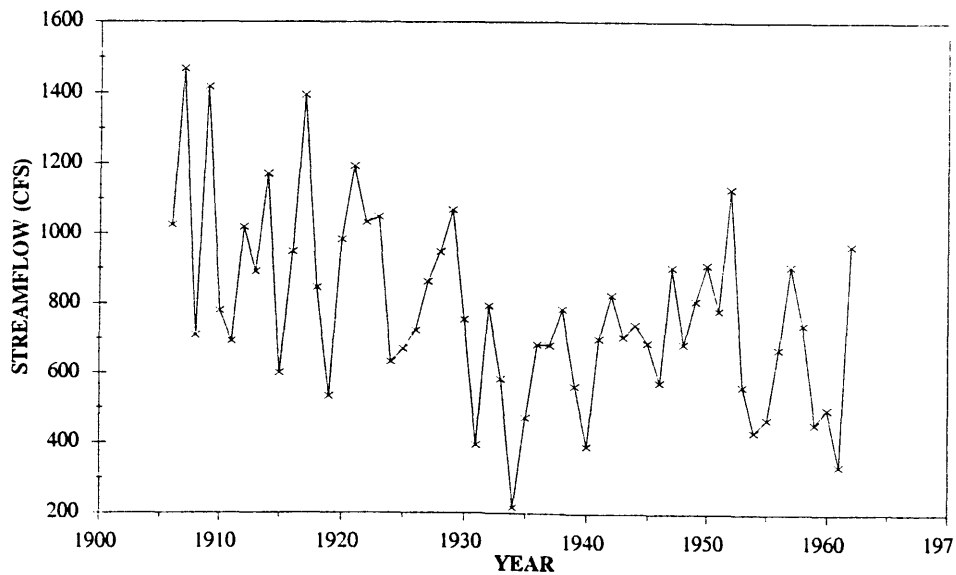
Fig 3-18 Historical annual flows for Green River at Green River, Utah.
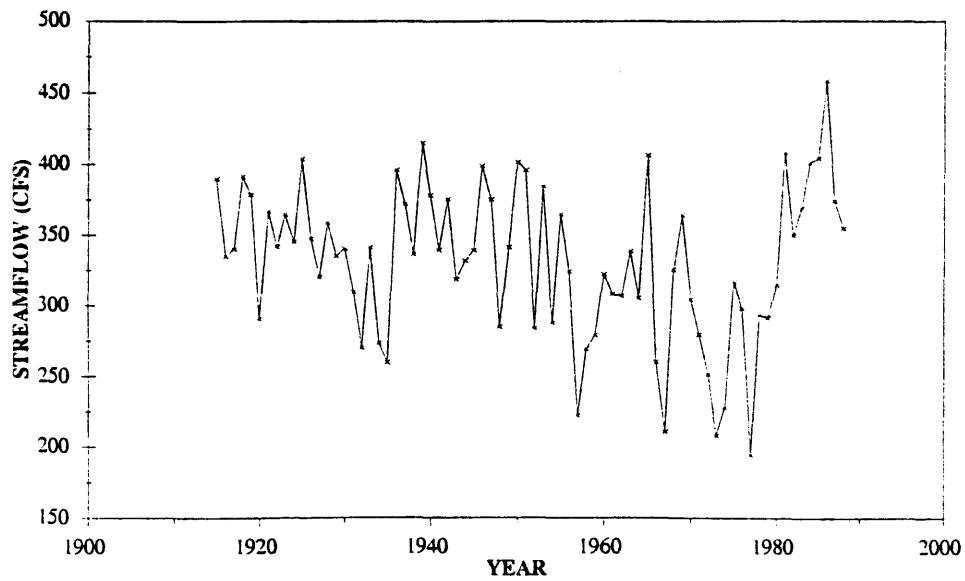
Fig 3-19 Historical annual flows for Beaver River near Beaver, Utah.
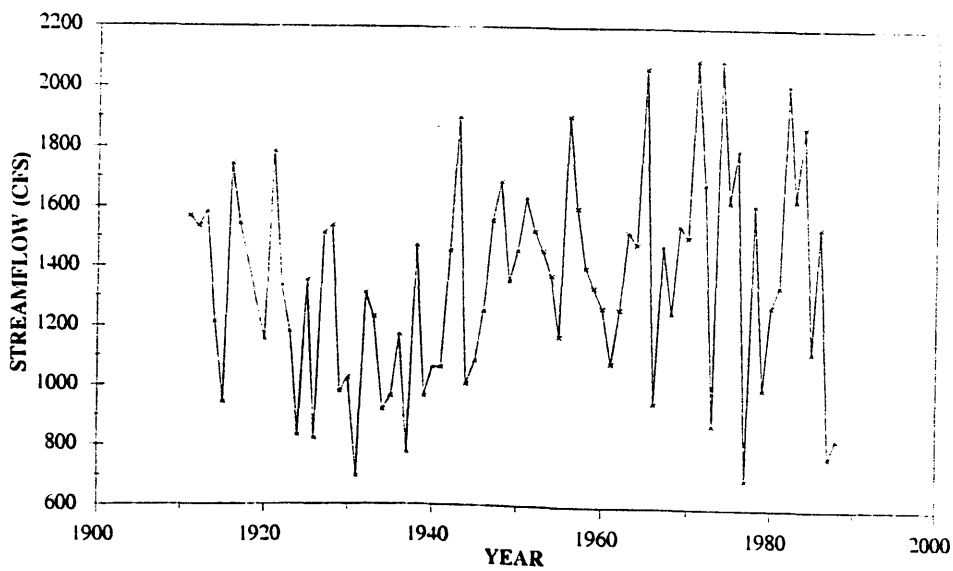
Fig 3-20 Historical annual flows for Salmon River at White Bird, Idaho.

44
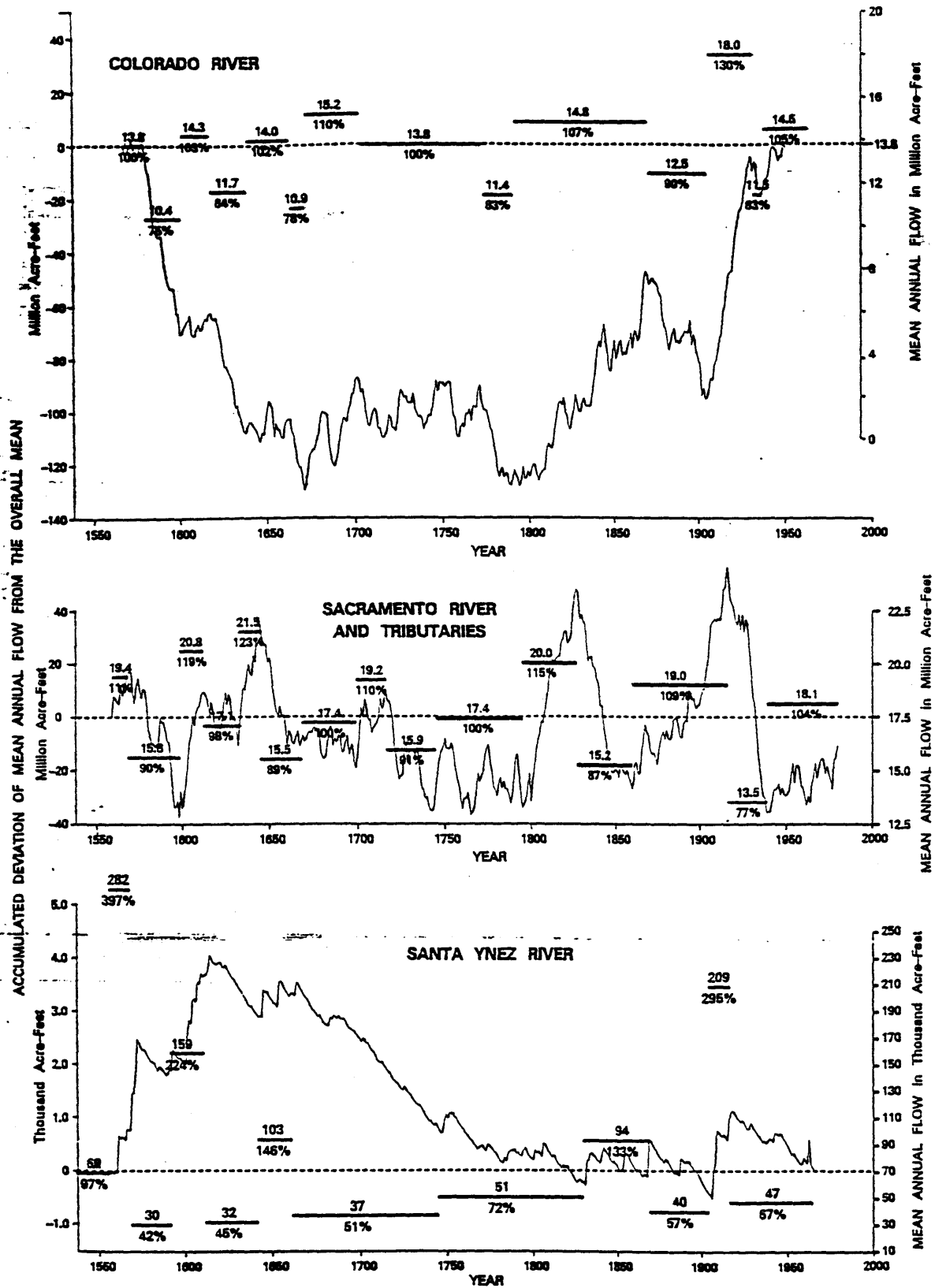
Fig 3-21 Accumulated Deviations of Mean Annual Flow from the Overall Mean, and Mean Annual Flows for Generally wet, Norma;, and Dry Periods from Dendrohydrologically Unimpaired Streamflows. After Turner (1996).

Northeast South America

Jul(0) - Mar(+)
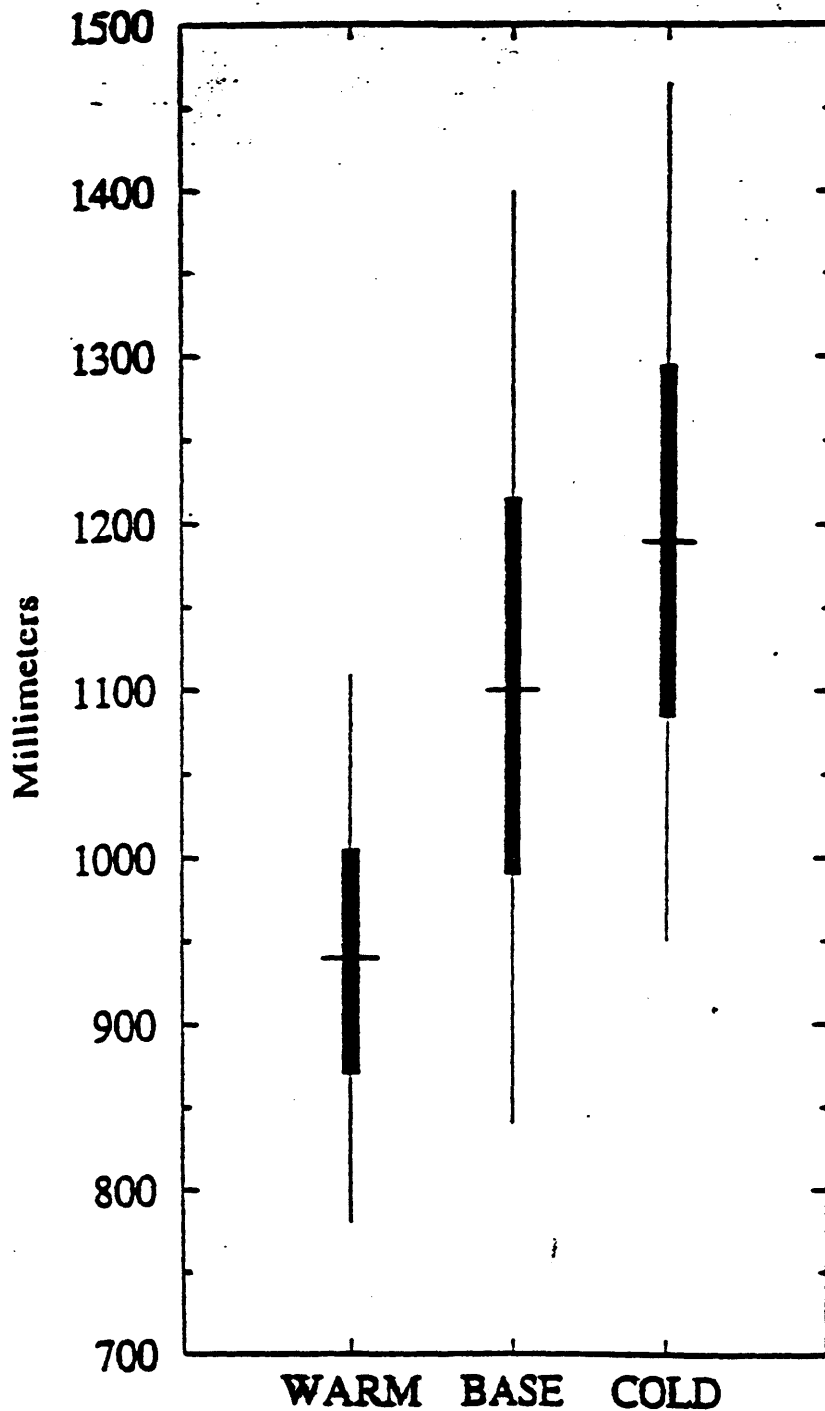
Fig 3-22 Precipitation distributions for (from left to right warm episodes, the base period, and cold episodes for northeastern South America. The horizantal line on each solid box represents median(50th percentile) precipitation amounts. Each solid box delineates the 70th (top) and 30th (bottom precipitation percentiles. The vertical line delineates the 90th and 10th percentile values. After Ropelewski and Halpert (1996).

Gulf and Northern Mexico

Oct(0) - Mar(+)
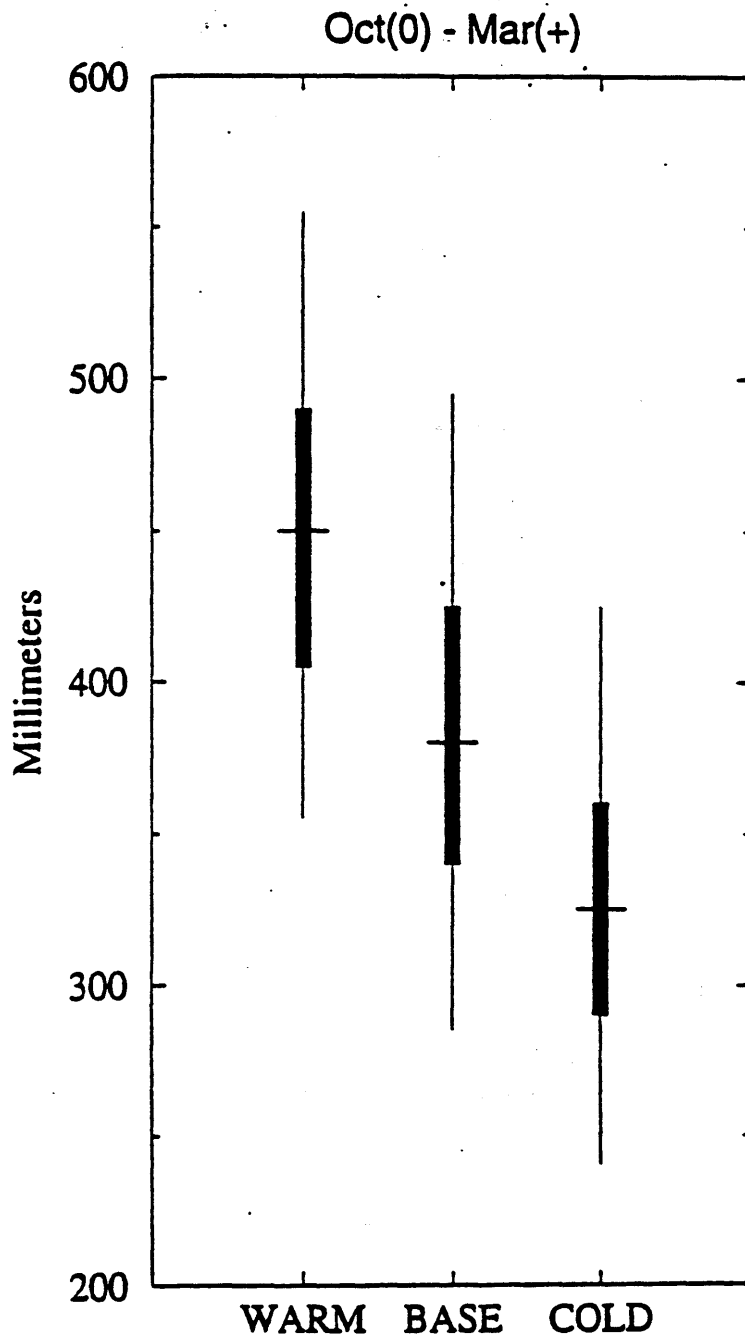
Fig 3-23 Precipitation distributions for (from left to right warm episodes, the base period, and cold episodes for northeastern South America. The horizantal line on each solid box represents median(50th percentile) precipitation amounts. Each solid box delineates the 70th (top) and 30th (bottom precipitation percentiles. The vertical line delineates the 90th and 10th percentile values. After Ropelewski and Halpert (1996).

Central Pacific

May(0) - Mar(+)
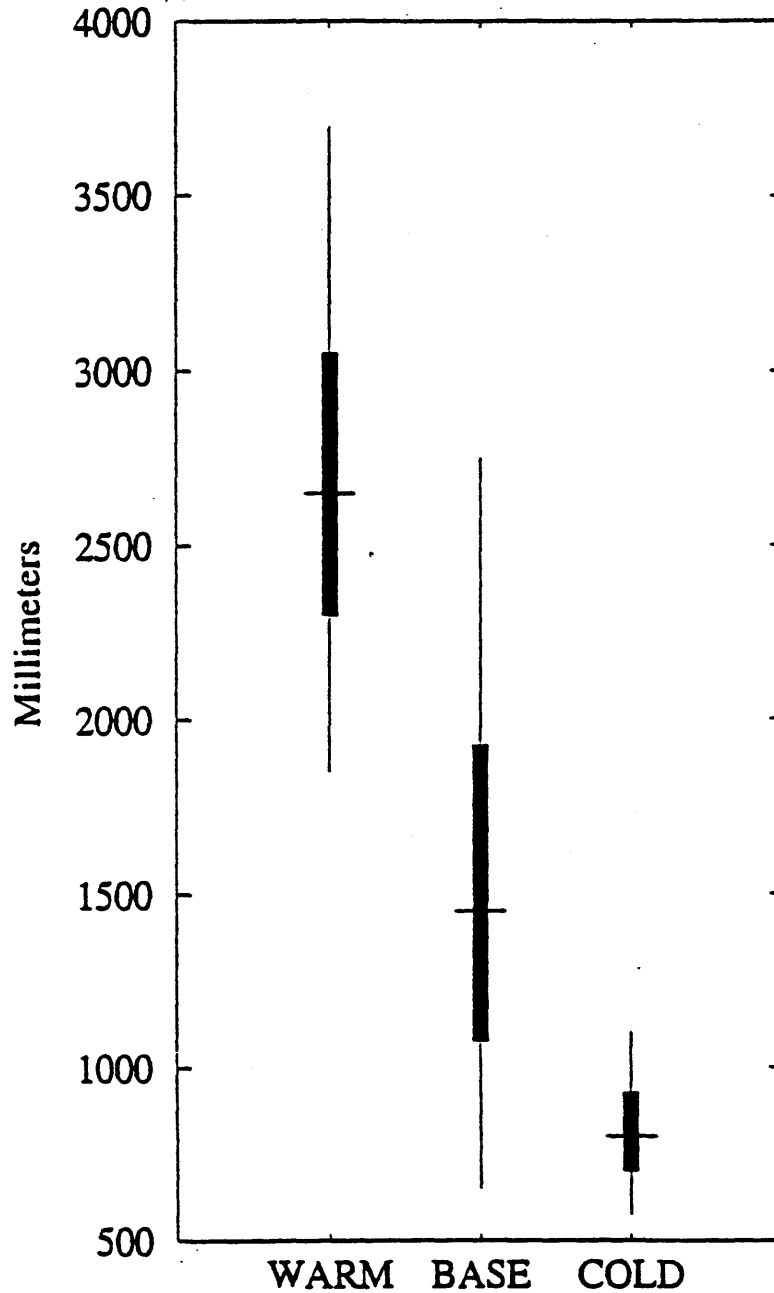
Fig 3-24 Precipitation distributions for (from left to right warm episodes, the base period, and cold episodes for northeastern South America. The horizantal line on each solid box represents median(50th percentile) precipitation amounts. Each solid box delineates the 70th (top) and 30th (bottom precipitation percentiles. The vertical line delineates the 90th and 10th percentile values. After Ropelewski and Halpert (1996).

48

Fiji/New Caledonia

Oct(0) - Mar(+)
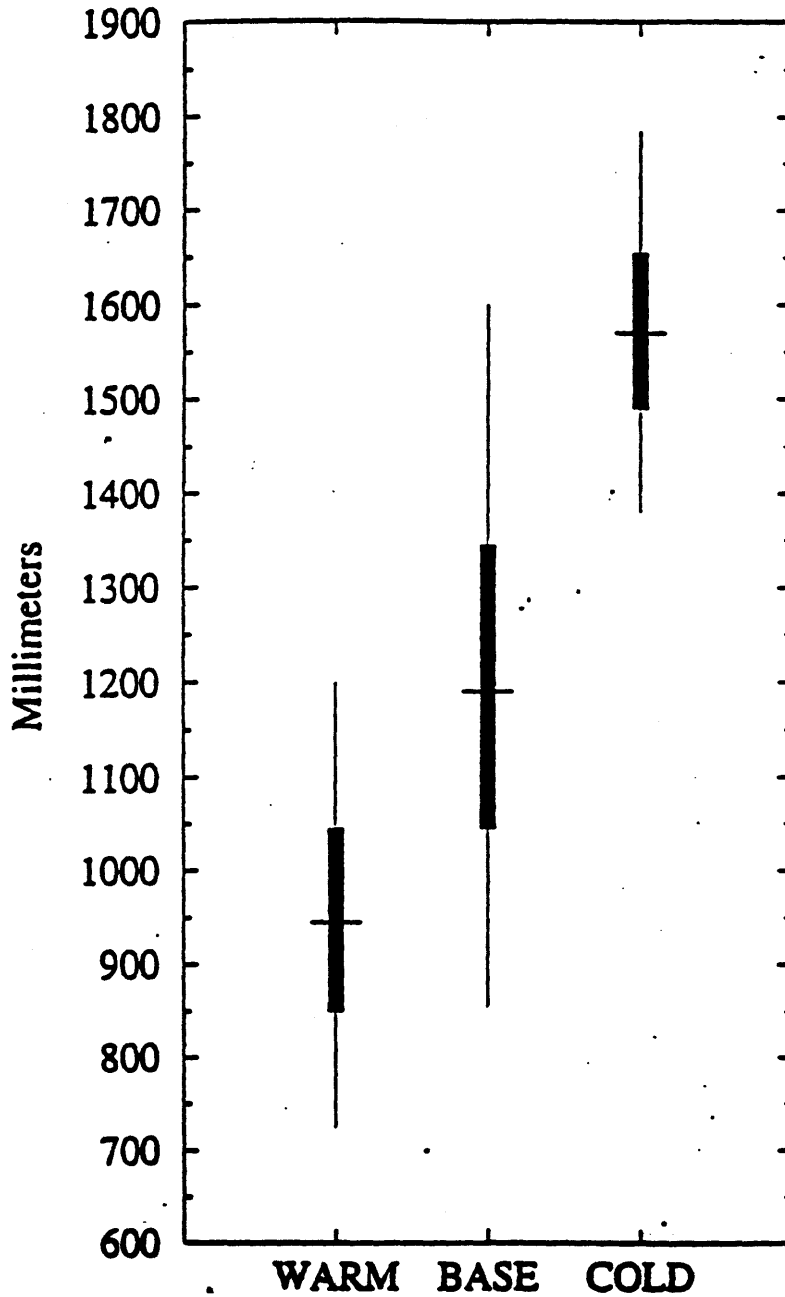
Fig 3-25 Precipitation distributions for (from left to right warm episodes, the base period, and cold episodes for northeastern South America. The horizantal line on each solid box represents median(50th percentile) precipitation amounts. Each solid box delineates the 70th (top) and 30th (bottom precipitation percentiles. The vertical line delineates the 90th and 10th percentile values. After Ropelewski and Halpert (1996).

PERCENTILE INDEX

100

75

50

25

0

1875 1885 1895 1905 1915 1925 1935 1945 1955 1965 1975 1985

GULF AND MEXICAN REGION
(OCT-MAR)
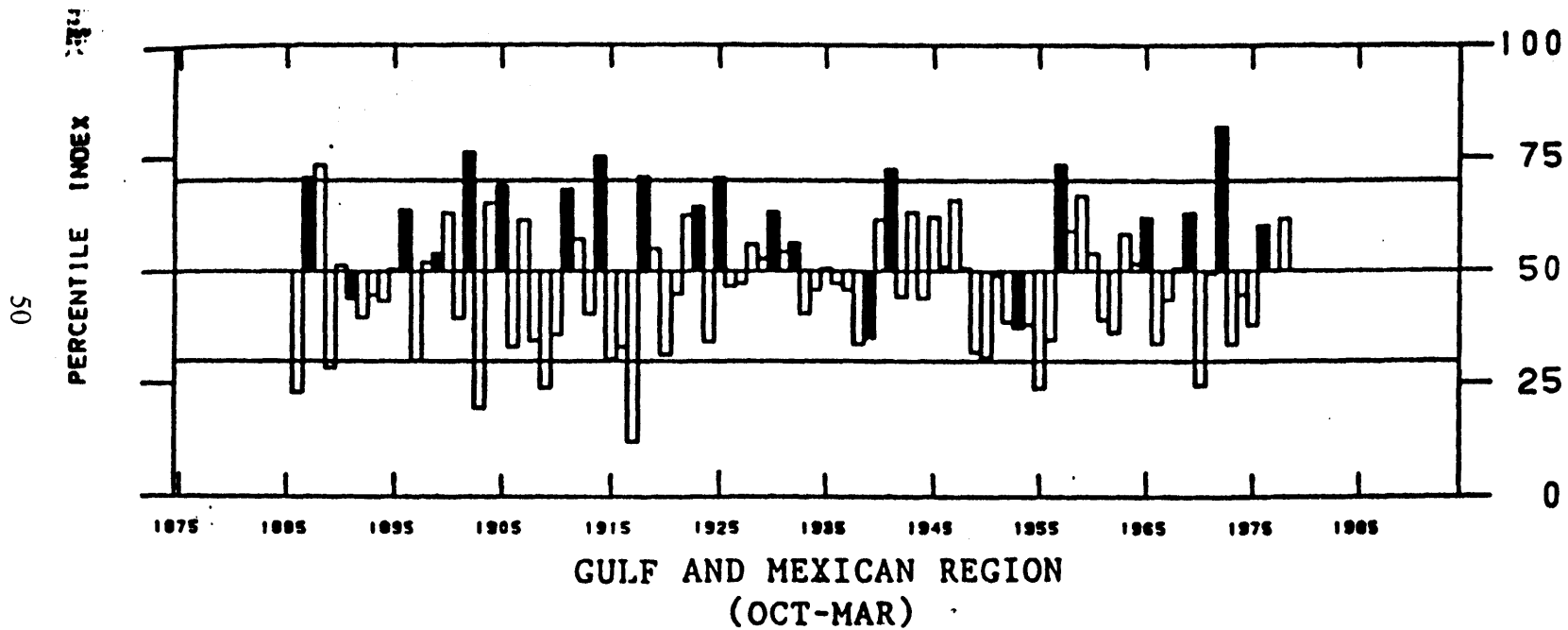
Fig 3-26 Time series of the GM region, i.e., Gulf and Mexican, precipitation for the October(0) to March(+) season. Precipitation is represented by the average of the precipitation percentiles for each of the stations within the area; ENSO years are represented by the dark bars. Of the 22 ENSO episodes shown here, 18 are associated with above normal precipitation in the Gulf area. After Ropelewski and Halpert (1986).

50
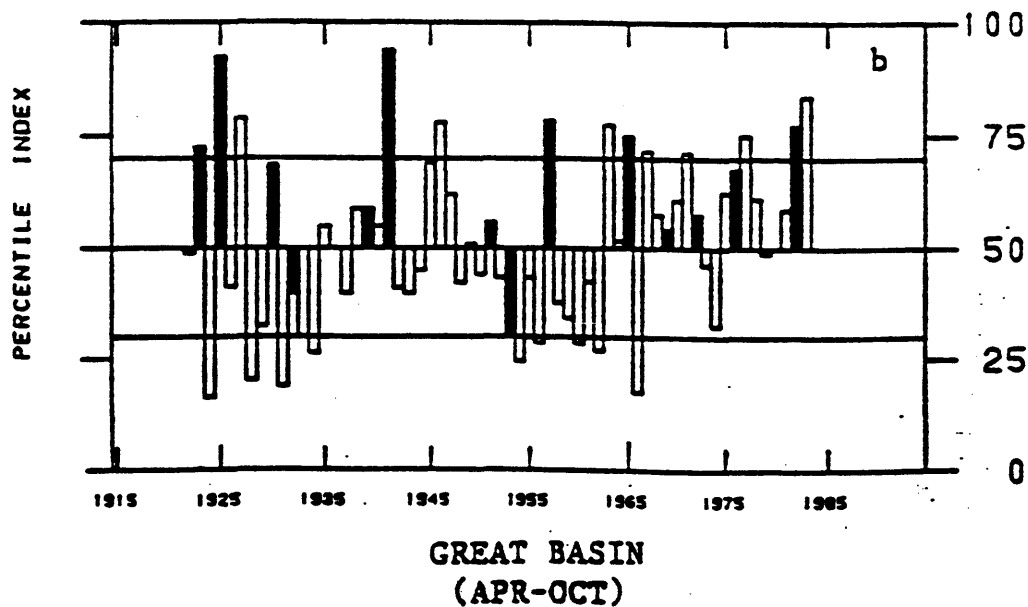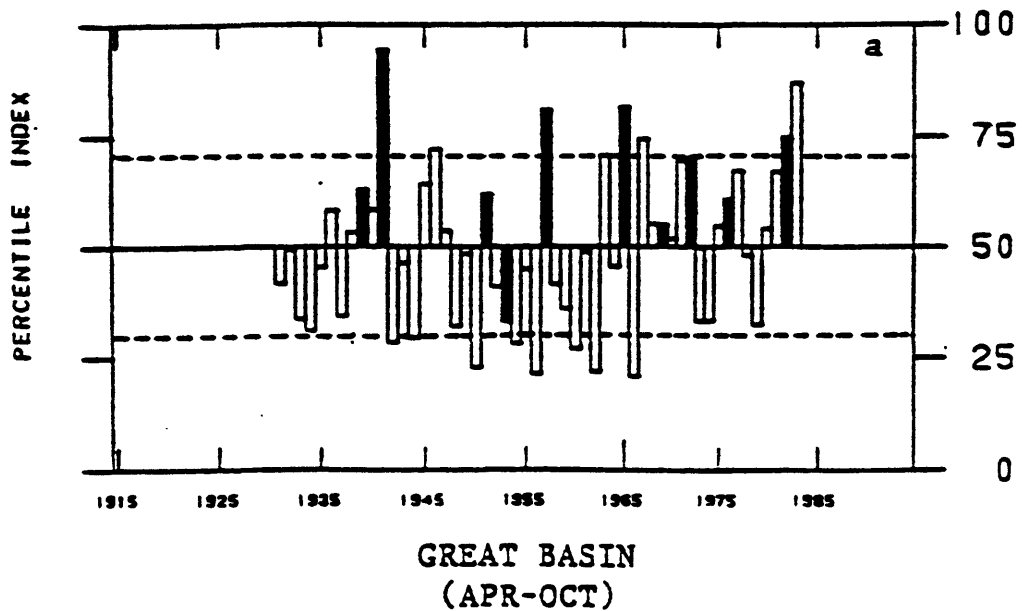
Fig 3-27 As in Fig 3.26 except in the Great Basin area for the April(0) to October)0) "season" based on (a) the climate division and (b) station precipitation. Of the 11 (14) ENSO episodes in the climate division (station) time series, 9 (12) were associated with above normal precipitation. After Ropelewski and Halpert (1986).

51
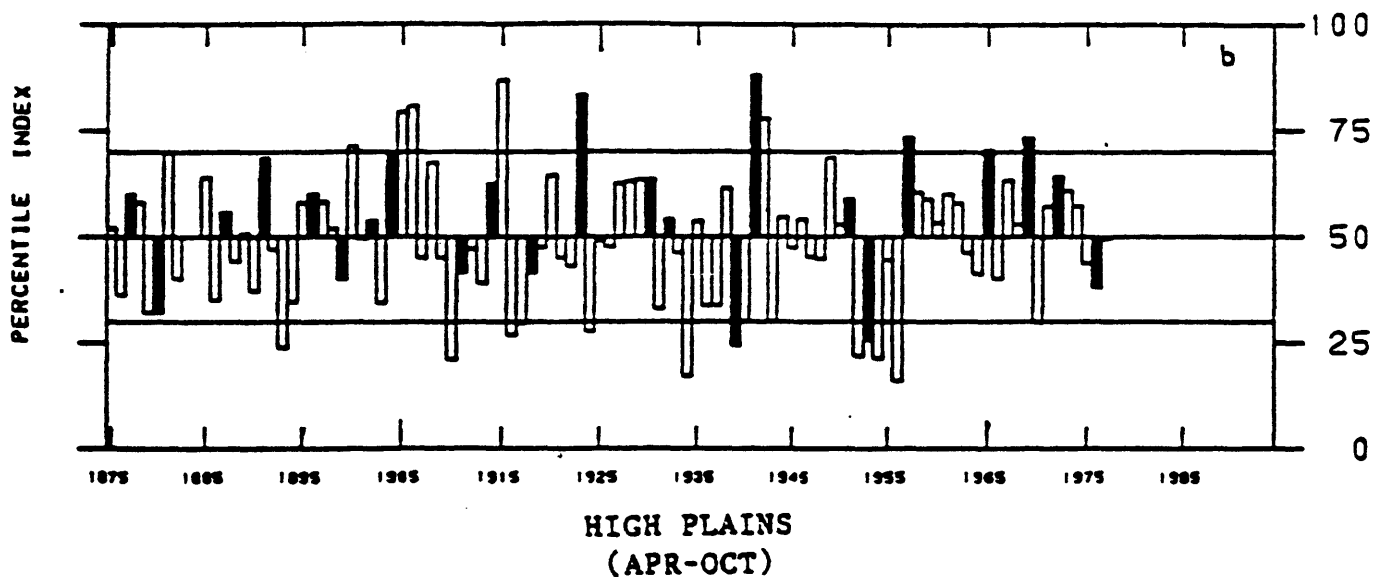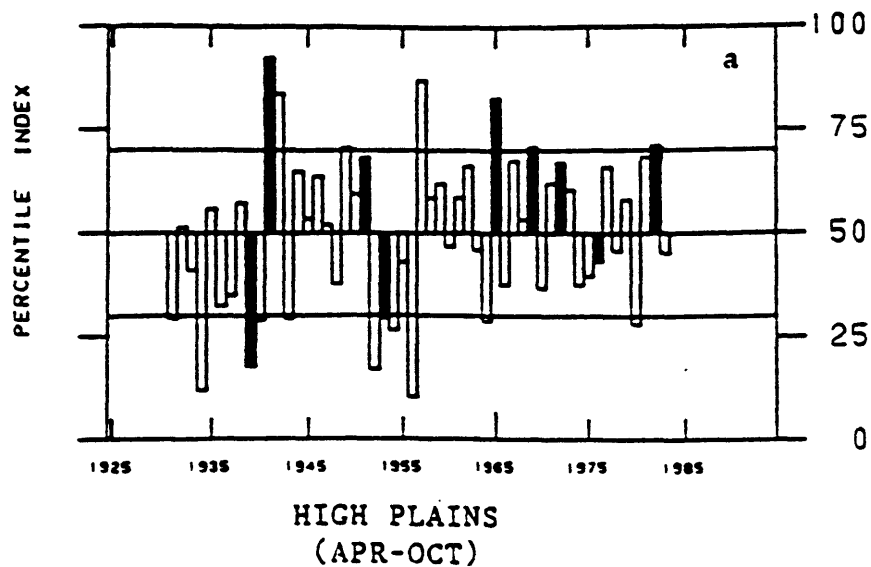
Fig 3-28 As in Fig 3.26 except for the High Plains region, based on (a) the climate division and (b) station data. A consistent ENSO-related response is not evident for this area. After Ropelewski and Halpert (1986).

52

CHAPTER IV

UNIVARIATE SHIFTING MEAN PROCESS

4.1 The GNN Model

In this chapter, the GNN model will be defined and formulated. Parameter estimation methods will be introduced and analyzed. Generation experiments will be done to test the performance of the estimation methods. This will be done by generating data from GNN using preselected population parameters. The sample parameters estimated from the generated data will be compared with the population parameters. Then, the GNN model will be applied to observed hydrologic data. The observed data will be fitted by appropriate GNN models. The parameters of the models will be estimated using the methods of parameter estimation mentioned above. Simulation studies will be conducted with the fitted models. This will be done by generating data using the estimated parameters. The generated statistics will be compared to the observed statistics to test whether the GNN model is capable of simulating series similar to the observed data. In addition, the GNN-1 model will be formulated and tested by simulation experiments.

4.1.1 Model Formulation

Let $\{N_i\}$ be a sequence of iid positive integer - valued random variables (Boes and Salas, 1978; Boes and Salas, 1980). These random variables represent the random time span of common means. Let $\{M_j\}$ be a sequence of iid random variables with mean $= \mu$ and variance $= \sigma_m^2$. These random variables represent the different means for the different time

spans. Let $\{Z_n\}$ be a sequence of iid random variables with mean=0 and variance= σ_z^2 .

These random variables represent the noise terms. Also assume that the sequences $\{N_i\}$, $\{M_j\}$ and $\{Z_i\}$ are independent. Now form the following series (Boes and Salas, 1978):

$$M_1 + Z_1 , M_1 + Z_2 , ... M_1 + Z_{N_1} ,$$

$$M_2 + Z_{N_1 + 1} , M_2 + Z_{N_1 + 2} , ... M_2 + Z_{N_1 + N_2} ,$$

$$M_3 + Z_{N_1 + N_2 + 1} , ... M_3 + Z_{N_1 + N_2 + N_3} , ...$$

The nth term of the above series is defined as: $X_n = M_n^* + Z_n$ where

$$M_n^* = \begin{cases} M_1 & \text{if } n \leq N_1 \\ M_2 & \text{if } N_1 < n \leq N_1 + N_2 \\ .. \\ .. \\ .. \\ M_n & \text{if } S_{n-1} < n \leq S_n \end{cases}$$

$$= \sum_{j=1}^{n} M_j I_{(S_{j-1}, S_j]}(n)$$

where

$S_j = N_1 + N_2 + ... + N_j$; $S_0 = 0$, and I is the indicator function. Now when $\{N\}$ is geometrically distributed and $\{M\}$ and $\{Z\}$ are normally distributed, the shifting level model is called the GNN model.

4.1.2 Moment Equations of The GNN Model

The GNN model can be expressed at time t as:

$$X_t = M_t^* + Z_t \qquad\qquad (4\text{-}1)$$

The mean of the X process can be found by taking expectation in Eq. (4-1)

$$E(X_t) = E(M_t^* + Z_t) = \mu$$

which suggests that the GNN model is stationary in the mean. Multiplying both sides of Eq. (4-1) by X_t gives

$$X_t X_t = M_t^* M_t^* + M_t^* Z_t + Z_t M_t^* + Z_t Z_t \qquad\qquad (4\text{-}2)$$

Taking expectation of Eq. (4-2)

$$E(X_t X_t) = E(M_t^* M_t^*) + E(M_t^* Z_t) + E(Z_t M_t^*) + E(Z_t Z_t) \qquad (4\text{-}3)$$

and since Z and M are independent, i.e. $E(M_t^* Z_t)$ and $E(Z_t M_t^*)$ are both equal to zero, Eq. (4-3) reduces to:

$$E(X_t X_t) = E(M_t^* M_t^*) + E(Z_t Z_t) \qquad\qquad (4\text{-}4)$$

In addition, since $E\left(M_t^* M_t^*\right) = E\left(M_t M_t\right)$ (Boes and Salas, 1978) and

$$E\left(X_t X_t\right) = \sigma_x^2 + \mu^2 ,$$

$$E\left(M_t M_t\right) = \sigma_m^2 + \mu^2 \text{ and}$$

$$E\left(Z_t Z_t\right) = \sigma_z^2$$

then from Eq. (4-4) it follows

$$\sigma_x^2 = \sigma_m^2 + \sigma_z^2 \tag{4-5}$$

which indicates that the GNN model is stationary in the variance.

Multiplying both sides of Eq. (4-1) by X_{t-k} gives

$$X_t X_{t-k} = M_t^* M_{t-k}^* + M_t^* Z_{t-k} + Z_t M_{t-k}^* + Z_t Z_{t-k} \tag{4-6}$$

Taking expectation of Eq. (4-6) and considering that M and Z are independent, it follows

$$E\left(X_t X_{t-k}\right) = E\left(M_t^* M_{t-k}^*\right) + E\left(Z_t Z_{t-k}\right) \tag{4-7}$$

Furthermore, since Z is uncorrelated , Eq. (4-7) simplifies to:

$$E\left(X_t X_{t-k}\right) = E\left(M_t^* M_{t-k}^*\right) \tag{4-8}$$

In addition

$$cov \ (M_t^* , \ M_{t-k}^*) \ = \ E \ (M_t^* \ M_{t-k}^*) \ - \ E \ (M_t^* \ ) \ E \ (M_{t-k}^*)$$

$$= \ E \ (M_t^* \ M_{t-k}^*) \ - \ \mu^2$$

$$cov \ (X_t , \ X_{t-k}) \ = \ E \ (X_t \ X_{t-k}) \ - \ E \ (X_t) \ E \ (X_{t-k})$$

$$= \ E \ (X_t \ X_{t-k}) \ - \ \mu^2$$

so that

$$cov \ (X_t , \ X_{t-k}) \ = \ cov \ (M_t^* , \ M_{t-k}^*) \qquad\qquad (4\text{-}9)$$

Moreover, assuming that the N_i's are geometrically distributed Boes and Salas (1978) showed that:

$$cov \ (M_t^* , \ M_{t-k}^*) \ = \ \sigma_m^2 \ P(N_1 > k) \ = \ \sigma_m^2 \ (1 \ - \ p)^k \qquad\qquad (4\text{-}10)$$

where p is the geometric distribution parameter. Then Eq. (4-9) can be written as

$$\sigma_x^2 \ \rho_x(k) \ = \ \sigma_m^2 \ (1 \ - \ p)^k \qquad\qquad k > 0 \qquad\qquad (4\text{-}11)$$

where $\rho_x(k)$ = lag-k autocorrelation coefficient of the X process.

Considering time lags i and k such that i < k the following can be written from

57

Eq. (4-11):

$$\frac{\rho_x(k)\; \sigma_x^2}{\rho_x(i)\; \sigma_x^2} = \frac{\sigma_m^2\; (1 - p)^k}{\sigma_m^2\; (1 - p)^i} \tag{4-12}$$

and after simplifying it yields

$$\frac{\rho_x(k)}{\rho_x(i)} = (1 - p)^{k-i} \tag{4-13}$$

Then the parameter p can be expressed as a function of the lag-i and lag-k correlations as

$$p = 1 - \left[\frac{\rho_x(k)}{\rho_x(i)}\right]^{\frac{1}{k-i}} \quad \textit{for } i > 0,\; k > 0,\; \textit{and}\;\; i < k \tag{4-14}$$

Furthermore, from Eq. (4-11) one can get σ_m^2 as

$$\sigma_m^2 = \frac{\rho_x(i)\; \sigma_x^2}{(1 - p)^i} \quad \textit{for } i > 0 \tag{4-15}$$

Also, from Eq. (4-5), we can express σ_z^2 as follows:

$$\sigma_z^2 = \sigma_x^2 - \sigma_m^2 \tag{4-16}$$

4.1.3 Range Properties of the GNN Model

In discussing the range properties of the GNN model, Obeysekera (1980) showed that the

58

following asymptotic equations for the mean range R_n and the mean rescaled range R_n^{**} that were developed by Troutman (1976) for continuous models are applicable for the GNN model:

$$E\ [R_n]\ \approx\ 2\ (\frac{2}{\pi})^{\frac{1}{2}}\ \gamma\ n^{\frac{1}{2}} \qquad (4\text{-}17)$$

$$E[R_n^{**}]\ \approx\ (\frac{2}{\pi})^{\frac{1}{2}}\ \beta\ n^{\frac{1}{2}} \qquad (4\text{-}18)$$

where

$$\beta^2\ =\ 1\ +\ 2\ \sum_{k=1}^{\infty}\ \rho_x(k) \qquad (4\text{-}19)$$

$$\gamma^2\ =\ \sigma_x^2\ (\ 1\ +\ 2\ \sum_{k=1}^{\infty}\ \rho_x(k)\ ) \qquad (4\text{-}20)$$

From Eqs. (4-19) and (4-20)

$$\gamma^2\ =\ \sigma_x^2\ \beta^2 \qquad (4\text{-}21)$$

Obeysekera (1980) showed that for the GNN model, γ^2 of Eq. (4-20) can be expressed as:

$$\gamma^2\ =\ \sigma_x^2\ \left[\frac{2\ \rho_x(1)\ +\ p}{p} \right] \qquad (4\text{-}22)$$

59

or

$$\beta^2 = \frac{2 \ \rho_x(1) + p}{p} \qquad (4\text{-}23)$$

Then the parameter p of the GNN model can be expressed as:

$$p = \frac{2 \ \rho_x(1)}{\beta^2 - 1} \qquad (4\text{-}24)$$

4.1.4 Run Properties of the GNN Model

The expected value of the positive run length for the GNN model can be expressed as (Obeysekera, 1980):

$$E(T^+) = \frac{\displaystyle\int_{-\infty}^{\infty} \frac{G(Q-m)}{1 - \rho \ G(Q-m)} f_M(m) \ dm}{1 - (1-\rho) \displaystyle\int_{-\infty}^{\infty} \frac{G(Q-m)}{1 - \rho \ G(Q-m)} f_M(m) \ dm} \qquad (4\text{-}25)$$

where $E(T^+)$ = expected value of the positive run length at a truncation level Q, ρ = 1-p, and

$$G(Q-m) = \int_{Q-m}^{\infty} f_Z(z) \ dz \qquad (4\text{-}26)$$

$$f_M(x) = f_Z(x) = \frac{1}{\sqrt{2 \ \pi}} \ e^{-\frac{1}{2}x^2} \qquad (4\text{-}27)$$

60

4.1.5 Parameter Estimation of The GNN Model

The parameters to be estimated are: p (geometric distribution parameter), σ_m^2 (variance of the M process), and σ_z^2 (variance of the Z process).

Method of moments (MOM)

In this method, the population parameters are equated to their estimates from the observed sample. Therefore, from Eq (4-14), the parameter p can be estimated as

$$\hat{p} = 1 - \left[\frac{\hat{\rho}_x(k)}{\hat{\rho}_x(i)}\right]^{\frac{1}{k-i}} \quad \textit{for } i > 0, \ k > 0, \textit{ and } i < k \tag{4-28}$$

In this dissertation, the choice is made that i is always less than k. This choice is adapted throughout this dissertation unless otherwise indicated. From Eq. (4-15) σ_m^2 can be estimated as:

$$\hat{\sigma}_m^2 = \frac{\hat{\rho}_x(i)\ \hat{\sigma}_x^2}{(1 - \hat{p})^i} \quad \textit{for } i > 0 \tag{4-29}$$

Likewise, from Eq. (4-16) σ_z^2 can be estimated as:

$$\hat{\sigma}_z^2 = \hat{\sigma}_x^2 - \hat{\sigma}_m^2 \tag{4-30}$$

where $\hat{\sigma}_x^2$ and $\hat{\rho}_x(k)$ are estimated from the observed series as follows:

$$\hat{\sigma}_x^2 = \frac{1}{n} \sum_{t=1}^{n} (x_t - \bar{x})^2 \qquad\qquad (4\text{-}31)$$

$$\hat{\rho}_x(k) = \frac{1}{n} \frac{\sum_{t=1}^{n-k} (x_t - \bar{x})(x_{t-k} - \bar{x})}{\hat{\sigma}_x^2} \qquad\qquad (4\text{-}32)$$

in which $\bar{x}$ is the sample mean and n = sample size.

It is clear from Eqs. (4-28) through (4-30) that the estimated parameters depend very much on the choice of i and k. Different estimates can be obtained by choosing different i and k combinations. It is possible that the chosen i and k will produce unreasonable estimates of the parameters such as negative variance for the {M} or the {Z} process. To identify which values of i and k can be used to estimate the model parameters, the following (parameter space) conditions must be met by the GNN model:

(1) $0 < \hat{p} < 1$, (2) $\hat{\sigma}_m^2 > 0$, and (3) $\hat{\sigma}_z^2 > 0$.

Equation (4-28) suggests that for the condition (1) to be met, the following should be satisfied

$$0 < \left[\frac{\hat{\rho}_x(k)}{\hat{\rho}_x(i)} \right]^{\frac{1}{k-i}} < 1 \qquad\qquad \text{for i<k and k>0}$$

Since i is less than k, the following should be true

$$\hat{\rho}_x(k) < \hat{\rho}_x(i)$$

and $\hat{\rho}_x(i)$ *and* $\hat{\rho}_x(k)$ are both positive or negative.

Equation (4-29) suggests that condition (2) is always true if

$$\frac{\hat{\rho}_x(i)}{(1-\hat{p})^i} > 0$$

and since $0<p<1$, then the lag-i correlation coefficient $\hat{\rho}_x(i)$ used in the above equation must satisfy the following condition: $\hat{\rho}_x(i) > 0$. Furthermore, from condition (1) both lag-i and lag-k correlation coefficients must be both positive or negative, then the following must also be true

$$\hat{\rho}_x(k) > 0.$$

Equation (4-30) suggests that condition(3) is always true if $\hat{\sigma}_m^2 < \hat{\sigma}_x^2$. Then from Eq.(4-29), we can conclude that

$$\frac{\hat{\rho}_x(i)}{(1-\hat{p})^i} < 1$$

Furthermore from this condition to and Eq. (4-28), we can also conclude that

$$\hat{\rho}_x(i) < \left[\frac{\hat{\rho}_x(k)}{\hat{\rho}_x(i)}\right]^{\frac{i}{k-i}}$$

After some algebra it gives

63

$$[\hat{\rho}_x(i)]^{\frac{k}{i}} < \hat{\rho}_x(k)$$

In summary, the sample correlations that are used to estimate the model parameters using MOM must comply with the following:

(a) $\hat{\rho}_x(i) > 0$

(b) $\hat{\rho}_x(k) > 0$

(c) $\hat{\rho}_x(k) < \hat{\rho}_x(i)$

(d) $(\hat{\rho}_x(i))^{\frac{k}{i}} < \hat{\rho}_x(k)$

Parameter Estimation Using Regression Analysis

The least squares method is widely used in estimating the parameters of several stochastic models such as the ARMA(p,q) model. Since traditional least squares methods are not applicable in the case of the GNN model, the regression analysis was used instead.

Taking natural logarithm of both sides of Eq. (4-13), one can write:

$$\log\left[\frac{\rho_x(k)}{\rho_x(i)}\right] = (k-i)\,\log(1-p) \qquad \text{for } i < k \qquad (4\text{-}33)$$

which is a linear regression equation of the form:

$$Y = b\,X + a$$

where $Y = \log\left[\dfrac{\rho_x(k)}{\rho_x(i)}\right]$; $X = k - i$; $b = \log(1-p)$ and a=0.

The coefficient b can be estimated using the least squares method as

$$\hat{b} = \frac{\sum\limits_{i=1}^{L} X_i \, Y_i}{\sum\limits_{i=1}^{L} X_i^2}$$

where L is the number of lags used in the regression model. Then the parameter p can be estimated as

$$\hat{p} = 1 - \exp[\hat{b}] \qquad (4\text{-}34)$$

Finally the parameters $\hat{\sigma}_m^2$ *and* $\hat{\sigma}_z^2$ can be estimated using Eqs. (4-29) and (4-30).

The difference between using this method of estimation over the direct MOM is that in the direct MOM only two correlation coefficients are used to estimate the model parameters as seen from Eqs. (4-28) through (4-30). On the other hand, in the regression method more than two lags are used to estimate the parameter p. This suggest that more information is used in estimating the model parameters and thus this method may be more reliable than the direct MOM. Note also that the parameter space of this method is exactly the same as the direct MOM. In other words the conditions imposed on the sample correlation coefficients that are used to estimate the model parameters in the direct MOM are the same for this method. Therefore, no further restrictions are imposed by using this method over the direct MOM method.

Parameter Estimation by Fitting the Autocorrelation Function

Another alternative to using the estimated values of $\rho_x(k)$ k=1,.... from Eq. (4-32) for the MOM estimation of the model parameters is to fit an autocorrelation function for the observed series and use the fitted function to estimate the model parameters using the method of MOM. The autocorrelation function can be fitted using the following equation:

$$\rho_x(k) = A \ e^{-Ck} \ , \ for \ k > 0 \tag{4-35}$$

Taking natural logarithm of both sides of Eq. (4-35)

$$\log_e \rho_x(k) = \log_e A - C \ k \tag{4-36}$$

Equation (4-36) is a linear regression equation of the form

$$Y = bX + a \tag{4-37}$$

where $Y = \log_e \rho_x(k)$, $X = k$, $b = -C$ and $a = \log_e A$. Coefficients a and b will be determined by the least square method. Only autocorrelation that are greater than Zero will be used because of the log in Eq. (4-36). The least square estimates of a and b can be found as:

$$\hat{a} = \overline{Y} - \hat{b} \ \overline{X}$$

$$\hat{b} = \frac{L \sum_{i=1}^{L} X_i Y_i - \sum_{i=1}^{L} X_i \sum_{i=1}^{L} Y_i}{L \sum_{i=1}^{L} X_i^2 - \sum_{i=1}^{L} Y_i^2}$$

where $\bar{X} = (1/L) \sum_{i=1}^{L} X_i$, $\bar{Y} = (1/L) \sum_{i=1}^{L} Y_i$, and L is the number of lags used in fitting the autocorrelation function. Then A and C can be estimated as:

$$\hat{A} = \exp [\hat{a}] \tag{4-38}$$

$$\hat{C} = - \hat{b} \tag{4-39}$$

Thus, the parameters p, σ_m^2, and σ_z^2 can be estimated using Eqs. (4-28), (4-29), and (4-30) respectively in which the correlations are taken from the fitted equation (4-35) instead of using the raw values from Eq.(4-32). Note that in this case the estimated parameters are independent from the choice of lags i and k to be used. In other words, the estimated value of p from Eq. (4-28) is the same when, for example, we use the fitted correlations for lags i=1 and k=2 and when we use the correlations for lags i=3 and k=4.

The above procedure was applied to the sites Malakal of the Nile River System and Lake St.Clair of the Great Lakes System. Figure 4-1 shows the autocorrelogram for Malakal annual flows for the period (1912-1989) derived from Eq.(4-32) and the fitted autocorrelograms where the number of lags (L) is taken as 10, 20, and 30 lags. Figure 4-2 is a similar plot for Lake St.Clair for the period (1900-1989). These figures show that for

Malakal, the fitted correlograms for L=10, 20, and 30 are about the same. So the GNN model parameters estimated by using this method will be about the same regardless of the choice of L. On the other hand, the fitted correlograms for lake St.Clair seem to vary with L and the GNN model parameters will depend on the choice of L.

Parameter Estimation Using Range Properties

Equation (4-24) can be used to estimate the parameter p as follows:

$$\hat{p} = \frac{2\ \hat{\rho}_x(1)}{\hat{\beta}^2 - 1} \tag{4-40}$$

where $\hat{\rho}_x(1)$ is estimated from Eq. (4-32) and $\hat{\beta}^2$ is estimated from Eq. (4-19) as follows

$$\hat{\beta}^2 = 1 + 2 \sum_{k=1}^{L} \hat{\rho}_x(k) \tag{4-41}$$

in which a cut-off maximum number of lags is used to approximate the infinite sum in Eq. (4-19). Then Eqs. (4-29) and (4-30) are used to estimate $\hat{\sigma}_m^2$ *and* $\hat{\sigma}_z^2$.

To find the parameter space for the GNN model using this method, the three conditions mentioned earlier for the MOM method must be satisfied. From Eq. (4-40), condition(1), i.e. $0 < \hat{p} < 1$, (refer to page 58) is satisfied if

$$0 < \frac{2\ \hat{\rho}_x(1)}{\hat{\beta}^2 - 1} < 1$$

68

The left hand side of this expression implies

$$\frac{2\,\hat{\rho}_x(1)}{\hat{\beta}^2-1} > 0$$

Two cases are recognized here: (a) if $\hat{\rho}_x(1) > 0$, then $\hat{\beta}^2 > 1$. Thus from Eq.(4-41) one can write $1 + 2\sum\limits_{k=1}^{L}\hat{\rho}_x(k) > 1$ which is equivalent to $\sum\limits_{k=1}^{L}\hat{\rho}_x(k) > 0$. (b) if $\hat{\rho}_x(1) < 0$, then $\hat{\beta}^2 < 1$,and $\sum\limits_{k=1}^{L}\hat{\rho}_x(k) < 0$

The right hand side of the foregoing expression implies

$$\frac{2\,\hat{\rho}_x(1)}{\hat{\beta}^2-1} < 1$$

Substituting $\hat{\beta}^2$ of Eq. (4-41) in the above expression, one can write

$$2\,\hat{\rho}_x(1) < 2\sum\limits_{k=1}^{L}\hat{\rho}_x(k) \quad \text{or} \quad \hat{\rho}_x(1) < \sum\limits_{k=1}^{L}\hat{\rho}_x(k). \text{ It follows } 0 < \sum\limits_{k=2}^{L}\hat{\rho}_x(k).$$

Summarizing, condition (1) requires the following :

$$\sum\limits_{k=2}^{L}\hat{\rho}_x(k) > 0 \text{ if } \hat{\rho}_x(1) > 0.$$

$$|\hat{\rho}_x(1)| > \sum\limits_{k=2}^{L}\hat{\rho}_x(k) \text{ if } \hat{\rho}_x(1) < 0.$$

From Eqs. (4-29) and (4-30), conditions (2) and (3) are satisfied if $\hat{\rho}_x(i) > 0$ and

69

$\hat{\rho}_x(i) < (1 - \hat{p})^i$ where $\hat{p}$ is estimated from Eq. (4-40). Therefore, in using the range method of estimation the following restrictions apply:

(a) When $\hat{\rho}_x(1) > 0$, then

   i. $0 < \sum\limits_{k=2}^{\infty} \hat{\rho}_x(k)$

   ii. $\hat{\rho}_x(i) > 0$

   iii. $\hat{\rho}_x(i) < (1 - \hat{p})^i$

(b) When $\hat{\rho}_x(1) < 0$

   i. $|\hat{\rho}_x(1)| > \sum\limits_{k=2}^{\infty} \hat{\rho}_x(k)$

   ii. $\hat{\rho}_x(i) > 0$

   iii. $\hat{\rho}_x(i) < (1 - \hat{p})^i$

Parameter Estimation Using Run Properties

In cases where the variances of the {M} and {Z} processes are both equal to one (i.e. $\hat{\sigma}_m^2 = \hat{\sigma}_z^2 = 1.0$) then Eq. (4-25) can be solved analytically to estimate p. It can be shown that the analytical solution can be expressed as:

$$\hat{E}\ (T^+) = \frac{-\hat{p} - \log_e|1-\hat{p}|}{\hat{p}^2 - (1-\hat{p})\ (-\hat{p} - \log_e|1-\hat{p}|\ )}$$

where $\hat{E}\ (T^+)$ = an estimate of the expected positive run length at truncation level Q as was defined in Eq. (4-25). However, in practice the parameters σ_m^2 and σ_z^2 could be different from one, so this method has limited application for estimating the model parameters.

Summary of Alternative Estimation Techniques

The parameters of the GNN model to be estimated are p, σ_m^2, and σ_z^2. Five parameter estimation methods have been introduced. They are:

1. Method of moments (MOM)

In this method, the parameter p is estimated by using Eq. (4-28). The variances σ_m^2, and σ_z^2 are estimated by using Eqs.(4-29) and (4-30), respectively.

2. Using regression analysis via MOM

In this method, the parameter p is estimated by using Eq. (4-34). σ_m^2, and σ_z^2 are estimated by using Eqs. (4-29) and (4-30), respectively.

3. Using fitted autocorrelation function via MOM

In this method, the parameter p is estimated by using Eq. (4-28) using the fitted correlations. σ_m^2, and σ_z^2 are estimated by using Eqs. (4-29) and (4-30) respectively.

4. Using range properties of the GNN model

In this method, the parameter p is estimated by using Eq. (4-40). σ_m^2, and σ_z^2 are estimated by using Eqs. (4-29) and (4-30), respectively.

5. using run properties of the GNN model

In this method the parameter p can be estimated by using the analytical solution of Eq. (4-25) which was discussed earlier. Note however that the analytical solution can be used only in cases where the parameters σ_m^2 and σ_z^2 are both equal to one.

4.1.6 Performance of the Estimation Methods

The performance of the estimation methods to estimate the model parameters was

71

tested by using the data generation method. Data was generated by using known (population) GNN model parameters. The generated samples were then used to estimate the model parameters using the different estimation methods mentioned in the previous section. Then estimated parameters from the generated samples will be compared to the population parameters.

The population parameters used in the generation experiment were $\sigma_m^2 = 1.0$, $\sigma_z^2 = 1.0$, and $p = 0.05, 0.15,$ and 0.25. Using these parameters, 100 samples of lengths 50, 100, 200, and 500 were generated. The generated data was then used to estimate the model parameters using each method of estimation. The results of the generation experiments are shown in Tables 4-1 through 4-15. These tables are self explanatory. Each table shows the population parameters, the generated parameters, bias, and the square root of the mean square error (RMSE) for each estimation method. These quantities are estimated as follows. Consider a model that has a number of (population) parameters and statistical properties. Let θ represent any such parameter or property. The model is used to generate n samples of size N. From each generated sample, θ is estimated, so that an array of estimates $\hat{\theta}_1, \hat{\theta}_2, ..., \hat{\theta}_n$ is obtained and the mean $m(\hat{\theta})$ and standard deviation $s(\hat{\theta})$ of $\hat{\theta}$ are determined from such an array. Then, the bias, $Bias(\hat{\theta})$ and the root mean square error, $RMSE(\hat{\theta})$, are determined by:

$$Bias(\hat{\theta}) = \theta - \bar{\theta}$$

$$RMSE(\hat{\theta}) = \sqrt{Bias^2 + S^2(\theta)}$$

In our data generation experiment we considered n=100 and various values of N as above noted. In some cases, some samples resulted in estimating unreasonable parameters such as p>1. In such cases, more samples were generated so that we will have 100 samples for which the estimated parameters for each sample are reasonable. The number of generated samples that was needed to achieve the 100 samples is shown in column 5 in the tables.

In general, results show that the MOM performance depends on the choice of the lags i and k which are used in Eqs. (4-28) and (4-29) to estimate the model parameters. The MOM, method generally, did not perform well in reproducing the population parameters of the GNN model. This method resulted, in general, in high bias in estimating the model parameters. The biases for p and σ_z^2 decreased as N increases for all cases. For σ_m^2 the biases, in general, decreased as N increases for p = 0.15 and p = 0.25. However, there is no consistent behavior for the biases for σ_m^2 when p =0.05. The regression analysis method performed better than MOM in reproducing the population parameters in terms of bias and RMSE as shown in Tables 4-1 through 4-15. In general, this method produced the smallest bias in estimating the parameters p and σ_z^2 among all other methods except for p=0.25. The performance of the regression analysis method depends on the lag i used in Eq. (4-29) to estimate σ_m^2. The best performance in terms of bias and RMSE is when i =1. The results, however deteriorate in terms of bias and RMSE when i >1. The method based on the fitted autocorrelation function performed well in reproducing the parameters p and σ_z^2. This method however, resulted in high bias and RMSE in estimating σ_m^2 as shown in Tables 4-3, 4-8, and 4-13. The performance of this method depend on the choice of the number of lags (L) which is used in fitting the autocorrelation function. Using the range properties to

estimate the model parameters resulted in relatively high bias in estimating the parameter p (for p=0.05, and 0.15) when compared with the regression method as shown in Tables 4-4 and 4-9. However, the range method produced less bias than the regression method for p = 0.25 as shown in Table 4-14. The performance of this method depends on the choice of i in Eq. (4-29). The best results were obtained when i =1 and the results deteriorated when i >1. Using the run properties to estimate the model performance did not perform well in reproducing the model parameters. This method resulted in high bias and RMSE in estimating the model parameters.

In summary the regression analysis method performed better than the other methods in reproducing the population parameters for p=0.05 and p=0.15. The range method performed relatively better than the other methods for p = 0.25. In general, these two methods performed better than the MOM and fitting autocorrelation methods.

4.1.7 Application of the GNN Model to Observed Hydrologic Data

In this section, data generation will be used to test whether the GNN model is capable of preserving some basic statistics of the observed historical data. Two sets of observed data were used: the annual flows of the white Nile River at Malakal and the annual net basin supplies (NBS) at Lake St.Clair of the Great Lakes System. Figure 4-3 shows the observed annual flows for Malakal for the period 1912 - 1989 and Fig. 4-4 shows the historical NBS for Lake St.Clair for the period 1900 - 1989. The GNN model was fitted to the historical data for these two sites. The parameters were estimated by using the methods discussed in

74

section 4.1.5. Once the parameters were estimated, 100 samples of the same length as the historical record were generated. The generated sequences were analyzed statistically. The statistics that were used in the analysis are:

1. mean and standard deviation

2. Skewness coefficient

3. Lag-1 autocorrelation coefficient

4. The partial sum of the autocorrelation function (ACF) defined as $\sum_{i=1}^{L} \rho_x(i)$ where L is the number of lags used in the sum and is taken as 10 and 20.

5. The longest drought at truncation level equal to the mean annual flow.

6. The rescaled adjusted range.

7. Hurst coefficient.

8. Surplus.

9. Deficit.

The storage-related statistics are particularly important in modeling time series for simulation studies of reservoir systems. Since such characteristics are functions of the dependence structure of a series, they are also useful in identifying the degree of temporal dependence of a series. Consider the time series y_i, $i = 1, ..., N$ and a subsample $y_1, ..., y_n$ with $n \leq N$. Form the sequence of partial sums S_i as

$$S_i = S_{i-1} + (y_i - \bar{y}_n) \qquad i = 1, ..., n$$

where $S_0 = 0$ and $\bar{y}_n$ is the sample mean determined by

75

$$\bar{y}_n = \left(\frac{1}{n}\right) \sum_{i=1}^{n} y_i$$

Define the sample standard deviation by

$$s_n = \left[\frac{1}{n} \sum_{i=1}^{n} (y_i - \bar{y})^2 \right]^{1/2}$$

Then the adjusted range and the rescaled adjusted range are calculated by

$$R_n^* = \max(S_0, S_1, ..., S_n) - \min(S_0, S_1, ..., S_n)$$

$$R_n^{**} = \frac{R_n^*}{s_n}$$

respectively. The Hurst coefficient for a series is estimated by

$$K = \frac{\ln(R_n^{**})}{\ln(n/2)}, \qquad n > 2$$

Another important storage related static is the surplus. Consider the time series y_i, $i = 1, ..., N$, and the demand level d is equal to the sample mean $\bar{y}$. A surplus occurs when $y_i > d$ consecutively during one or more years until $y_i \leq d$ again. Assuming that u surpluses occur in a given hydrologic sample, the maximum surplus or simply surplus is given by

$$U^* = \max(U_1, ..., U_u)$$

Likewise, drought-related statistics are also important in modeling hydrologic time series. A deficit occurs when $y_i < d$ consecutively during one or more years until $y_i \geq d$ again. Such a deficit can be defined by its duration L, its magnitude M, and its intensity $I = M/L$. Assuming that m deficits occur in a given hydrologic sample, the longest deficit duration (longest drought or maximum run-length) is given by

$$L^* = \max(L_1, \ldots, L_m)$$

and the maximum deficit magnitude (maximum run-sum) is

$$M^* = \max(M_1, \ldots, M_m)$$

The above statistics were obtained from the generated samples. Then average statistics were computed based on all generated samples. Such average statistics were compared with the corresponding historical statistics.

White Nile River at Malakal

The GNN model was fitted to the annual flows of the white Nile river at Malakal for the period 1912 - 1989. The MOM, regression analysis, fitting the autocorrelation function, and using the range properties methods were used to estimate the model parameters. The estimated parameters are shown in Table 4-16. For each estimation method the estimated parameters were used to generate 100 samples of annual flows of equal length to the

historical flows. The average statistics of the generated data were compared to the observed statistics of the historical sample. Tables 4-17 through 4-19 show the results of the generation experiment. The results are presented in three columns for each tested statistic (i.e. mean, standard deviation, longest drought ....etc), as shown in these tables. The second column represents the mean value of each statistic calculated from the generated series. The first and the third columns are simply the generated value of column 2 plus or minus one standard deviation of the generated value.

Table 4-17 shows that the GNN model did preserve the mean very well for all the methods of estimation tested. The standard deviation was also well preserved by all methods except perhaps for method 2 (using regression analysis) where there is some bias. The skewness however, was not preserved because the original skewed data was not transformed. The lag-1 autocorrelation coefficient was well preserved using all estimation methods except for the method based on fitting of the autocorrelation function, (labeled method 3 in the tables). The preservation of the partial sum of the ACF is important in water resources since it is directly related to the range as shown in section 4.1.3. The range is an important statistic in water resources as it is related to reservoir sizing. Hence, it is expected that the ability of any model to preserve the range is directly related to the ability of the model to preserve the partial sum of the ACF. In general, the partial sums of the ACF for L=10 is reasonably well preserved by all estimation methods except for method 3 (based on fitting the ACF) and method 1 (using MOM with i=2, k=6). However, for L=20 the results are not as good and only method 2 (using regression analyssis) was relatively successful. The rescaled range was well preserved using all methods of estimation except for method 3. Similarly the Hurst

coefficient was well preserved using all methods of estimation except for method 3. The regression method performed the best (in terms of bias) in preserving the rescaled range and the Hurst coefficient at Malakal. The longest drought was also well preserved using all methods of estimation except for method 3 which underestimated the longest drought. In general, the regression method performed the best in preserving the historical statistics at Malakal.

Similar generation experiments were made using ARMA(p,q) models of different orders. The parameters of the ARMA models were estimated using the methods of moments (MOM) and least squares (LS). The estimated parameters are shown in Table 4-20 and the results are shown in Tables 4-21 to 4-23. The ARMA models performed quite well in preserving the mean and the standard deviation. The skewness was not preserved because the analysis was made based on the original data which is skewed. The lag-1 autocorrelation coefficient was well preserved using all ARMA models. ARMA(1,1) model resulted in the least bias among all other ARMA models. However, AR(1) performed better than the other ARMA models in preserving the partial sums of the ACF. AR(1) model also performed better than the other ARMA models in preserving the longest drought, rescaled range, and the Hurst coefficient. In comparing the performances of the GNN and ARMA models in preserving the lag-1 autocorrelation coefficient, the ARMA models resulted in less bias than the GNN model. However, for the partial sums of ACF, the GNN model based on the regression analysis estimation method performed better (in terms of bias) than ARMA models as shown in Tables 4-18 and 4-22. Similarly the rescaled range and the Hurst coefficient were best preserved using the GNN model based on the regression analysis estimation method. In

79

preserving the longest drought ARMA models had, in general, a better performance than the GNN model.

Figures 4-5 through 4-11 compare the historical and generated autocorrelogram for the GNN model using the different methods which were used to estimate the model parameters. Likewise, Figs. 4-12 through 4-18 show the generated autocorrelograms for the ARMA models. From these figures, one can conclude that all the models studied did perform well in preserving the overall shape of the historical autocorrelogram except for the GNN model based on fitting the ACF estimation method. The decay of the autocorrelogram is slower for the AR(1) than other ARMA models. For the GNN model, the decay of the autocorrelogram was slowest for the case based on the regression analysis estimation method as shown in Fig. 4-9.

The correlograms plotted in Figures 4-5 through 4-11 for the GNN model 4-12 through 4-18 for ARMA models were obtained based on the 100 samples of generated data. In other words, the plotted correlogram is the average of the 100 correlograms obtained from the generated samples. Figures 4-5, 4-9, and 4-11 for the GNN model and 4-12, 4-14, and 4-15 for ARMA models also show another correlogram which was obtained based on one (generated) sample. This sample was formed by putting the 100 generated samples mentioned above next to each other thus forming one long sample of 100 x 78 = 7800 data points. Then a correlogram was obtained based on such long series. By examining these figures one can see that the correlation estimated based on the 100 samples were biased (always underestimated). One can explain this bias by the fact that if one uses the long time series (7800 data points) to calculate the correlations, the overall mean based on the 7800 data

points will be used to calculate the correlations. On the other hand, if one uses 100 samples each consisting of 78 data points and calculates the correlation for each sample and then computes the average of all these correlations, these correlations will always be underestimated because the mean value for each sample (based on the 78 data points) will be different. This is always true regardless if the sample mean (based on 78 data points) is higher or lower than the overall mean (based on 7800 data points). This is one of the disadvantages associated with testing the ACF in this manner which is probably the standard method among hydrologists in conducting simulation experiments. However, for evaluating the preservation of other properties such as drought and storage related statistics, the only feasible choice is to generate samples of length equal or less than the length of the historic sample. This is because storage and drought related statistics are quantities that vary with the sample length.

At this point, it would be interesting to look at the generated flows from the GNN model and compare them with the historical flows to see if the GNN model was able to generate flows that, in general, resemble the historical flows. Figure 4-19 shows the historical flows at Malakal along with 7 generated samples of equal length using the GNN model where the parameters were estimated using the regression analysis method. These figures suggest that in at least 6 of the 7 samples the GNN model was capable, in general, of producing the basic characteristics of the historical flows (i.e. some kind of shift in the mean). Likewise Figs. 4-20 and 4-21 show 7 generated samples from the AR(1) and ARMA(1,1) models, respectively. Out of the 7 samples, only two or three samples show some kind of shift in the mean for the AR(1) and ARMA(1,1) models. Similar results were obtained for

81

higher order ARMA models (not shown here). The point here is that the GNN model appears to be more consistent than ARMA in picking up the shifts observed in the historical data.

As was discussed earlier, the skewness coefficient was not preserved using the GNN or the ARMA models. In an attempt to preserve the skewness coefficient, the historical data was transformed using a logarithmic transformation such as

$$X_t = \log_e [Y_t + C]$$

where X_t is the transformed series, Y_t is the observed series, and C is the transformation coefficient. The transformation coefficient for Malakal was estimated to be -21. The transformed data was then used to estimate the GNN model parameters using the four estimation methods discussed above. The estimated parameters are shown in Table 4-16. For each estimation method, the estimated parameters were used to generate 100 samples of equal length to the historical sample. These generated values were then transformed into skewed data by inverting the above equation as

$$Y_t = \exp [X_t] - C$$

The results of these generation experiments are presented in Tables 4-24 through 4-26 for the GNN model and 4-27 through 4-29 for the ARMA models. Table 4-24 shows that the mean, standard deviation and skewness coefficient are all well preserved except may be for the regression analysis method which resulted in the highest bias with respect to the preservation

82

of the standard deviation. MOM with i=1, k=8 resulted in the highest bias with respect to the preservation of the skewness coefficient. For lag-1 autocorrelation coefficient the fitting of the ACF resulted in the highest bias whereas MOM with i=1, k=8 produced the smallest bias. Notice that for lag-1 autocorrelation coefficient the biases based on the transformation case are higher than those obtained based on the original untransformed data except for method 3. The regression analysis and MOM methods had the best performance in preserving the partial sum of the ACF as shown in Table 4-25. Also, for the regression method, the bias based on the transformation case is lower than the bias based on the original untransformed data for L=10. The opposite is true for L=20. The regression analysis and MOM methods performed better than the other methods in preserving the rescaled range and the Hurst coefficient. However, the regression method resulted in the highest bias in preserving the longest drought. For the ARMA models, Table 4-27 shows that the mean, standard deviation, and skewness coefficient are well preserved. For lag-1 autocorrelation coefficient and the partial sum of ACF the biases based on the transformation case are higher than those obtained based on the original untransformed data for all estimation methods. For the longest drought, the ARMA models based on the transformation case overestimated this statistic for all estimation methods whereas they underestimated the longest drought based on the original untransformed case for all estimation methods. For the rescaled range and the Hurst coefficient, the biases based on the transformation case are higher than those obtained based on the original untransformed data for all estimation methods.

Monte Carlo experiments were also used to test the performance of the GNN and ARMA models to preserve the rescaled range, longest drought, maximum deficit, and surplus

for different data lengths. For this purpose, the GNN model fitted to the historical flows for Malakal was used to generate 100 samples of synthetic data of different lengths. These lengths are 30, 38, 46, 54, 62, 70, and 78 years. The average rescaled range, the longest drought, the maximum deficit, and the surplus based on the generated samples and the historical ones are shown in Figs. 4-22, 4-26, 4-30, and 4-34 respectively. The historical statistics are calculated based on averaging over a given data length. For example the historical rescaled range for a data length of say 30 years is calculated by finding the sample rescaled range based on the record for years 1 through 30 and then from year 2 through 31 and so on. The average of all these rescaled ranges is calculated and that is what is shown in the figures as the historical rescaled range. These figures show that the GNN model did perform well in preserving these important statistics for different sample lengths. Similar experiments with ARMA(p,q) were also conducted. The results are shown in Figs. 4-23 through 4-25 for the rescaled range, Figs. 4-27 through 4-29 for the longest drought, Figs. 4-31 through 4-33 for the maximum deficit, and Figs. 4-35 through 4-37 for the surplus . From these figures, it is shown that, in general, the GNN model performed better than the ARMA(p,q) models in preserving these important statistics for Malakal. However, the ARMA(1,1) has performed slightly better than GNN in preserving the maximum deficit.

Lake St. Clair

The GNN model was fitted to the historical Net Basin Supply (NBS) data for the period (1900-1989). As in the first example, 100 samples were generated of equal length to the historical sample using the estimated parameters for each method. The results are shown

in Tables 4-30 through 4-32. Table 4-30 shows that the mean and the standard deviation are well preserved for all cases. The MOM, regression analysis, and range methods performed better than the fitting autocorrelation method in preserving the rescaled range and the Hurst coefficient as shown in Tables 4-30 through 4-32. The MOM method performance changes with the choice of i and k. The longest drought was overestimated except for the fitting of the ACF estimation method. The rescaled range was underestimated for all estimation methods. Overall, the GNN model did perform well in preserving the basic short and long term statistics.

Several ARMA models were also fitted to the historical data at lake St.Clair. The method of moments (MOM) was used to estimate the model parameters. The least squares method was used to estimate the model parameters in cases when the MOM estimated parameters did not satisfy the stationarity condition for the fitted model. As in the previous case, 100 samples were generated using the estimated parameters. Each sample was of equal length to the historical series. The results of the simulation experiments for the ARMA models are shown in Tables 4-33 through 4-35. The ARMA models performed slightly better than the GNN model in preserving the longest drought. The GNN, however performance was outstanding when compared with the ARMA models in preserving the rescaled range. This is one of the advantages of using the GNN model when dealing with data that exhibit apparent shifts such as the NBS record at lake St.Clair. The range is a very important statistic especially in reservoir operations and design. The simulation model used should be able to preserve the range in such cases. As a result the GNN model can be used in cases where the preservation of the range is very important.

Figures 4-38 through 4-46 show the generated autocorrelograms for the GNN model using the different methods of estimation. Figures 4-47 through 4-53 show the generated autocorrelograms for the ARMA models. The regression analysis method results show that the historical autocorrelogram is well preserved by using this method especially for the high lags as shown in Fig 4-46. This is probably the reason that the rescaled range was very well preserved using this method. Figure 4-54 show 7 generated samples for lake St.Clair from the fitted GNN model where the parameters were estimated using the regression analysis method. Figures 4-55 through 4-57 show generated samples from the ARMA(1,1), ARMA(2,2), and ARMA(3,1) models respectively for lake St.Clair. These figures show that the GNN model was capable of reproducing the basic characteristics of the historical NBS especially the apparent shift which is observed in the historical NBS. ARMA(2,2) and ARMA(3,1) were also capable of mimicking the shifting phenomena observed in the historical NBS. Note however that the number of parameters that are needed for the GNN model (three parameters) is less than the number of parameters for the ARMA(2,2) or the ARMA(3,1) (four parameters).

As was done for Malakal, the historical NBS record was transformed using the log transformation presented in the previous section. The coefficient C was estimated as 106. Simulation experiments were done using the transformed data to check weather the skewness can be preserved and to check the effect of the transformation on the performance of the GNN model to preserve the historical statistics. The results are presented in Tables 4-36 through 4-38. Results show that the log transformation of the historical data performed well in preserving the skewness. As for the Malakal case, the transformation resulted in slightly

86

higher biases in preserving the other historical statistics especially for the lag-1 autocorrelation coefficient and the longest drought.

Monte Carlo experiments were also used to test the performance of the GNN and ARMA models to preserve the rescaled range, longest drought, maximum deficit, and surplus for different data lengths for lake St.Clair. The results are shown in Figs. 4-58 through 4-73. These figures show that for St.Clair the GNN clearly performed better than the ARMA models in preserving the rescaled range and the surplus. However, ARMA(1,1) and ARMA(2,1) models has, in general, performed slightly better than GNN in preserving the longest drought and the maximum deficit. The conclusion here is that both the GNN, ARMA(1,1) and ARMA(2,1) all performed well in mimicking its historical record. The GNN has the edge over ARMA(1,1) and ARMA(2,1) models in preserving the rescaled range and the surplus. The latter models have a somewhat slight edge in preserving the drought duration and the maximum deficit.

4.2 The GNN-1 model

The term representing the noise in the GNN model was assumed to be an uncorrelated process (white noise). One way to introduce correlation is by adding a moving average term. The resulting model is called GNN-1. Adding this extra parameter might help in better modeling the observed data in terms of preserving its statistical characteristics. Also this model can be used as an alternative to GNN in cases where GNN could not be fit to the data so model.

87

4.2.1 Model formulation

The GNN-1 model is defined as :

$$X_t = M_t^* + Z_t - \theta \, Z_{t-1} \tag{4-42}$$

Note that the difference between this model and the GNN model is the addition of a moving average term to the GNN model. The assumption about the M_t^* and the Z_t processes are as defined in section (4.1.1) for the GNN model. Therefore, following the same procedure as was done for the GNN model, one can write:

$$E\,(\,M_t^*\,M_{t-k}^*\,)\,=\,\sigma_m^2\,(\,1-p\,) \qquad \textit{for } k> 0$$

and

$$E\,(\,M_t^*\,Z_{t-k}\,)\,=\,E\,(\,M_{t-k}^*\,Z_t)\,=\,0 \qquad \textit{for } k \geq 0$$

Following the same procedure as for the GNN model , we can find the moment equations of the GNN-1 model. Multiplying both sides of Eq. (4-42) by X_t and taking expectation:

$$\begin{aligned}
E\,(\,X_t\,X_t\,) &= E\,(\,M_t^*\,X_t\,) + E\,(\,Z_t\,X_t\,) - \theta\,E\,(\,Z_{t-1}\,X_t\,) \\
&= E\,(\,M_t^*\,M_t^*\,) + E\,(\,Z_t\,Z_t\,) + \theta^2\,E\,(\,Z_{t-1}\,Z_{t-1}\,) \\
&= \sigma_m^2 + \sigma_z^2 + \theta^2\,\sigma_z^2
\end{aligned}$$

And finally, the above equation can be written as:

$$\sigma_x^2 = \sigma_m^2 + (1 + \theta^2) \sigma_z^2 \tag{4-43}$$

Multiplying both sides of Eq. (4-42) by X_{t-1} and taking expectation gives

$$\sigma_x^2 \rho_x(1) = \sigma_m^2 (1-p) - \theta \sigma_z^2 \tag{4-44}$$

In general, multiplying both sides of Eq. (4-42) by X_{t-k} and taking expectation, one can show that:

$$\sigma_x^2 \rho_x(k) = \sigma_m^2 (1-p)^k \qquad for \ k > 1$$

4.2.2 Range Properties of The GNN-1 Model

The range properties developed by Troutman (1976) which were mentioned in section 4.1.3 apply for the GNN-1 model. It can be shown (see proof in Appendix A) that the following equation applies:

$$p = \frac{2 \rho_x(2)}{\beta^2 - 1 - 2 \rho_x(1)} \tag{4-45}$$

where $\hat{\beta}^2$ is as defined as in Eq. (4-22).

4.2.3 Parameter Estimation of the GNN-1 Model

The parameters to be estimated are p (geometric distribution parameter), σ_m^2 (variance of the M process), σ_z^2 (variance of the Z process), and θ (moving average coefficient).

Method of Moments

Following the same procedure as for the GNN model, the MOM parameters of the GNN-1 model can be estimated as:

$$\hat{p} = 1 - \left[\frac{\hat{\rho}_x(k)}{\hat{\rho}_x(i)}\right]^{\frac{1}{k-i}} \quad for \ i<k \ , \ i > 1 \tag{4-46}$$

and

$$\hat{\sigma}_m^2 = \frac{\hat{\rho}_x(i) \ \hat{\sigma}_x^2}{(1-\hat{p})^i} \tag{4-47}$$

and we can use Eqs. (4-43) and (4-44) to estimate θ and σ_z^2 as follows:

$$\hat{\sigma}_z^2 = \frac{\hat{\sigma}_x^2 - \hat{\sigma}_m^2}{(1 + \hat{\theta}^2)} \tag{4-48}$$

$$\hat{\theta} = \frac{\hat{\sigma}_m^2 \ (1-\hat{p}) - \hat{\sigma}_x^2 \ \hat{\rho}_x(1)}{\hat{\sigma}_z^2} \tag{4-49}$$

Substituting Eq. (4-48) in (4-49), it can be shown that $\hat{\theta}$ can be found by solving the following quadratic equation:

$$A \ \hat{\theta}^2 + B \ \hat{\theta} + C = 0 \quad \text{where } A= \sigma_m^2 \ (1-p) - \sigma_x^2 \ \rho_x(1) \ , \ B= \sigma_m^2 - \sigma_x^2 \ , \text{and}$$

$$C = \sigma_m^2 \ (1-p) - \sigma_x^2 \ \rho_x(1) \ .$$

90

Using Regression Analysis

Following the same procedure as for the GNN model, the parameter p can be estimated using the following regression model :

$$
\log \left[\frac{\rho_x(k)}{\rho_x(i)} \right] = (k-i) \log(1 - p) \qquad \textit{for } i > 1, \ k > 1, \textit{ and } i < k
$$

$$(4\text{-}50)$$

The parameter p can be estimated from the above regression model using the least squares method as was done in the GNN model. Note that the difference between the above equation and Eq. (4-34) of the GNN model is that i and k in the above model can't be equal to 1. In other words the lag-1 autocorrelation coefficient will not be used in the estimation of the parameter p of the GNN-1 model.

Using Range Properties

From Eq. (4-45), we can write:

$$
\hat{p} = \frac{2 \, \hat{\rho}_x(2)}{\hat{\beta}^2 - 1 - 2 \, \hat{\rho}_x(1)}
$$

$$(4\text{-}51)$$

Once $\hat{p}$ is estimated, then $\hat{\sigma}_m^2$, $\hat{\theta}$, $\hat{\sigma}_z^2$ can be estimated using the MOM method mentioned in section 4.5.1 above.

4.2.4 Application of the GNN-1 Model to Observed Hydrologic Data

Lake St.Clair

The GNN-1 model was fitted to the historical Net Basin Supply (NBS) data for the period 1900 to 1989. 100 samples were generated of equal length to the historical sample using the estimated parameters. Tables 4-39 and 4-40 show the results of this experiment. Table 4-39 show that the mean and the standard deviation are well preserved for all cases. The storage and drought related statistics are also well preserved as shown in Table 4-40. The generated and the historical ACF functions for lake St.Clair for each estimation method are shown in Figs. 4-74 through 4-79. These figures suggest that this model, in general, performed well in preserving the ACF for Lake St.Clair. By comparing the results for the GNN and the GNN-1 models, one can argue that the performance of the GNN-1 model was similar to that of the GNN model. The MOM method performance changes with the choice of i and k. As in the previous example, the longest drought was overestimated while the rescaled range was underestimated. The regression analysis method performed well in preserving the mean and the standard deviation as well as the longest drought and the rescaled range. Overall, the GNN-1 did perform well in preserving the basic short and long term statistics. In fact the performance of the GNN-1 model is very close to the performance of the GNN model in preserving the short and long term properties of the NBS record for lake St.Clair. One may wonder, however, that since GNN-1 model has an additional parameter over the GNN model, one expects that the GNN-1 should perform better. One should know, however, that this case is similar to the case of the ARMA(1,1) and ARMA(2,1) models as an example. It is not unusual to find that in some cases the ARMA(1,1) performs better than

the ARMA(2,1) even though ARMA(2,1) has an additional parameter. It should be mentioned here that the GNN-1 model could not fit to the Malakal flows. The use of Eqs. (4-48) and (4-49) resulted in negative discriminant when solving the quadratic equations formulated form the above equations. As a result no generation was done for Malakal.

As for the GNN model, the Monte Carlo experiments were also used to test the performance of the GNN-1 model to preserve the rescaled range of observed time series for different data lengths. The average rescaled range and the longest drought based on the generated samples was compared to the historical ones are shown in Figs. 4-80 and 4-81. When comparing these figures with those for the GNN model, it is apparent that the GNN and GNN-1 models had similar performances in preserving the rescaled range and the longest drought.

Lake Ontario

The GNN and GNN-1 models were also fitted to the historical Net Basin Supply (NBS) data for lake Ontario for the period 1900 to 1989. The regression method was used to estimate the parameters of the fitted models. 100 samples of equal length to the historical sample using the estimated parameters were generated. Tables 4-41 and 4-42 show the results for the GNN and GNN-1 models respectively. The performance of the GNN model was slightly better than GNN-1 in preserving the rescaled range, surplus, and maximum deficit. The models have similar performance in preserving the longest drought. It is apparent that the GNN model has a slight edge over the GNN-1 model. The extra parameter that GNN-1 model has over the GNN model does not appear to improve the performance of

93

the GNN-1 model. However, the GNN-1 can be used as an alternative in cases where the

GNN model could not be fitted to the data.

Table 4-1   Results of the Generation Experiment for the GNN model with population parameters

$p = 0.05$, $\sigma_m^2 = 1.0$ and $\sigma_z^2 = 1.0$. The method of Moments (MOM) estimation are

used in which p is estimated using Eq. (4-28) where $i = 1$, $k = 2$, and σ_m^2 and σ_z^2 are

estimated using Eqs. (4-29) and (4-30) .

Parameter	Sample Length	Population θ	Generated $m(\hat{\theta})$	Bias	RMSE	# Samples
p	50	0.050	0.282	-0.232	0.287	247
	100	0.050	0.250	-0.200	0.273	172
	200	0.050	0.165	-0.115	0.193	143
	500	0.050	0.097	-0.047	0.087	133
σ_m^2	50	1.000	1.053	-0.053	0.707	247
	100	1.000	0.917	0.083	0.461	172
	200	1.000	1.015	-0.015	0.503	143
	500	1.000	0.982	0.018	0.368	133
σ_z^2	50	1.000	0.726	0.274	0.440	247
	100	1.000	0.828	0.172	0.339	172
	200	1.000	0.878	0.122	0.243	143
	500	1.000	0.948	0.052	0.123	133

95

Table 4-2   Results of the Generation Experiment for the GNN model with population parameters p=0.05, $\sigma_m^2 = 1.0$ and $\sigma_z^2 = 1.0$. The method of regression analysis estimation is used in which p is estimated using Eq. (4-34) where the number of lags (L) = 20, and σ_m^2 and σ_z^2 are estimated using Eqs. (4-29) and (4-30) where i = 1.

Parameter	Sample Length	Population θ	Generated $m(\hat{\theta})$	Bias	RMSE	# Samples
p	50	0.050	0.116	-0.066	0.151	135
	100	0.050	0.119	-0.069	0.105	113
	200	0.050	0.096	-0.046	0.082	100
	500	0.050	0.071	-0.021	0.042	100
σ_m^2	50	1.000	0.733	0.267	0.793	135
	100	1.000	0.718	0.222	0.686	113
	200	1.000	0.851	0.149	0.503	100
	500	1.000	0.915	0.085	0.383	100
σ_z^2	50	1.000	0.912	0.089	0.322	135
	100	1.000	0.952	0.048	0.220	113
	200	1.000	0.986	0.014	0.146	100
	500	1.000	0.983	0.017	0.093	100

Table 4-3   Results of the Generation Experiment for the GNN model with population parameters $p=0.05$, $\sigma_m^2 = 1.0$ and $\sigma_z^2 = 1.0$. The method of fitting the autocorrelation function is used in which the number of lags used (L)=20, p is estimated using Eq. (4-28) where $i=1$, $k=2$, and σ_m^2 and σ_z^2 are estimated using Eqs. (4-29) and (4-30).

Parameter	Sample Length	Population θ	Generated $m(\hat{\theta})$	Bias	RMSE	# Samples
p	50	0.050	0.093	-0.043	0.080	130
	100	0.050	0.106	-0.056	0.090	116
	200	0.050	0.088	-0.038	0.069	104
	500	0.050	0.071	-0.021	0.042	100
σ_m^2	50	1.000	0.504	0.496	0.776	130
	100	1.000	0.715	0.285	0.653	116
	200	1.000	0.899	0.101	0.549	104
	500	1.000	0.957	0.043	0.416	100
σ_z^2	50	1.000	0.971	0.029	0.334	130
	100	1.000	0.921	0.079	0.307	116
	200	1.000	0.937	0.063	0.265	104
	500	1.000	0.941	0.059	0.184	100

Table 4-4   Results of the Generation Experiment for the GNN model with population parameters $p = 0.05$, $\sigma_m^2 = 1.0$ and $\sigma_z^2 = 1.0$. The method of using the range properties estimation is used in which p is estimated using Eq. (4-40) where the number of lags (L) = 20, and σ_m^2 and σ_z^2 are estimated using Eqs. (4-29) and (4-30) where i = 1.

Parameter	Sample Length	Population θ	Generated $m(\hat{\theta})$	Bias	RMSE	# Samples
p	50	0.050	0.157	-0.107	0.139	126
	100	0.050	0.136	-0.086	0.112	108
	200	0.050	0.107	-0.057	0.072	100
	500	0.050	0.090	-0.040	0.045	100
σ_m^2	50	1.000	0.719	0.281	0.782	126
	100	1.000	0.757	0.243	0.670	108
	200	1.000	0.858	0.142	0.510	100
	500	1.000	0.934	0.066	0.390	100
σ_z^2	50	1.000	0.920	0.080	0.303	126
	100	1.000	0.958	0.042	0.205	108
	200	1.000	0.978	0.022	0.139	100
	500	1.000	0.964	0.036	0.093	100

Table 4-5 Results of the Generation Experiment for the GNN model with population parameters $p=0.05$, $\sigma_m^2 = 1.0$ and $\sigma_z^2 = 1.0$. The method of using the run properties estimation is used in which p is estimated using Eq. (4-25), and σ_m^2 and σ_z^2 are estimated using Eqs. (4-29) and (4-30) where $i=1$.

Parameter	Sample Length	Population θ	Generated $m(\hat{\theta})$	Bias	RMSE	# Samples
p	50	0.050	0.068	-0.018	0.061	126
	100	0.050	0.056	-0.056	0.040	108
	200	0.050	0.043	0.007	0.026	100
	500	0.050	0.037	0.013	0.114	100
σ_m^2	50	1.000	0.602	0.398	0.710	126
	100	1.000	0.681	0.319	0.641	108
	200	1.000	0.797	0.203	0.501	100
	500	1.000	0.881	0.119	0.377	100
σ_z^2	50	1.000	1.037	-0.037	0.254	126
	100	1.000	1.034	-0.034	0.190	108
	200	1.000	1.040	-0.040	0.130	100
	500	1.000	1.017	-0.017	0.081	100

Table 4-6   Results of the Generation Experiment for the GNN model with population parameters $p=0.15$, $\sigma_m^2 = 1.0$ and $\sigma_z^2 = 1.0$. The method of Moments (MOM) estimation are used in which p is estimated using Eq. (4-28) where $i=1$, $k=2$, and σ_m^2 and σ_z^2 are estimated using Eqs. (4-29) and (4-30) .

Parameter	Sample Length	Population θ	Generated $m(\hat{\theta})$	Bias	RMSE	# Samples
p	50	0.150	0.315	-0.165	0.240	180
	100	0.150	0.284	-0.134	0.206	138
	200	0.150	0.219	-0.069	0.149	115
	500	0.150	0.176	-0.026	0.088	108
σ_m^2	50	1.000	1.192	-0.192	0.685	180
	100	1.000	1.078	-0.078	0.492	138
	200	1.000	1.023	-0.023	0.329	115
	500	1.000	1.018	-0.018	0.283	108
σ_z^2	50	1.000	0.763	0.237	0.467	180
	100	1.000	0.827	0.173	0.358	138
	200	1.000	0.925	0.075	0.260	115
	500	1.000	0.960	0.040	0.142	108

Table 4-7 Results of the Generation Experiment for the GNN model with population parameters

p=0.15, $\sigma_m^2 = 1.0$ and $\sigma_z^2 = 1.0$. The method of regression analysis estimation

is used in which p is estimated using Eq. (4-34) where the number of lags (L) = 20,and

σ_m^2 and σ_z^2 are estimated using Eqs. (4-29) and (4-30) where i=1.

Parameter	Sample Length	Population θ	Generated $m(\hat{\theta})$	Bias	RMSE	# Samples
p	50	0.150	0.131	0.019	0.092	115
	100	0.150	0.141	0.009	0.086	107
	200	0.150	0.161	-0.011	0.076	100
	500	0.150	0.159	-0.090	0.077	100
σ_m^2	50	1.000	0.832	0.168	0.754	115
	100	1.000	0.909	0.091	0.534	107
	200	1.000	0.925	0.075	0.390	100
	500	1.000	0.992	0.008	0.302	100
σ_z^2	50	1.000	0.992	0.008	0.305	115
	100	1.000	1.015	-0.015	0.265	107
	200	1.000	1.009	-0.009	0.188	100
	500	1.000	0.971	0.029	0.154	100

Table 4-8   Results of the Generation Experiment for the GNN model with population parameters p=0.15, $\sigma_m^2 = 1.0$ and $\sigma_z^2 = 1.0$. The method of fitting the autocorrelation function is used in which the number of lags used (L)=20, p is estimated using Eq. (4-28) where i=1, k=2, and σ_m^2 and σ_z^2 are estimated using Eqs. (4-29) and (4-30).

Parameter	Sample Length	Population θ	Generated $m(\hat{\theta})$	Bias	RMSE	# Samples
p	50	0.150	0.117	0.034	0.081	119
	100	0.150	0.133	0.017	0.083	111
	200	0.150	0.158	-0.008	0.076	103
	500	0.150	0.153	-0.003	0.066	104
σ_m^2	50	1.000	0.629	0.371	0.801	119
	100	1.000	0.759	0.241	0.657	111
	200	1.000	0.866	0.134	0.536	103
	500	1.000	0.957	0.043	0.448	104
σ_z^2	50	1.000	1.137	-0.137	0.398	119
	100	1.000	1.111	-0.111	0.416	111
	200	1.000	1.051	-0.051	0.395	103
	500	1.000	1.005	-0.005	0.355	104

Table 4-9   Results of the Generation Experiment for the GNN model with population parameters $p=0.15$, $\sigma_m^2 = 1.0$ and $\sigma_z^2 = 1.0$. The method of using the range properties estimation is used in which p is estimated using Eq. (4-40) where the number of lags (L) = 20, and σ_m^2 and σ_z^2 are estimated using Eqs. (4-29) and (4-30) where $i=1$.

Parameter	Sample Length	Population θ	Generated $m(\hat{\theta})$	Bias	RMSE	# Samples
p	50	0.150	0.217	-0.067	0.117	109
	100	0.150	0.196	-0.046	0.098	107
	200	0.150	0.195	-0.045	0.090	100
	500	0.150	0.173	-0.023	0.059	100
σ_m^2	50	1.000	0.845	0.151	0.766	109
	100	1.000	0.920	0.080	0.523	107
	200	1.000	0.955	0.045	0.376	100
	500	1.000	0.999	0.001	0.280	100
σ_z^2	50	1.000	0.924	0.076	0.322	109
	100	1.000	0.975	0.025	0.273	107
	200	1.000	0.979	0.021	0.190	100
	500	1.000	0.964	0.036	0.119	100

Table 4-10   Results of the Generation Experiment for the GNN model with population parameters $p=0.15$, $\sigma_m^2 = 1.0$ and $\sigma_z^2 = 1.0$. The method of using the run properties estimation is used in which p is estimated using Eq. (4-25), and σ_m^2 and σ_z^2 are estimated using Eqs. (4-29) and (4-30) where $i=1$.

Parameter	Sample Length	Population θ	Generated $m(\hat{\theta})$	Bias	RMSE	# Samples
p	50	0.150	0.055	0.095	0.105	108
	100	0.150	0.047	0.103	0.108	102
	200	0.150	0.044	0.106	0.108	100
	500	0.150	0.037	0.113	0.114	100
σ_m^2	50	1.000	0.677	0.323	0.656	108
	100	1.000	0.763	0.237	0.487	102
	200	1.000	0.797	0.203	0.372	100
	500	1.000	0.858	0.142	0.286	100
σ_z^2	50	1.000	1.102	-0.102	0.287	108
	100	1.000	1.134	-0.134	0.252	102
	200	1.000	1.137	-0.137	0.199	100
	500	1.000	1.106	-0.106	0.138	100

Table 4-11   Results of the Generation Experiment for the GNN model with population parameters $p=0.25$, $\sigma_m^2 = 1.0$ and $\sigma_z^2 = 1.0$. The method of Moments (MOM) estimation are used in which p is estimated using Eq. (4-28) where $i=1$, $k=2$, and σ_m^2 and σ_z^2 are estimated using Eqs. (4-29) and (4-30) .

Parameter	Sample Length	Population θ	Generated $m(\hat{\theta})$	Bias	RMSE	# Samples
p	50	0.250	0.372	-0.122	0.240	205
	100	0.250	0.341	-0.091	0.221	133
	200	0.250	0.323	-0.073	0.164	114
	500	0.250	0.273	-0.023	0.103	101
σ_m^2	50	1.000	1.175	-0.175	0.634	205
	100	1.000	1.090	-0.090	0.437	133
	200	1.000	1.094	-0.094	0.423	114
	500	1.000	1.039	-0.039	0.263	101
σ_z^2	50	1.000	0.818	0.182	0.470	205
	100	1.000	0.841	0.156	0.437	133
	200	1.000	0.874	0.126	0.312	114
	500	1.000	0.953	0.047	0.185	101

Table 4-12   Results of the Generation Experiment for the GNN model with population parameters

$p = 0.25$, $\sigma_m^2 = 1.0$ and $\sigma_z^2 = 1.0$. The method of regression analysis estimation

is used in which p is estimated using Eq. (4-34) where the number of lags (L) = 20, and

σ_m^2 and σ_z^2 are estimated using Eqs. (4-29) and (4-30) where i = 1.

Parameter	Sample Length	Population θ	Generated $m(\hat{\theta})$	Bias	RMSE	# Samples
p	50	0.250	0.118	0.132	0.164	112
	100	0.250	0.105	0.145	0.150	100
	200	0.250	0.136	0.114	0.144	102
	500	0.250	0.156	0.094	0.118	101
σ_m^2	50	1.000	0.731	0.269	0.602	112
	100	1.000	0.723	0.277	0.471	100
	200	1.000	0.824	0.176	0.394	102
	500	1.000	0.886	0.114	0.266	101
σ_z^2	50	1.000	1.126	-0.126	0.380	112
	100	1.000	1.156	-0.156	0.302	100
	200	1.000	1.131	-0.131	0.238	102
	500	1.000	1.100	-0.100	0.181	101

Table 4-13   Results of the Generation Experiment for the GNN model with population parameters $p=0.25$, $\sigma_m^2 = 1.0$ and $\sigma_z^2 = 1.0$. The method of fitting the autocorrelation function is used in which the number of lags used (L)=20, p is estimated using Eq. (4-28) where i=1, k=2, and σ_m^2 and σ_z^2 are estimated using Eqs. (4-29) and (4-30).

Parameter	Sample Length	Population θ	Generated $m(\hat{\theta})$	Bias	RMSE	# Samples
p	50	0.250	0.114	0.136	0.164	113
	100	0.250	0.103	0.147	0.159	109
	200	0.250	0.130	0.120	0.144	105
	500	0.250	0.152	0.098	0.119	103
σ_m^2	50	1.000	0.498	0.502	0.718	113
	100	1.000	0.454	0.546	0.646	109
	200	1.000	0.559	0.441	0.595	105
	500	1.000	0.597	0.403	0.507	103
σ_z^2	50	1.000	1.347	-0.347	0.576	113
	100	1.000	1.418	-0.418	0.542	109
	200	1.000	1.382	-0.382	0.532	105
	500	1.000	1.383	-0.383	0.487	103

Table 4-14   Results of the Generation Experiment for the GNN model with population parameters

$p=0.25$, $\sigma_m^2 = 1.0$ and $\sigma_z^2 = 1.0$. The method of using the range properties

estimation is used in which p is estimated using Eq. (4-40) where the number of lags (L)

= 20, and σ_m^2 and σ_z^2 are estimated using Eqs. (4-29) and (4-30) where i=1.

Parameter	Sample Length	Population θ	Generated $m(\hat{\theta})$	Bias	RMSE	# Samples
p	50	0.250	0.234	0.016	0.105	103
	100	0.250	0.232	0.018	0.102	100
	200	0.250	0.245	0.005	0.084	100
	500	0.250	0.234	0.016	0.056	100
σ_m^2	50	1.000	0.798	0.202	0.592	103
	100	1.000	0.837	0.163	0.490	100
	200	1.000	0.938	0.062	0.384	100
	500	1.000	0.969	0.031	0.239	100
σ_z^2	50	1.000	1.043	-0.043	0.402	103
	100	1.000	1.023	-0.023	0.315	100
	200	1.000	1.021	-0.021	0.223	100
	500	1.000	0.964	0.036	0.093	100

Table 4-15   Results of the Generation Experiment for the GNN model with population parameters $p=0.25$, $\sigma_m^2 = 1.0$ and $\sigma_z^2 = 1.0$. The method of using the run properties estimation is used in which p is estimated using Eq. (4-25), and σ_m^2 and σ_z^2 are estimated using Eqs. (4-29) and (4-30) where $i=1$.

Parameter	Sample Length	Population θ	Generated $m(\hat{\theta})$	Bias	RMSE	# Samples
p	50	0.250	0.062	0.188	0.193	102
	100	0.250	0.058	0.192	0.194	100
	200	0.250	0.048	0.202	0.203	100
	500	0.250	0.044	0.206	0.206	100
σ_m^2	50	1.000	0.619	0.381	0.557	102
	100	1.000	0.651	0.349	0.479	100
	200	1.000	0.726	0.274	0.382	100
	500	1.000	0.773	0.227	0.291	100
σ_z^2	50	1.000	1.234	-0.234	0.391	102
	100	1.000	1.209	-0.209	0.322	100
	200	1.000	1.232	-0.232	0.276	100
	500	1.000	1.213	-0.213	0.234	100

Table 4-16: Estimated GNN model parameters for the original and transformed data for Malakal. The estimation methods are:

Method 1: Using MOM (i and k of Eq. (4-28) and (4-29))

Method 2: Using Regression analysis (L=number of lags used in regression, i=lag used in Eq. (4-29))

Method 3: Using fitted autocorrelation function (L=number of lags used in fitting the autocorrelation function)

Method 4: Using range poperties (L=number of lags used in approximating the infinite sum of Eq. (4-41), i=lag used in Eq. (4-29))

	Original data			Transformed data		
	p	σ^2_m	σ^2_z	p	σ^2_m	σ^2_z
METHOD 1 (i=1,k=8)	0.199	5.376	0.941	0.144	0.546	0.228
METHOD 1 (i=1,k=9)	0.178	5.305	1.284	0.116	0.537	0.248
METHOD 1 (i=2,k=6)	0.277	5.268	1.427	0.221	0.518	0.286
METHOD 2 (i=1, L=30)	0.116	5.117	1.898	0.099	0.532	0.259
METHOD 3 (L=30)	0.116	3.606	4.097	0.099	0.411	0.426
METHOD 4 (i=1, L=30)	0.220	5.448	0.328	0.179	0.557	0.199

Table 4-17: Comparison of historical and generated mean in milliards of cubic meters (mcm), st.deviation (mcm),and skewness coefficient based on the GNN model fitted to the original data of annual flows at Malakal. Eestimation methods:

Method 1: Using MOM (i and k of Eq. (4-28) and (4-29))

Method 2: Using Regression analysis (L=number of lags used in regression, i=lag used in Eq. (4-29))

Method 3: Using fitted autocorrelation function (L=number of lags used in fitting the autocorrelation function)

Method 4: Using range properties (L=number of lags used in approximating the infinite sum of Eq. (4-41), i=lag used in Eq. (4-29)). Note that GEN means the average of generated values.

	MEAN (mcm)			STANDARD DEVIATION (mcm)			SKEWNESS COEF		
	GEN - STD	GEN	GEN+ STD	GEN - STD	GEN	GEN+ STD	GEN - STD	GEN	GEN + STD
HISTORICAL		29.81			5.46			1.41	
METHOD 1 (i=1,k=8)	27.89	29.88	31.88	3.91	5.14	5.89	-0.51	-0.06	0.40
METHOD 1 (i=1,k=9)	27.94	29.96	31.98	3.73	4.94	6.15	-0.52	0.00	0.52
METHOD 1 (i=1,k=10)	28.13	29.91	31.68	3.93	5.07	6.21	-0.59	-0.05	0.49
METHOD 1 (i=2,k=6)	28.35	29.89	31.43	4.37	5.19	6.00	-0.50	-0.01	0.47
METHOD 2 (i=1, L=30)	27.70	29.64	31.59	3.50	4.66	5.83	-0.56	0.05	0.67
METHOD 3 (L=30)	28.31	29.72	31.14	4.39	5.09	5.80	-0.38	-0.04	0.30
METHOD 4 (i=1, L=30)	27.93	29.72	31.50	3.94	5.04	6.14	-0.57	0.04	0.65

Table 4-18: Comparison of historical and generated lag-1 autocorrelation coefficient and the partial sum of the autocorrelation function for lags 10 and 20 based on GNN model fitted to original data of annual flows at Malakal. Estimation methods:
Method 1: Using MOM (i and k of Eq. (4-28) and (4-29))
Method 2: Using Regression analysis (L=number of lags used in regression, i=lag used in Eq. (4-29))
Method 3: Using fitted autocorrelation function (L=number of lags used in fitting the autocorrelation function)
Method 4: Using range properties (L=number of lags used in approximating the infinite sum of Eq. (4-41), i=lag used in Eq. (4-29)). Note that GEN means the average of generated values.

	$\rho_1(x)$			$\sum\limits_{i=1}^{10} \rho_i(x)$			$\sum\limits_{i=1}^{20} \rho_i(x)$		
	GEN-STD	GEN	GEN+STD	GEN-STD	GEN	GEN+STD	GEN-STD	GEN	GEN+STD
HISTORICAL		0.777			2.604			3.288	
METHOD 1 (i=1,k=8)	0.625	0.721	0.817	0.916	2.411	3.906	-0.495	1.568	3.632
METHOD 1 (i=1,k=9)	0.599	0.703	0.807	0.685	2.359	4.032	-0.335	1.984	4.302
METHOD 1 (i=1,k=10)	0.604	0.710	0.815	0.945	2.511	4.078	-0.321	1.852	4.025
METHOD 1 (i=2,k=6)	0.538	0.634	0.730	0.278	1.622	2.966	-0.695	1.013	2.722
METHOD 2 (i=1, L=30)	0.508	0.652	0.797	1.321	2.963	4.606	-0.162	2.497	5.157
METHOD 3 (L=30)	0.122	0.276	0.430	0.132	1.180	2.228	-0.530	0.922	2.374
METHOD 4 (i=1, L=30)	0.637	0.724	0.810	0.930	2.229	3.527	-0.510	1.407	3.324

112

Table 4-19: Comparison of historical and generated longest drought, rescaled range, and Hurst coefficient based on GNN model fitted to original data of annual flows at Malakal. Estimation methods:

Method 1: Using MOM (i and k of Eq. (4-28) and (4-29))

Method 2: Using Regression analysis (L=number of lags used in regression, i=lag used in Eq. (4-29))

Method 3: Using fitted autocorrelation function (L=number of lags used in fitting the autocorrelation function)

Method 4: Using range prperties (L=number of lags used in approximating the infinite sum of Eq. (4-41), i=lag used in Eq. (4-29)). Note that GEN means the average of generated values.

	LONGEST DROUGHT (years)			RESCALED RANGE			HURST COEFICIENT		
	GEN-STD	GEN	GEN+STD	GEN-STD	GEN	GEN+STD	GEN-STD	GEN	GEN+STD
HISTORICAL		14.00			24.42			0.872	
METHOD 1 (i=1,k=8)	10.15	16.49	22.83	17.95	22.40	26.85	0.788	0.843	0.898
METHOD 1 (i=1,k=9)	9.23	15.89	22.55	16.94	22.26	27.58	0.771	0.839	0.907
METHOD 1 (i=1,k=10)	10.30	15.49	20.68	17.86	22.65	27.43	0.785	0.845	0.906
METHOD 1 (i=2,k=6)	8.05	12.59	17.13	14.91	19.32	23.74	0.739	0.801	0.864
METHOD 2 (i=1, L=30)	9.49	16.15	22.81	18.18	23.50	28.81	0.791	0.855	0.918
METHOD 3 (L=30)	5.45	8.43	11.41	12.25	16.69	21.14	0.683	0.758	0.834
METHOD 4 (i=1, L=30)	10.47	17.38	24.29	16.96	21.36	25.75	0.774	0.830	0.886

113

Table 4-20: Estimated ARMA(p,q) model parameters for the original and transformed data for Malakal based on the method of moments and method of least squares(*).

	Original data						Transformed data					
	ϕ_1	ϕ_2	ϕ_3	θ_1	θ_2	σ_ϵ^2	ϕ_1	ϕ_2	ϕ_3	θ_1	θ_2	σ_ϵ^2
ARMA (1,0)	0.777	------	------	------	------	11.814	0.729	------	------	------	------	0.164
ARMA (1,0) *	0.779	------	------	------	------	11.580	0.734	------	------	------	------	0.161
ARMA (2,0)	1.006	-0.295	------	------	------	10.787	0.831	-0.140	------	------	------	0.161
ARMA (2,0) *	1.000	-0.289	------	------	------	10.479	0.817	-0.126	------	------	------	0.152
ARMA (1,1)	0.626	------	------	-0.423	------	10.590	0.639	------	------	-0.196	------	0.161
ARMA (1,1) *	0.640	------	------	-0.401	------	10.552	0.660	------	------	-0.168	------	0.159
ARMA (2,1)	0.716	-0.070	------	-0.320	------	10.682	0.457	0.132	------	-0.382	------	0.160
ARMA (2,1) *	0.114	0.464	------	-0.929	------	10.172	0.258	0.279	------	-0.625	------	0.148
ARMA (2,2) *	0.059	0.452	------	-1.001	-0.116	10.114	0.289	0.277	------	-0.584	0.038	0.148
ARMA (3,0)	1.033	-0.389	0.093	------	------	10.693	0.838	-0.184	0.053	------	------	0.161
ARMA (3,0) *	1.028	-0.369	0.077	------	------	10.382	0.848	-0.162	0.027	------	------	0.149
ARMA (3,1)	0.397	0.251	-0.094	-0.646	------	10.627	0.255	0.261	0.025	-0.611	------	0.148

Table 4-21: Comparison of historical and generated mean in milliards of cubic meters (mcm), st.deviation (mcm), and skewness coefficient based on ARMA(p,q) model fitted to the original data of annual flows at Malakal. Estimation methods: method of moments and least squares (*).

	MEAN (mcm)			STANDARD DEVIATION (mcm)			SKEWNESS COEF		
	GEN-STD	GEN	GEN+STD	GEN-STD	GEN	GEN+STD	GEN-STD	GEN	GEN+STD
HISTORICAL		29.81			5.46			1.41	
ARMA (1,0)	28.05	29.92	31.78	4.44	5.11	5.78	-0.31	0.06	0.43
ARMA (1,0) *	28.05	29.92	31.79	4.42	5.08	5.75	-0.31	0.06	0.43
ARMA (2,0)	28.59	29.87	31.14	4.59	5.18	5.81	-0.34	0.00	0.33
ARMA (2,0) *	28.61	29.87	31.12	4.47	5.08	5.70	-0.34	-0.01	0.33
ARMA (1,1)	28.38	29.90	31.42	4.58	5.19	5.80	-0.36	0.06	0.43
ARMA (1,1) *	28.35	29.90	31.45	4.58	5.20	5.81	-0.31	0.06	0.43
ARMA (2,1)	28.52	29.88	31.23	4.53	5.16	5.79	-0.35	-0.01	0.32
ARMA (2,1) *	28.29	29.89	31.50	4.52	5.21	5.89	-0.36	-0.02	0.32
ARMA (2,2) *	28.36	29.89	31.41	4.54	5.21	5.88	-0.35	-0.02	0.32
ARMA (3,0)	28.46	29.89	31.32	4.51	5.16	5.80	-0.32	0.02	0.35
ARMA (3,0) *	28.49	29.89	31.29	4.46	5.09	5.73	-0.32	0.02	0.35
ARMA (3,1)	28.51	29.90	31.28	4.51	5.17	5.82	-0.33	0.01	0.34

Table 4-22: Comparison of historical and generated lag-1 autocorrelation coefficient and the partial sum of the autocorrelation function for lags 10 and 20 based on ARMA(p,q) model fitted to the annual flows at Malakal. Estimation methods: method of moments and least squares (*).

	$\rho_1(x)$			$\sum\limits_{i=1}^{10} \rho_i(x)$			$\sum\limits_{i=1}^{20} \rho_i(x)$		
	GEN-STD	GEN	GEN+STD	GEN-STD	GEN	GEN+STD	GEN-STD	GEN	GEN+STD
HISTORICAL		0.777			2.604			3.288	
ARMA (1,0)	0.659	0.733	0.806	1.151	2.288	3.425	-0.244	1.476	3.196
ARMA (1,0) *	0.662	0.735	0.808	1.173	2.316	3.459	-0.235	1.501	3.236
ARMA (2,0)	0.683	0.742	0.800	0.215	1.173	2.132	-0.484	0.692	1.867
ARMA (2,0) *	0.681	0.740	0.799	0.224	1.185	2.145	-0.481	0.700	1.881
ARMA (1,1)	0.691	0.746	0.802	0.548	1.494	2.440	-0.460	0.812	2.083
ARMA (1,1) *	0.695	0.750	0.806	0.592	1.557	2.521	-0.450	0.857	2.163
ARMA (2,1)	0.678	0.739	0.800	0.391	1.394	2.397	-0.426	0.859	2.144
ARMA (2,1) *	0.666	0.737	0.808	0.834	1.965	3.096	-0.264	1.331	2.925
ARMA (2,2) *	0.677	0.743	0.810	0.687	1.780	2.872	-0.325	1.164	2.653
ARMA (3,0)	0.681	0.740	0.798	0.386	1.419	2.451	-0.472	0.858	2.189
ARMA (3,0) *	0.685	0.743	0.801	0.370	1.402	2.433	-0.478	0.843	2.164
ARMA (3,1)	0.681	0.740	0.800	0.325	1.336	2.347	-0.468	0.810	2.088

116

Table 4-23: Comparison of historical and generated longest drought, rescaled range, and Hurst coefficient based on ARMA(p,q) model fitted to the original data of annual flows at Malakal. Estimation methods: method of moments and least squares (*).

	LONGEST DROUGHT (years)			RESCALED RANGE			HURST COEFICIENT		
	GEN-STD	GEN	GEN+STD	GEN-STD	GEN	GEN+STD	GEN-STD	GEN	GEN+STD
HISTORICAL		14.00			24.42			0.872	
ARMA (1,0)	9.04	13.40	17.76	17.21	20.97	24.73	0.775	0.826	0.877
ARMA (1,0) *	9.12	13.46	17.80	17.27	21.04	24.82	0.776	0.827	0.878
ARMA (2,0)	8.34	11.86	15.38	14.53	18.22	21.90	0.728	0.786	0.845
ARMA (2,0) *	8.39	11.90	15.41	14.55	18.25	21.94	0.728	0.787	0.845
ARMA (1.1)	8.37	11.80	15.39	15.45	18.84	22.23	0.746	0.797	0.848
ARMA (1.1) *	8.23	12.19	16.15	15.61	19.04	22.46	0.749	0.800	0.851
ARMA (2,1)	8.25	11.97	15.69	14.96	18.82	22.68	0.736	0.795	0.854
ARMA (2,1) *	9.02	13.14	17.26	16.24	20.41	24.59	0.759	0.817	0.875
ARMA (2,2) *	8.78	12.84	16.90	15.81	19.92	24.03	0.752	0.811	0.869
ARMA (3,0)	8.59	12.21	15.83	15.04	18.95	22.85	0.739	0.797	0.855
ARMA (3,0) *	8.65	12.21	15.77	15.03	18.92	22.81	0.739	0.797	0.855
ARMA (3,1)	8.64	12.18	15.72	14.93	18.78	22.62	0.736	0.794	0.853

117

Table 4-24: Comparison of historical and generated mean in milliards of cubic meters (mcm), st.deviation (mcm), and skewness coefficient based on the GNN model fitted to the transformed data of annual flows at Malakal.
Estimation methods:
Method 1: Using MOM (i and k of Eq. (4-28) and (4-29))
Method 2: Using Regression analysis (L=number of lags used in regression, i=lag used in Eq. (4-29))
Method 3: Using fitted autocorrelation function (L=number of lags used in fitting the autocorrelation function)
Method 4: Using range properties (L=number of lags used in approximating the infinite sum of Eq. (4-41), i=lag used in Eq. (4-29)). Note that GEN means the average of generated values.

	MEAN (mcm)			STANDARD DEVIATION (mcm)			SKEWNESS COEF		
	GEN-STD	GEN	GEN+STD	GEN-STD	GEN	GEN+STD	GEN-STD	GEN	GEN+STD
HISTORICAL		29.81			5.46			1.41	
METHOD 1 (i=1,k=8)	27.76	29.96	32.16	2.87	5.03	7.19	0.40	1.13	1.87
METHOD 1 (i=1,k=9)	27.67	29.56	31.44	2.45	4.64	6.83	0.45	1.18	1.93
METHOD 1 (i=1,k=10)	27.70	29.87	32.03	2.97	4.90	6.83	0.45	1.21	1.96
METHOD 1 (i=2,k=6)	28.12	29.77	31.43	3.45	5.20	6.95	0.84	1.39	1.93
METHOD 2 (i=1, L=30)	27.41	29.69	31.97	2.74	4.62	6.50	0.43	1.15	1.87
METHOD 3 (L=30)	27.99	29.76	31.54	3.62	5.09	6.57	0.79	1.50	2.22
METHOD 4 (i=1, L=30)	27.82	29.68	31.53	3.03	4.83	6.63	0.51	1.32	2.12

Table 4-25: Comparison of historical and generated lag-1 autocorrelation coefficient and the partial sum of the autocorrelation function for lags 10 and 20 based on the GNN model fitted to the transformed data of annual flows at Malakal. Estimation methods:
Method 1: Using MOM (i and k of Eq. (4-28) and (4-29))
Method 2: Using Regression analysis (L=number of lags used in regression, i=lag used in Eq. (4-29))
Method 3: Using fitted autocorrelation function (L=number of lags used in fitting the autocorrelation function)
Method 4: Using range properties (L=number of lags used in approximating the infinite sum of Eq. (4-41), i=lag used in Eq. (4-29)). Note that GEN means the average of generated values.

	$\rho_1(x)$			$\sum_{i=1}^{10} \rho_i(x)$			$\sum_{i=1}^{20} \rho_i(x)$		
	GEN- STD	GEN	GEN+ STD	GEN- STD	GEN	GEN+ STD	GEN - STD	GEN	GEN + STD
HISTORICAL		0.777			2.604			3.288	
METHOD 1 (i=1,k=8)	0.499	0.611	0.723	1.103	2.508	3.913	0.183	2.205	4.227
METHOD 1 (i=1,k=9)	0.395	0.555	0.715	0.968	2.492	4.016	-0.225	2.079	4.382
METHOD 1 (i=1,k=10)	0.484	0.595	0.706	1.017	2.358	3.699	-0.079	1.866	3.811
METHOD 1 (i=2,k=6)	0.358	0.490	0.622	0.378	1.434	2.489	-0.638	0.815	2.269
METHOD 2 (i=1, L=30)	0.417	0.558	0.699	1.227	2.552	3.877	0.257	2.360	4.463
METHOD 3 (L=30)	0.128	0.285	0.441	0.278	1.204	2.130	-0.217	1.047	2.311
METHOD 4 (i=1, L=30)	0.469	0.599	0.728	0.596	1.942	3.289	-0.356	1.472	3.301

Table 4-26: Comparison of historical and generated longest drought, rescaled range, and Hurst coefficient based on the GNN model fitted to the transformed data of annual flows at Malakal. Estimation methods:
Method 1: Using MOM (i and k of Eq. (4-28) and (4-29))
Method 2: Using Regression analysis (L=number of lags used in regression, i=lag used in Eq. (4-29))
Method 3: Using fitted autocorrelation function (L=number of lags used in fitting the autocorrelation function)
Method 4: Using range prperties (L=number of lags used in approximating the infinite sum of Eq. (4-41), i=lag used in Eq. (4-29)). Note that GEN means the average of generated values.

	LONGEST DROUGHT (years)			RESCALED RANGE			HURST COEFICIENT		
	GEN- STD	GEN	GEN+ STD	GEN- STD	GEN	GEN+ STD	GEN - STD	GEN	GEN + STD
HISTORICAL		14.00			24.42			0.872	
METHOD 1 (i=1,k=8)	10.82	18.34	25.87	17.58	22.56	27.54	0.778	0.843	0.908
METHOD 1 (i=1,k=9)	9.65	18.96	28.27	16.56	21.87	27.18	0.762	0.833	0.905
METHOD 1 (i=1,k=10)	11.12	17.82	24.52	17.00	21.90	26.79	0.771	0.835	0.899
METHOD 1 (i=2,k=6)	9.61	15.54	21.48	14.02	18.21	22.41	0.720	0.785	0.849
METHOD 2 (i=1, L=30)	12.11	19.85	27.59	17.94	22.70	27.46	0.788	0.846	0.904
METHOD 3 (L=30)	7.33	13.04	18.75	13.58	17.50	21.42	0.712	0.774	0.836
METHOD 4 (i=1, L=30)	10.41	17.86	25.31	15.65	20.46	25.27	0.751	0.816	0.882

120

Table 4-27: Comparison of historical and generated mean in milliards of cubic meters (mcm), st.deviation (mcm), and skewness coefficient based on ARMA(p,q) model fitted to the transformed data of annual flows at Malakal. Estimation methods: method of moments and least squares (*).

	MEAN (mcm)			STANDARD DEVIATION (mcm)			SKEWNESS COEF		
	GEN-STD	GEN	GEN+STD	GEN-STD	GEN	GEN+STD	GEN-STD	GEN	GEN+STD
HISTORICAL		29.81			5.46			1.41	
ARMA (1,0)	28.32	29.92	31.52	3.96	5.37	6.78	0.84	1.53	2.21
ARMA (1,0) *	28.30	29.92	31.54	3.95	5.36	6.77	0.84	1.52	2.20
ARMA (2,0)	28.51	29.84	31.18	3.84	5.29	6.75	0.76	1.43	2.10
ARMA (2,0) *	28.46	29.74	31.01	3.69	5.03	6.36	0.74	1.38	2.02
ARMA (1,1)	28.47	29.90	31.33	4.06	5.43	6.80	0.85	1.58	2.30
ARMA (1,1) *	28.44	29.91	31.39	4.05	5.44	6.82	0.85	1.57	2.28
ARMA (2,1)	28.48	29.85	31.28	3.81	5.28	6.74	0.76	1.42	2.09
ARMA (2,1) *	28.44	29.79	31.15	3.71	5.12	6.54	0.75	1.39	2.04
ARMA (2,2) *	28.42	29.79	31.16	3.69	5.10	6.51	0.73	1.38	2.03
ARMA (3,0)	28.45	29.88	31.31	3.80	5.32	6.85	0.79	1.46	2.13
ARMA (3,0) *	28.42	29.82	31.22	3.71	5.18	6.65	0.77	1.42	2.07
ARMA (3,1)	28.39	29.80	31.22	3.63	5.11	6.58	0.74	1.39	2.05

Table 4-28: Comparison of historical and generated lag-1 autocorrelation coefficient and the partial sum of the autocorrelation function for lags 10 and 20 based on ARMA(p,q) model fitted to the transformed data of annual flows at Malakal. Estimation methods: method of moments and least squares (*).

	$\rho_1 (x)$			$\sum\limits_{i=1}^{10} \rho_j (x)$			$\sum\limits_{i=1}^{20} \rho_j (x)$		
	GEN-STD	GEN	GEN+STD	GEN-STD	GEN	GEN+STD	GEN-STD	GEN	GEN+STD
HISTORICAL		0.777			2.604			3.288	
ARMA (1,0)	0.569	0.649	0.729	0.727	1.624	2.522	-0.275	0.991	2.256
ARMA (1,0) *	0.575	0.654	0.734	0.755	1.665	2.575	-0.266	1.023	2.312
ARMA (2,0)	0.562	0.644	0.727	0.322	1.142	1.962	-0.438	0.671	1.780
ARMA (2,0) *	0.559	0.642	0.726	0.341	1.171	2.001	-0.434	0.693	1.819
ARMA (1.1)	0.588	0.657	0.727	0.477	1.277	2.078	-0.349	0.718	1.784
ARMA (1,1) *	0.593	0.663	0.733	0.536	1.363	2.189	-0.333	0.781	1.895
ARMA (2,1)	0.559	0.642	0.725	0.407	1.249	2.091	-0.407	0.755	1.917
ARMA (2,1) *	0.571	0.652	0.733	0.452	1.313	2.174	-0.392	0.805	2.002
ARMA (2,2) *	0.567	0.649	0.732	0.491	1.362	2.233	-0.377	0.846	2.068
ARMA (3,0)	0.560	0.643	0.725	0.384	1.247	2.109	-0.438	0.748	1.935
ARMA (3,0) *	0.578	0.658	0.738	0.395	1.273	2.152	-0.442	0.762	1.967
ARMA (3,1)	0.562	0.645	0.729	0.441	1.320	2.199	-0.401	0.819	2.039

122

Table 4-29: Comparison of historical and generated longest drought, rescaled range, and Hurst coefficient based on ARMA(p,q) model fitted to the transformed data of annual flows at Malakal. Estimation methods: method of moments and least squares (*).

	LONGEST DROUGHT (years)			RESCALED RANGE			HURST COEFICIENT		
	GEN- STD	GEN	GEN+ STD	GEN- STD	GEN	GEN+ STD	GEN - STD	GEN	GEN + STD
HISTORICAL		14.00			24.42			0.872	
ARMA (1,0)	10.42	15.42	20.42	15.36	18.91	22.46	0.744	0.797	0.850
ARMA (1,0) *	10.44	15.43	20.43	15.46	19.04	22.61	0.746	0.799	0.852
ARMA (2,0)	10.06	14.59	19.12	14.18	17.77	21.36	0.720	0.779	0.839
ARMA (2,0) *	9.97	14.46	18.95	14.25	17.87	21.48	0.721	0.781	0.840
ARMA (1,1)	9.57	14.39	19.21	14.59	17.88	21.17	0.731	0.782	0.834
ARMA (1,1) *	9.72	14.68	19.64	14.81	18.17	21.52	0.735	0.787	0.839
ARMA (2,1)	9.92	14.69	19.46	14.44	18.10	21.78	0.726	0.784	0.843
ARMA (2,1) *	10.13	15.14	20.15	14.60	18.32	22.04	0.729	0.788	0.847
ARMA (2,2) *	10.29	15.20	20.11	14.70	18.46	22.22	0.731	0.790	0.849
ARMA (3,0)	9.90	14.94	19.98	14.43	18.11	21.79	0.727	0.785	0.843
ARMA (3,0) *	10.26	15.14	20.03	14.57	18.27	21.97	0.729	0.787	0.845
ARMA (3,1)	10.24	15.13	20.02	14.67	18.39	22.11	0.731	0.789	0.847

123

Table 4-30: Comparison of historical and generated mean, st.deviation, and skewness coefficient based on the GNN model fitted to original data of annual NBS at Lake St.Clair. Estimation methods:
Method 1: Using MOM (i and k of Eq. (4-28) and (4-29))
Method 2: Using Regression analysis (L=number of lags used in regression, i=lag used in Eq. (4-29))
Method 3: Using fitted autocorrelation function (L=number of lags used in fitting the autocorrelation function)
Method 4: Using range properties (L=number of lags used in approximating the infinite sum of Eq. (4-41), i=lag used in Eq. (4-29)). Note that GEN means the average of generated values.

	MEAN (thousands cfs)			STANDARD DEVIATION (thousands cfs)			SKEWNESS COEF		
	GEN-STD	GEN	GEN+STD	GEN-STD	GEN	GEN+STD	GEN-STD	GEN	GEN+STD
HISTORICAL		51.71			26.87			0.31	
METHOD 1 (i=1,k=2)	46.96	51.74	56.52	22.90	26.12	29.34	-0.37	0.03	0.42
METHOD 1 (i=1,k=4)	43.43	51.40	59.47	20.60	24.57	28.54	-0.32	-0.03	0.26
METHOD 1 (i=2,k=3)	46.96	52.31	57.66	23.19	26.13	29.08	-0.31	-0.04	0.24
METHOD 1 (i=3,k=6)	45.99	52.10	58.20	22.66	26.08	29.51	-0.24	0.02	0.28
METHOD 1 (i=4,k=8)	44.70	52.95	61.19	20.34	24.76	29.17	-0.57	-0.05	0.48
METHOD 1 (i=4,k=9)	43.55	51.93	60.37	21.36	24.69	28.02	-0.37	-0.04	0.30
METHOD 2 (i=1, L=30)	42.79	52.84	62.89	19.94	24.15	28.36	-0.33	-0.01	0.31
METHOD 3 (L=30)	45.77	52.52	59.27	23.33	25.77	28.21	-0.27	-0.02	0.23
METHOD 4 (i=1, L=30)	44.60	52.25	59.90	21.78	25.41	29.03	-0.33	-0.02	0.29

124

Table 4-31: Comparison of historical and generated lag-1 autocorrelation coefficient and the partial sum of the autocorrelation function for lags 10 and 20 based on the GNN model fitted to original data of annual NBS at Lake St.Clair. Estimation methods:

Method 1: Using MOM (i and k of Eq. (4-28) and (4-29))[*]

Method 2: Using Regression analysis (L=number of lags used in regression, i=lag used in Eq. (4-29))

Method 3: Using fitted autocorrelation function (L=number of lags used in fitting the autocorrelation function)

Method 4: Using range properties (L=number of lags used in approximating the infinite sum of Eq. (4-41), i=lag used in Eq. (4-29)). Note that GEN means the average of generated values.

	$\rho_1(x)$			$\sum_{i=1}^{10} \rho_i(x)$			$\sum_{i=1}^{20} \rho_i(x)$		
	GEN-STD	GEN	GEN+STD	GEN-STD	GEN	GEN+STD	GEN-STD	GEN	GEN+STD
HISTORICAL		0.503			2.411			2.616	
METHOD 1 (i=1,k=2)	0.370	0.467	0.565	0.035	0.784	1.533	-0.456	0.471	1.399
METHOD 1 (i=1,k=4)	0.232	0.374	0.516	0.581	1.714	2.848	-0.143	1.581	3.305
METHOD 1 (i=2,k=3)	0.163	0.302	0.440	0.240	1.036	1.831	-0.246	0.741	1.728
METHOD 1 (i=3,k=6)	0.254	0.376	0.498	0.028	0.998	1.968	-0.566	0.728	2.022
METHOD 1 (i=4,k=8)	0.576	0.668	0.761	0.971	2.142	3.313	-0.269	1.405	3.078
METHOD 1 (i=4,k=9)	0.200	0.357	0.514	0.501	1.722	2.942	-0.014	1.711	3.436
METHOD 2 (i=1, L=30)	0.164	0.339	0.514	0.682	2.111	3.539	0.159	2.438	4.717
METHOD 3 (L=30)	0.005	0.127	0.249	-0.039	0.760	1.559	-0.395	0.819	2.034
METHOD 4 (i=1, L=30)	0.272	0.424	0.576	0.815	1.982	3.149	0.124	1.805	3.486

Table 4-32: Comparison of historical and generated longest drought, rescaled range, and Hurst coefficient based on GNN model fitted to original data of annual NBS at Lake St.Clair. Estimation methods:
Method 1: Using MOM (i and k of Eq. (4-28) and (4-29))
Method 2: Using Regression analysis (L=number of lags used in regression, i=lag used in Eq. (4-29))
Method 3: Using fitted autocorrelation function (L=number of lags used in fitting the autocorrelation function)
Method 4: Using range properties (L=number of lags used in approximating the infinite sum of Eq. (4-41), i=lag used in Eq. (4-29)). Note that GEN means the average of generated values.

	LONGEST DROUGHT (years)			RESCALED RANGE			HURST COEFICIENT		
	GEN-STD	GEN	GEN+STD	GEN-STD	GEN	GEN+STD	GEN-STD	GEN	GEN+STD
HISTORICAL		9.00			25.04			0.846	
METHOD 1 (i=1,k=2)	6.81	10.03	13.25	13.73	17.37	21.01	0.688	0.744	0.800
METHOD 1 (i=1,k=4)	6.44	10.96	15.49	15.93	21.20	26.46	0.727	0.794	0.861
METHOD 1 (i=2,k=3)	5.79	9.03	12.27	13.96	18.39	22.81	0.690	0.757	0.824
METHOD 1 (i=3,k=6)	6.50	9.78	13.06	13.52	17.80	22.09	0.685	0.749	0.813
METHOD 1 (i=4,k=8)	9.14	14.65	20.16	17.93	22.67	27.40	0.759	0.814	0.869
METHOD 1 (i=4,k=9)	5.41	10.57	15.73	15.83	21.28	26.73	0.727	0.795	0.862
METHOD 2 (i=1, L=30)	6.34	10.99	15.64	17.12	23.13	29.14	0.743	0.816	0.889
METHOD 3 (L=30)	5.00	7.23	9.46	12.52	16.79	21.06	0.665	0.732	0.800
METHOD 4 (i=1, L=30)	6.45	11.60	16.74	16.94	22.28	27.62	0.743	0.808	0.872

Table 4-33: Comparison of historical and generated mean, st.deviation, and skewness coefficient based on ARMA(p,q) model fitted to the original data of annual NBS at Lake St.Clair. Estimation methods: method of moments and least squares (*).

	MEAN (thousands cfs)			STANDARD DEVIATION (thousands cfs)			SKEWNESS COEF		
	GEN-STD	GEN	GEN+STD	GEN-STD	GEN	GEN+STD	GEN-STD	GEN	GEN+STD
HISTORICAL		51.71			26.87			0.31	
ARMA (1,0)	46.71	51.56	56.42	23.76	26.32	28.88	-0.27	0.03	0.32
ARMA (2,0)	45.83	51.26	56.68	23.29	25.19	28.35	-0.26	0.02	0.30
ARMA (1,1)	46.42	51.55	56.68	23.67	26.29	28.92	-0.27	0.03	0.32
ARMA* (2,1)	44.96	51.18	57.41	22.79	25.38	27.98	-0.26	0.01	0.28
ARMA* (2,2)	44.34	51.15	57.96	22.66	25.30	27.94	-0.26	0.01	0.27
ARMA (3,0)	45.21	51.25	57.30	23.12	25.69	28.25	-0.25	0.02	0.29
ARMA (3,1)	43.47	51.25	59.03	22.57	25.26	27.96	-0.25	0.00	0.25

127

Table 4-34: Comparison of historical and generated lag-1 autocorrelation coefficient and the partial sum of the autocorrelation function for lags 10 and 20 based on ARMA(p,q) model fitted to the original data of annual NBS at Lake St.Clair. Estimation methods: method of moments and least squares (*).

	$\rho_1(x)$			$\sum_{i=1}^{10} \rho_i(x)$			$\sum_{i=1}^{20} \rho_i(x)$		
	GEN-STD	GEN	GEN+STD	GEN-STD	GEN	GEN+STD	GEN-STD	GEN	GEN+STD
HISTORICAL		0.503			2.411			2.616	
ARMA (1,0)	0.374	0.472	0.569	0.080	0.735	1.390	-0.508	0.418	1.345
ARMA (2,0)	0.336	0.447	0.558	0.134	0.789	1.443	-0.434	0.414	1.262
ARMA (1,1)	0.366	0.469	0.572	0.164	0.868	1.571	-0.488	0.517	1.523
ARMA* (2,1)	0.305	0.427	0.549	0.372	1.136	1.901	-0.324	0.726	1.776
ARMA* (2,2)	0.306	0.428	0.551	0.484	1.307	2.130	-0.234	0.935	2.104
ARMA (3,0)	0.327	0.443	0.559	0.284	1.019	1.753	-0.376	0.613	1.601
ARMA (3,1)	0.309	0.430	0.552	0.632	1.531	2.429	-0.079	1.245	2.569

Table 4-35: Comparison of historical and generated longest drought, rescaled range, and Hurst coefficient based on ARMA(p,q) model fitted to the original data of annual NBS at Lake St.Clair. Estimation methods: method of moments and least squares (*).

	LONGEST DROUGHT (years)			RESCALED RANGE			HURST COEFICIENT		
	GEN-STD	GEN	GEN+STD	GEN-STD	GEN	GEN+STD	GEN-STD	GEN	GEN+STD
HISTORICAL		9.00			25.04			0.846	
ARMA (1,0)	6.29	9.41	12.53	12.76	16.64	20.52	0.669	0.731	0.794
ARMA (2,0)	6.60	9.58	12.56	13.66	17.11	20.56	0.688	0.741	0.794
ARMA (1,1)	6.47	9.81	13.15	13.20	17.19	21.19	0.678	0.740	0.802
ARMA* (2,1)	6.83	10.02	13.21	14.66	18.46	22.25	0.706	0.760	0.814
ARMA* (2,2)	7.00	10.22	13.47	15.10	19.07	23.05	0.714	0.769	0.824
ARMA (3,0)	7.15	10.07	12.99	14.31	18.04	21.78	0.700	0.754	0.809
ARMA (3,1)	7.06	10.61	14.16	15.57	19.91	24.26	0.721	0.779	0.838

129

Table 4-36: Comparison of historical and generated mean, st.deviation, and skewness coefficient based on the GNN model fitted to transformed data of annual NBS at Lake St.Clair. Estimation methods:
Method 1: Using MOM (i and k of Eq. (4-28) and (4-29))
Method 2: Using Regression analysis (L=number of lags used in regression, i=lag used in Eq. (4-29))
Method 3: Using fitted autocorrelation function (L=number of lags used in fitting the autocorrelation function)
Method 4: Using range properties (L=number of lags used in approximating the infinite sum of Eq. (4-41), i=lag used in Eq. (4-29)). Note that GEN means the average of generated values.

	MEAN (thousands cfs)			STANDARD DEVIATION (thousands cfs)			SKEWNESS COEF		
	GEN-STD	GEN	GEN+STD	GEN-STD	GEN	GEN+STD	GEN-STD	GEN	GEN+STD
HISTORICAL		51.71			26.87			0.31	
METHOD 1 (i=1,k=2)	46.22	51.40	56.58	22.73	26.62	30.52	0.07	0.46	0.85
METHOD 1 (i=1,k=4)	42.81	51.14	59.46	20.52	25.32	30.13	0.09	0.46	0.83
METHOD 1 (i=2,k=3)	46.30	52.55	58.80	23.12	26.30	29.49	0.09	0.43	0.77
METHOD 1 (i=3,k=6)	45.82	52.22	58.63	22.96	26.22	29.47	0.16	0.47	0.78
METHOD 1 (i=4,k=9)	42.67	51.08	59.48	20.59	25.38	30.17	0.06	0.43	0.80
METHOD 2 (i=1, L=30)	41.48	52.31	63.15	19.06	24.45	29.85	0.03	0.36	0.69
METHOD 3 (L=30)	45.05	52.18	59.32	22.88	26.10	29.31	0.12	0.42	0.71
METHOD 4 (i=1, L=30)	44.19	51.97	59.75	21.49	25.36	29.23	0.09	0.44	0.79

Table 4-37: Comparison of historical and generated lag-1 autocorrelation coefficient and the partial sum of the autocorrelation function for lags 10 and 20 based onthe GNN model fitted to the transformed data of annual NBS at Lake St.Clair. Estimation methods:
Method 1: Using MOM (i and k of Eq. (4-28) and (4-29))
Method 2: Using Regression analysis (L=number of lags used in regression, i=lag used in Eq. (4-29))
Method 3: Using fitted autocorrelation function (L=number of lags used in fitting the autocorrelation function)
Method 4: Using range properties (L=number of lags used in approximating the infinite sum of Eq. (4-41), i=lag used in Eq. (4-29)). Note that GEN means the average of generated values.

	$\rho_1(x)$			$\sum_{i=1}^{10} \rho_i(x)$			$\sum_{i=1}^{20} \rho_i(x)$		
	GEN- STD	GEN	GEN+ STD	GEN- STD	GEN	GEN+ STD	GEN - STD	GEN	GEN + STD
HISTORICAL		0.503			2.411			2.616	
METHOD 1 (i=1,k=2)	0.330	0.449	0.567	-0.031	0.734	1.500	-0.613	0.433	1.480
METHOD 1 (i=1,k=4)	0.206	0.387	0.569	0.567	1.926	3.285	-0.108	1.950	4.009
METHOD 1 (i=2,k=3)	0.156	0.284	0.412	0.267	1.080	1.893	-0.323	0.827	1.977
METHOD 1 (i=3,k=6)	0.213	0.346	0.479	0.143	1.001	1.860	-0.538	0.606	1.750
METHOD 1 (i=4,k=9)	0.226	0.393	0.560	0.593	1.920	3.248	-0.115	1.891	3.898
METHOD 2 (i=1, L=30)	0.135	0.326	0.516	0.627	2.053	3.479	0.076	2.344	4.611
METHOD 3 (L=30)	-0.030	0.114	0.259	-0.119	0.684	1.486	-0.525	0.709	1.944
METHOD 4 (i=1, L=30)	0.228	0.391	0.554	0.511	1.654	2.798	-0.277	1.480	3.237

131

Table 4-38: Comparison of historical and generated longest drought, rescaled range, and Hurst coefficient based on GNN model fitted to transformed data of annual NBS at Lake St.Clair. Estimation methods:
Method 1: Using MOM (i and k of Eq. (4-28) and (4-29))
Method 2: Using Regression analysis (L=number of lags used in regression, i=lag used in Eq. (4-29))
Method 3: Using fitted autocorrelation function (L=number of lags used in fitting the autocorrelation function)
Method 4: Using range properties (L=number of lags used in approximating the infinite sum of Eq. (4-41), i=lag used in Eq. (4-29)). Note that GEN means the average of generated values.

	LONGEST DROUGHT (years)			RESCALED RANGE			HURST COEFICIENT		
	GEN- STD	GEN	GEN+ STD	GEN- STD	GEN	GEN+ STD	GEN - STD	GEN	GEN + STD
HISTORICAL		9.00			25.04			0.846	
METHOD 1 (i=1,k=2)	6.86	10.54	14.22	13.13	16.83	20.53	0.680	0.736	0.792
METHOD 1 (i=1,k=4)	6.02	12.48	18.94	15.99	21.86	27.72	0.729	0.801	0.873
METHOD 1 (i=2,k=3)	6.25	9.87	13.49	13.71	18.14	22.57	0.687	0.753	0.820
METHOD 1 (i=3,k=6)	6.83	10.67	14.51	13.82	17.92	22.02	0.692	0.751	0.811
METHOD 1 (i=4,k=9)	6.33	12.80	19.27	16.48	22.06	27.65	0.737	0.804	0.872
METHOD 2 (i=1, L=30)	5.59	11.97	18.35	16.08	22.60	29.11	0.723	0.805	0.891
METHOD 3 (L=30)	4.93	7.65	10.37	11.16	15.93	20.71	0.636	0.715	0.795
METHOD 4 (i=1, L=30)	6.84	12.35	17.86	15.90	20.90	25.89	0.725	0.791	0.856

Table 4-39: Comparison of historical and generated mean and st.deviation based on GNN-1 model fitted to the original data of annual NBS at Lake St.Clair. Estimation methods:

Method 1: Using MOM (i and k of Eq. (4-46 and Eq. (4-47))

Method 2: Using Regression Analysis (L=Number of lags used in regression, i=lag used in Eq.(4-50))

Method 3: Using range poperties (L=number of lags used in approximating the infinite sum of Eq. (4-41), i=lag used in Eq. (4-51)). Note that GEN means the average of generated values.

	MEAN			STANDARD DEVIATION		
	GEN- STD	GEN	GEN+ STD	GEN- STD	GEN	GEN+ STD
HISTORICAL		51.71			26.87	
METHOD 1 (i=2, k=3)	46.77	52.35	57.93	23.02	26.11	29.20
METHOD 1 (i=3, k=6)	45.90	52.12	58.34	22.54	26.05	29.56
METHOD 1 (i=4,k=9)	43.54	51.94	60.33	21.33	24.69	28.04
METHOD 2 (i=4,L=30)	41.82	52.01	62.21	20.80	24.48	28.15
METHOD 2 (i=5,L=30)	41.67	52.01	62.35	20.63	24.39	28.15
METHOD 3 (i=2,L=30)	45.98	52.05	58.12	22.91	26.10	29.30

Table 4-40: Comparison of historical and generated longest drought, rescaled range, and Hurst coefficient based on GNN-1 model fitted to the original data of annual NBS at Lake St.Clair. Estimation methods:
Method 1: Using MOM (i and k of Eq. (4-46 and Eq. (4-47))
Method 2: Using Regression Analysis (L=Number of lags used in regression, i=lag used in Eq.(4-50))
Method 3: Using range properties (L=number of lags used in approximating the   infinite sum of Eq. (4-41), i=lag used in Eq. (4-51)). Note that GEN means the average of generated values.

	LONGEST DROUGHT (years)			RESCALED RANGE			HURST COEFICIENT		
	GEN-STD	GEN	GEN+STD	GEN-STD	GEN	GEN+STD	GEN-STD	GEN	GEN+STD
HISTORICAL		9.00			25.04			0.846	
METHOD 1 (i=2, k=3)	6.17	9.38	12.59	14.61	19.13	23.66	0.702	0.767	0.833
METHOD 1 (i=3, k=6)	6.92	10.10	13.26	13.90	18.20	22.49	0.693	0.755	0.817
METHOD 1 (i=4,k=9)	5.54	10.65	15.76	15.94	21.37	26.79	0.729	0.796	0.863
METHOD 2 (i=4,L=30)	5.15	9.74	14.33	14.78	21.29	27.81	0.701	0.790	0.878
METHOD 2 (i=5,L=30)	5.26	9.93	14.60	14.90	21.53	28.16	0.703	0.792	0.882
METHOD 3 (i=2,L=30)	6.50	10.28	14.00	15.44	20.28	25.13	0.718	0.783	0.848

134

Table 4-41:   Comparison of historical and generated statistics based on the GNN model

fitted to the original data of annual NBS at Lake Ontario. Estimation

method: Using Regression Analysis (i=1, L=30). Note that GEN means

the average of generated values.

	Historical	GEN-STD	GENERATED	GEN+STD
Mean	432.56	396.15	431.68	467.22
Standard Deviation	97.22	80.57	89.87	99.16
Longest Drought	10.00	4.64	7.93	11.22
Rescaled Range	17.49	11.33	17.20	23.08
Hurst Coefficient	0.752	0.639	0.732	0.824
Surplus	1700.00	787.61	1478.87	2170.13
Maximum Deficit	777.56	365.63	733.63	1101.44

Table 4-42: Comparison of historical and generated statistics based on the GNN-1 model fitted to the original data of annual NBS at Lake Ontario. Estimation method: Using Regression Analysis (i=2, L=30). Note that GEN means the average of generated values.

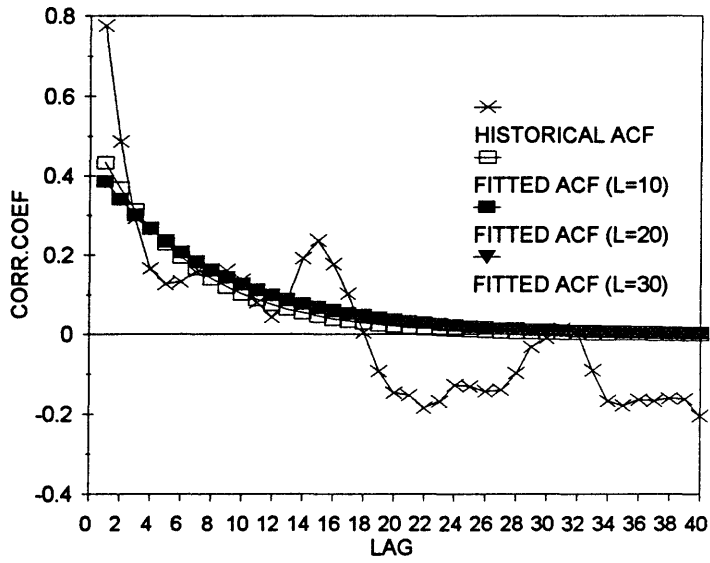|	Historical	GEN-STD	GENERATED	GEN+STD
Mean	432.56	396.85	432.49	468.12
Standard Deviation	97.22	80.28	89.15	98.02
Longest Drought	10.00	4.64	7.06	9.48
Rescaled Range	17.49	10.90	15.60	20.31
Hurst Coefficient	0.752	0.626	0.709	0.792
Surplus	1700.00	820.14	1338.92	1857.71
Maximum Deficit	777.56	391.92	636.89	881.87

Fig 4-1   Historical and fitted autocorrelation function (ACF)
           for the annual flows of the white Nile River at Malakal
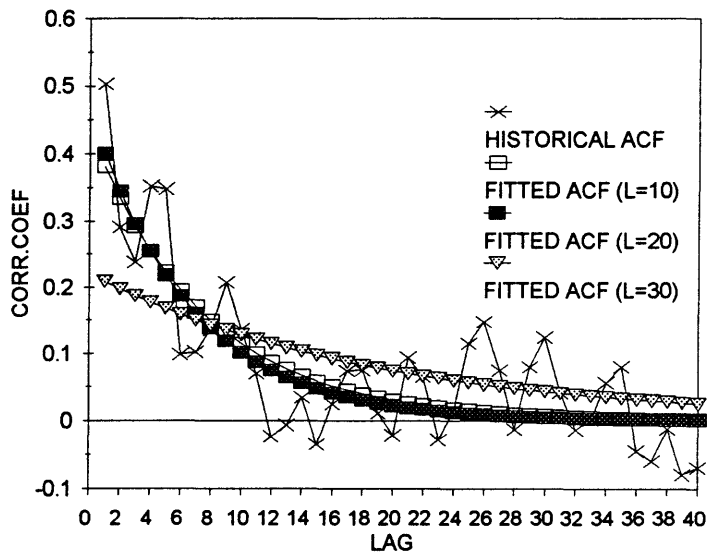

Fig 4-2   Historical and fitted autocorrelation function (ACF)
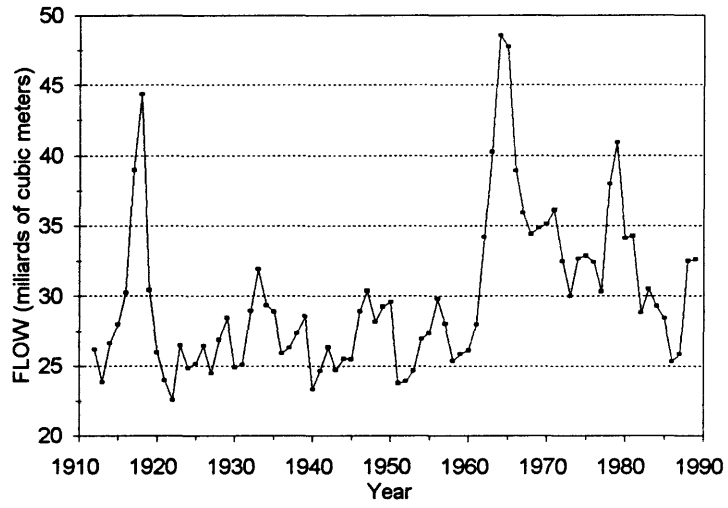   for the annual net basin supplies for lake St.Clair

137

Fig 4-3   Historical annual flows for the white Nile River
           at Malakal (1912 - 1989)
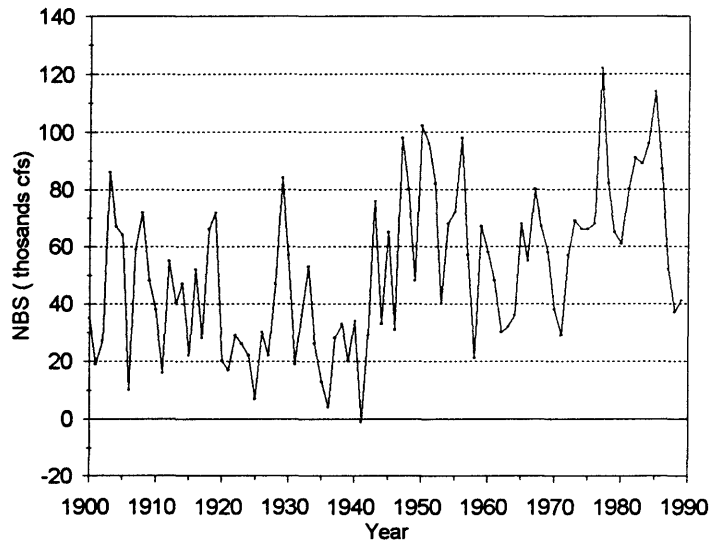

Fig 4-4   Historical annual NBS for Lake St.Clair (1900 - 1989)

138
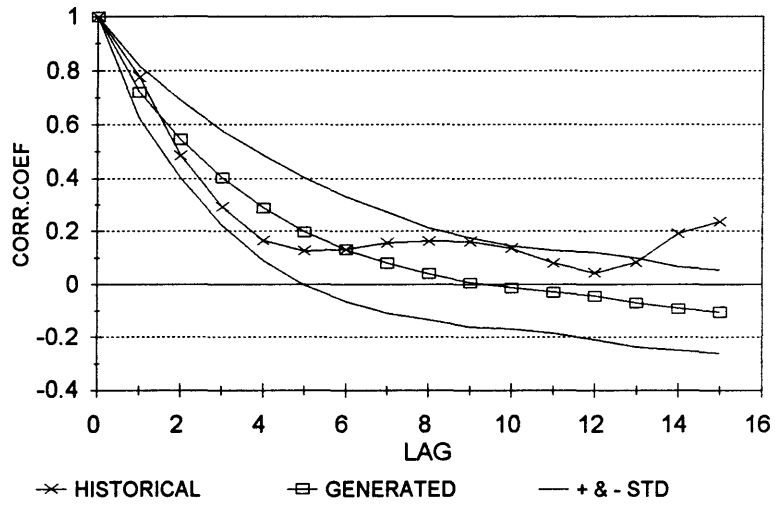
Fig 4-5   Historic and generated autocorrelogram for Malakal based on
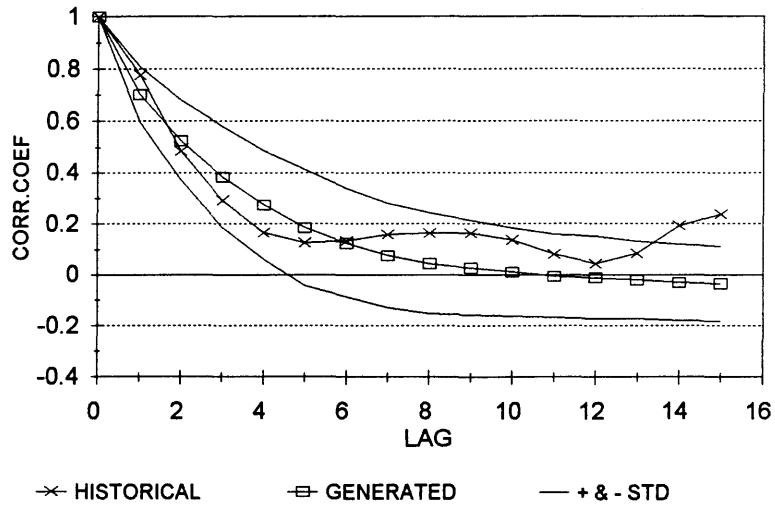   GNN model and MOM estimation method with i=1, k=8.

Fig 4-6   Historic and generated autocorrelogram for Malakal based on
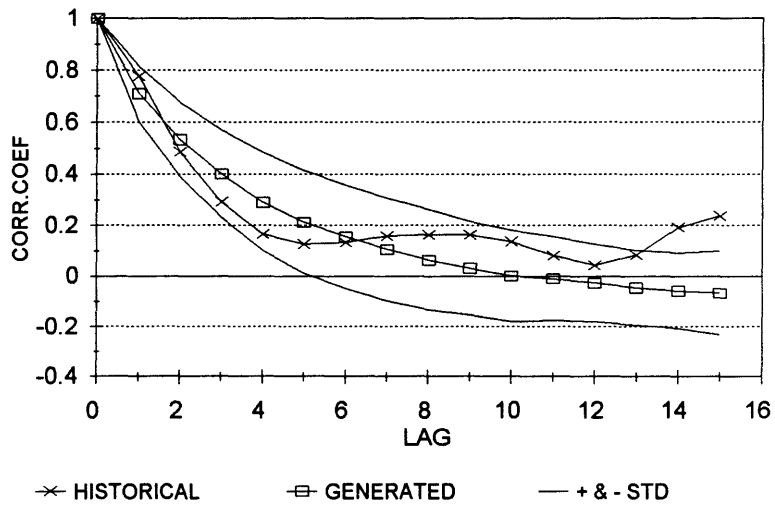   GNN model and MOM estimation method with i=1, k=9.

Fig 4-7   Historic and generated autocorrelogram for Malakal based on
           GNN model and MOM estimation method with i=1, k=10.
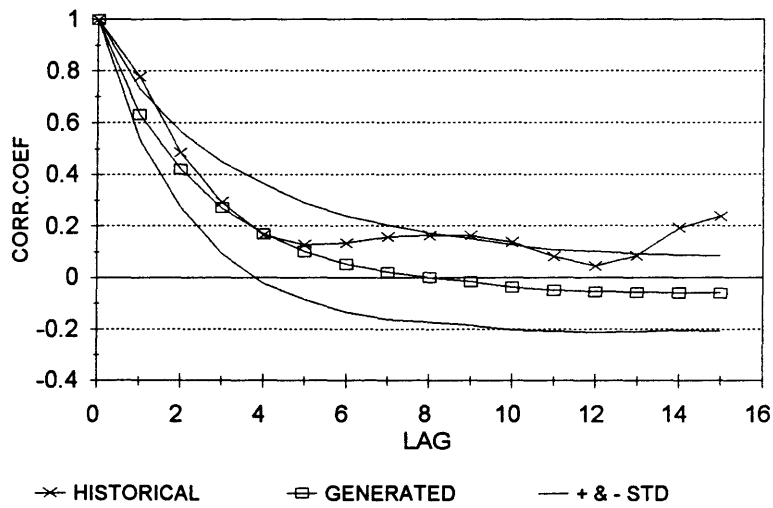

Fig 4-8   Historic and generated autocorrelogram for Malakal based on
           GNN model and MOM estimation method with i=2, k=6.
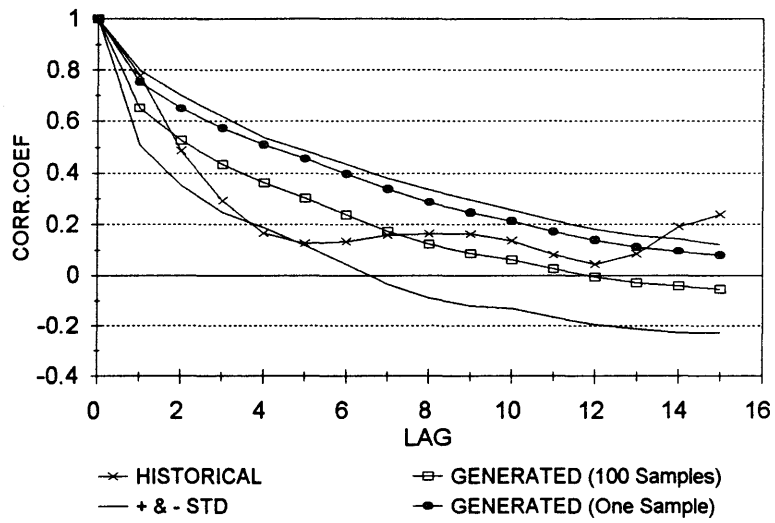
140

Fig 4-9   Historic and generated autocorrelogram for Malakal based on
           GNN model and regression analysis estimation method with
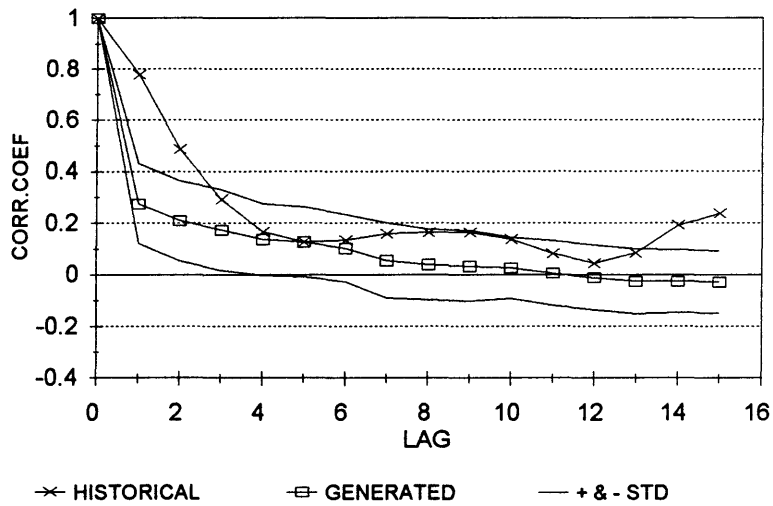   i=1, L=30.

Fig 4-10   Historic and generated autocorrelogram for Malakal based on
           GNN model and fitting the ACF estimation method with L=30.

141
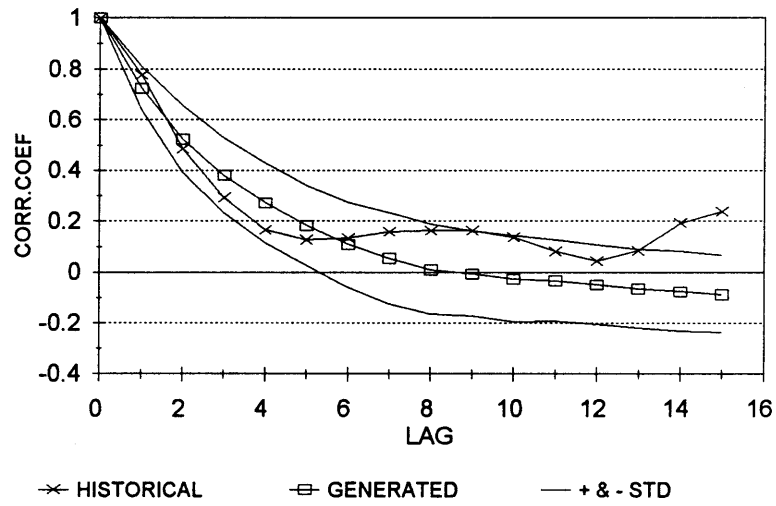
Fig 4-11 Historic and generated autocorrelogram for Malakal based on
GNN model and using the range properties estimation method with
i=1, L=30.

Fig 4-12   Historic and generated autocorrelogram for Malakal based on
AR(1) model and least squares estimation method.

Fig 4-13   Historic and generated autocorrelogram for Malakal based on
AR(2) model and least squares estimation method.

Fig 4-14   Historic and generated autocorrelogram for Malakal based on
ARMA(1,1) model and least squares estimation method.

Fig 4-15   Historic and generated autocorrelogram for Malakal based on
ARMA(2,1) model and least squares estimation method.

Fig 4-16   Historic and generated autocorrelogram for Malakal based on
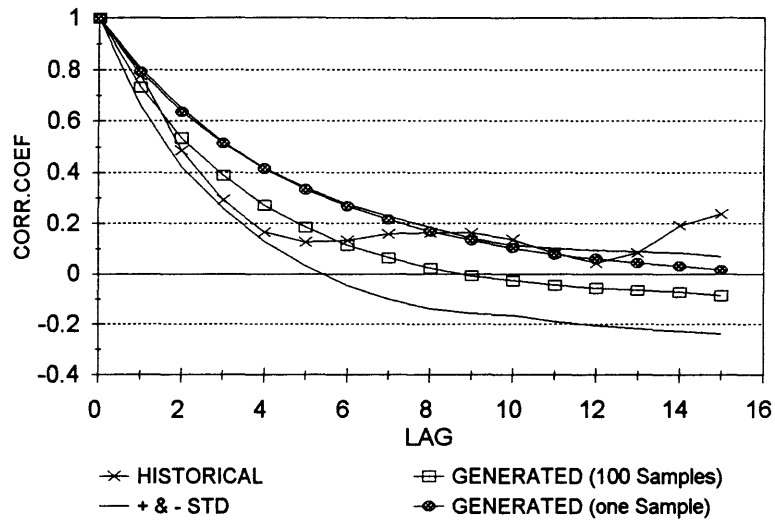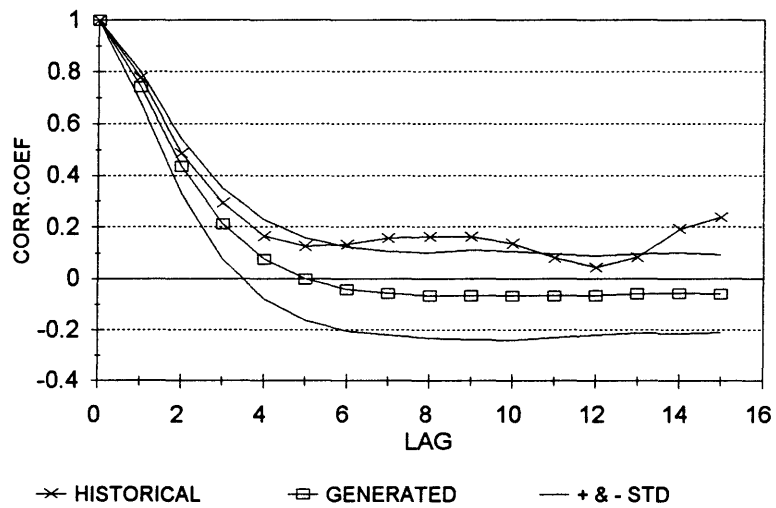ARMA(2,2) model and least squares estimation method.
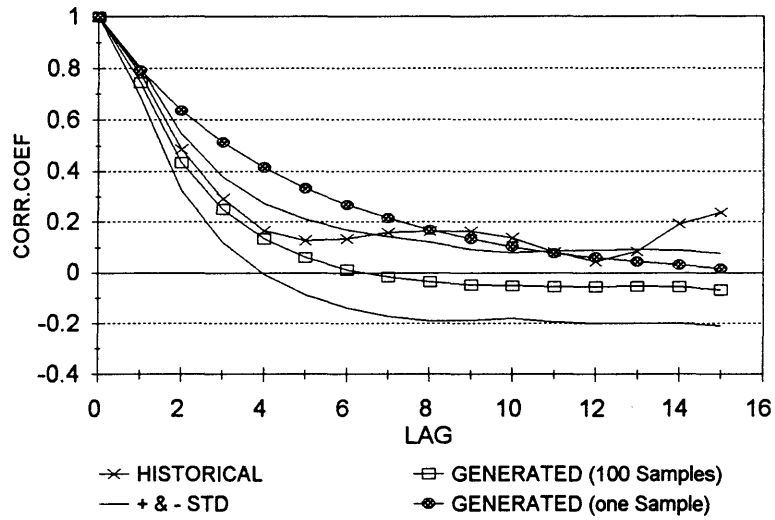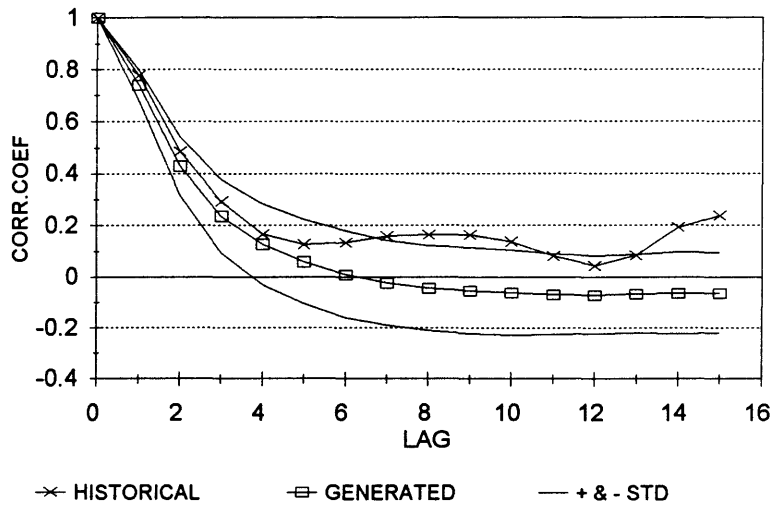

Fig 4-17   Historic and generated autocorrelogram for Malakal based on
AR(3) model and least squares estimation method.

145
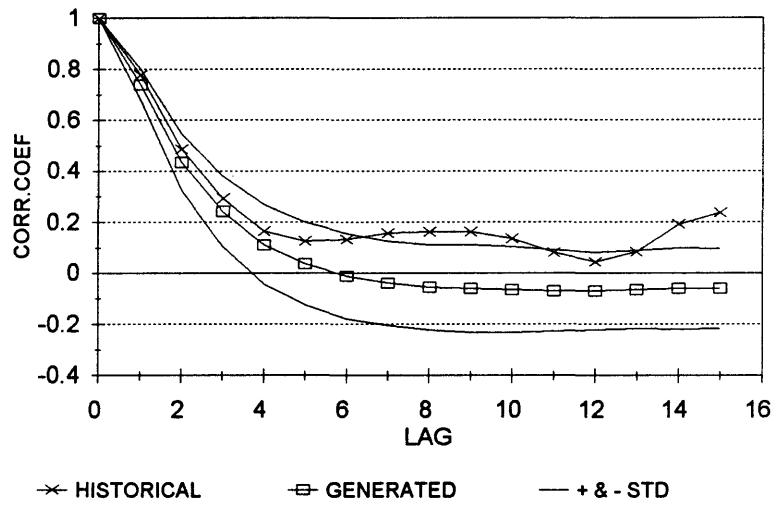
Fig 4-18   Historic and generated autocorrelogram for Malakal based on
ARMA(3,1) model and method of moments estimation method.
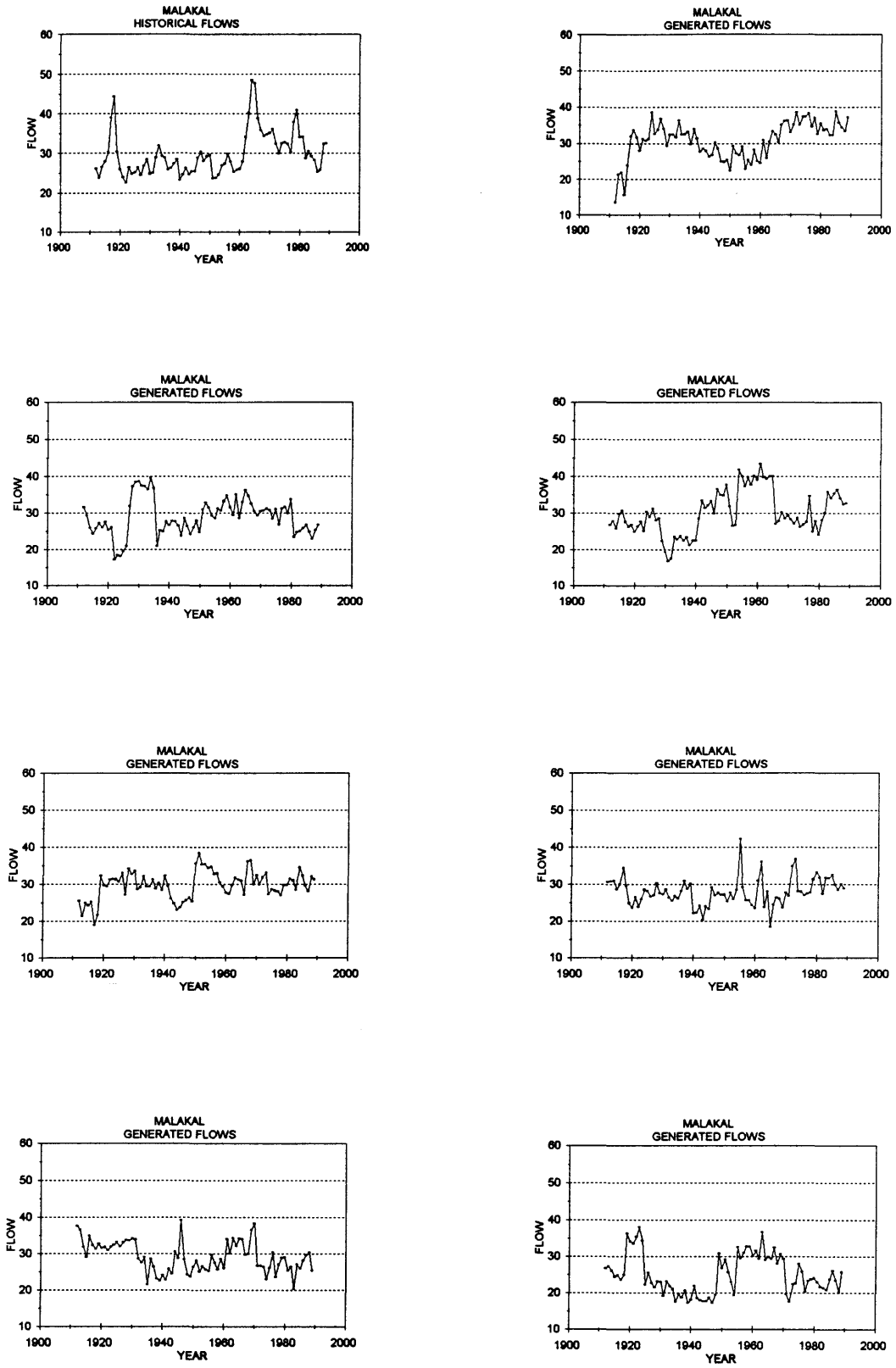
Fig 4-19   Historical and Generated Flows for Malakal based on GNN model
and regression analysis estimation method.

147
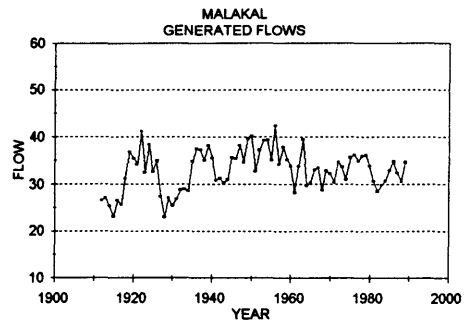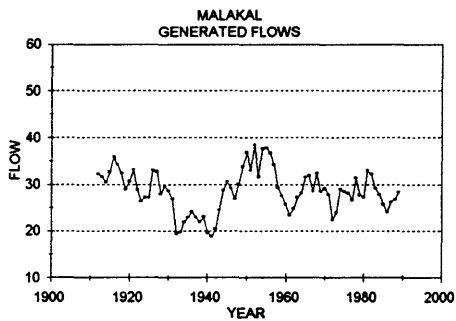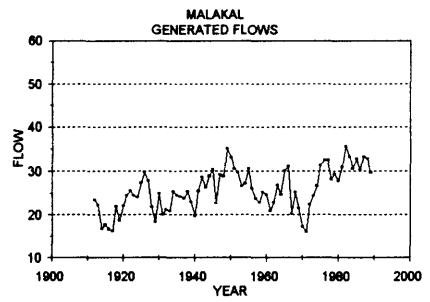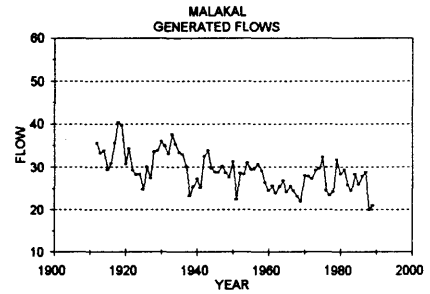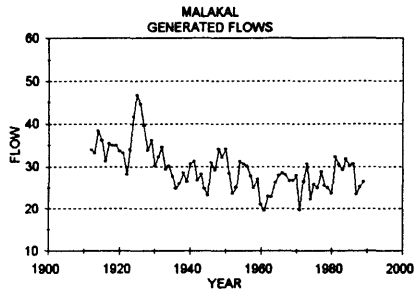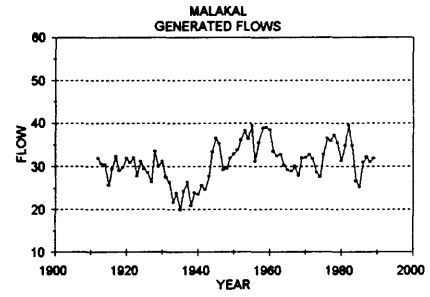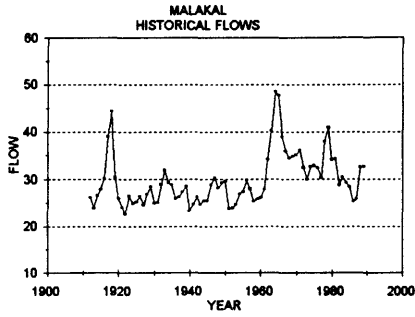
Fig 4-20 Historical and Generated Flows for Malakal based on AR(1) model
and least squares estimation method.

148
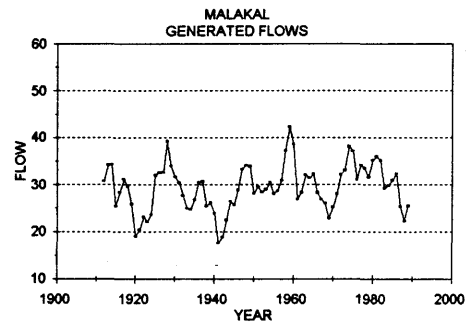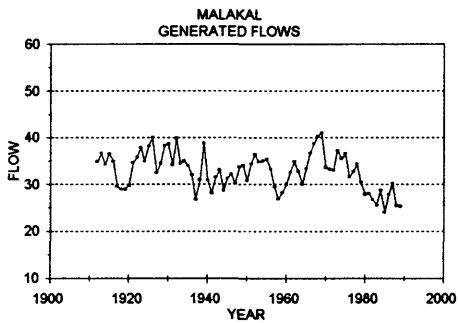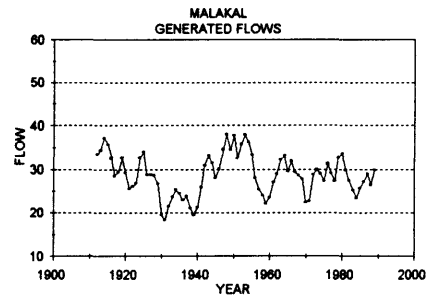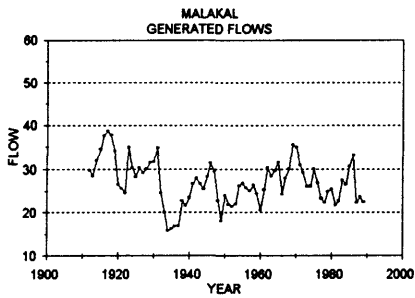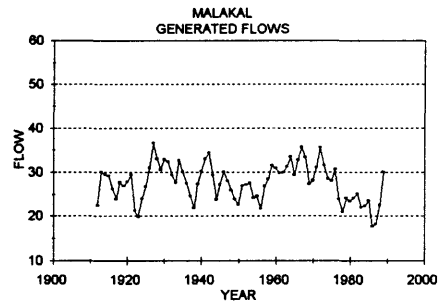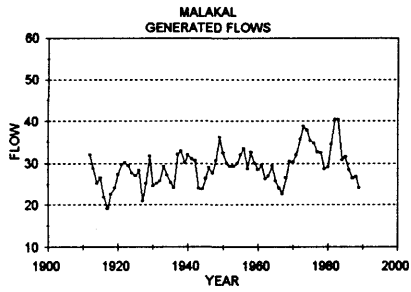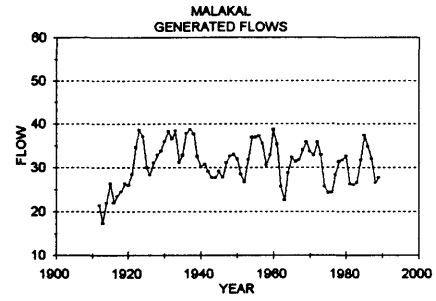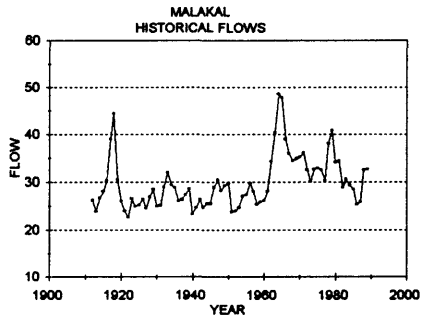
Fig 4-21   Historical and Generated Flows for Malakal based on ARMA(1,1) model
and least squares estimation method.
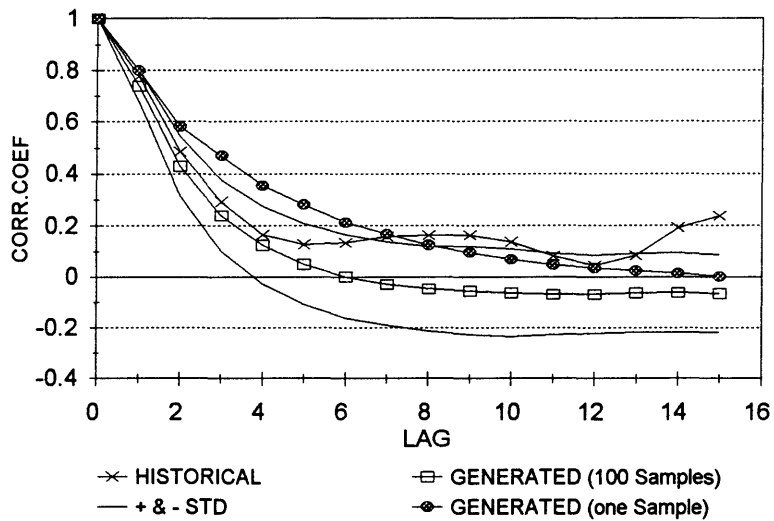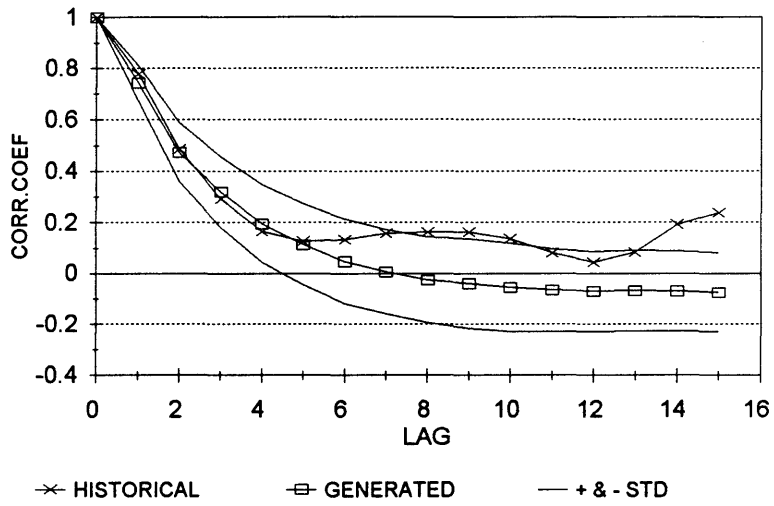
149
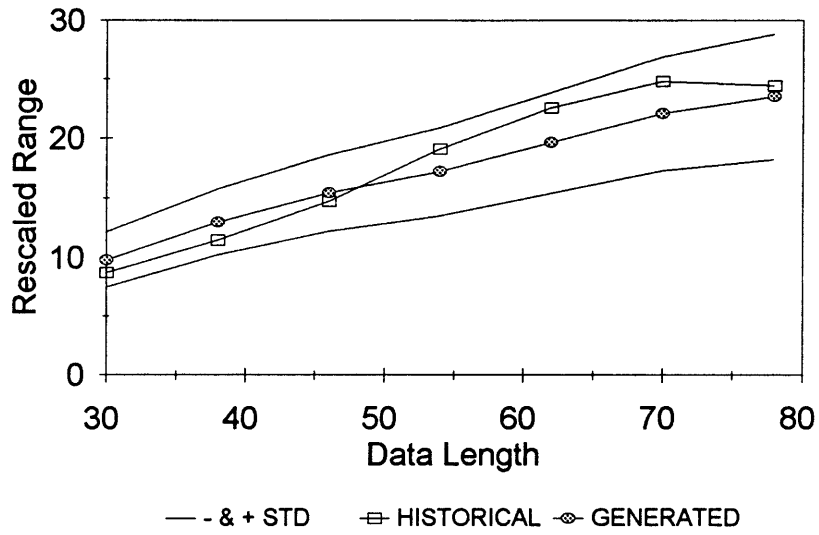
Fig 4-22   Historic and generated rescaled range for different data lengths for
           Malakal based on GNN Model and regression analysis estimation
           method with i=1, L=30.

Fig 4-23   Historic and generated rescaled range for different data lengths for
           Malakal based on AR(1) Model and least squares estimation method.

150
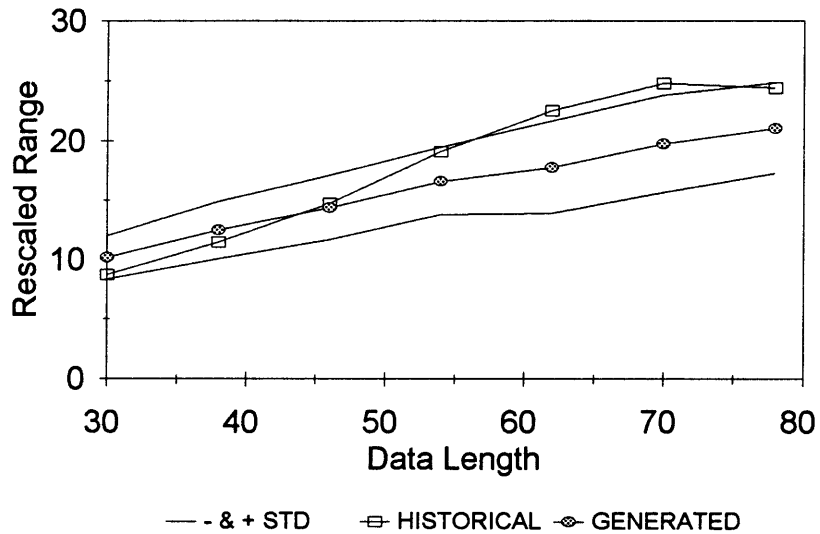
— - & + STD   -□- HISTORICAL -⊗- GENERATED

Fig 4-24   Historic and generated rescaled range for different data lengths for
Malakal based on ARMA(1,1) Model and least squares estimation
method.
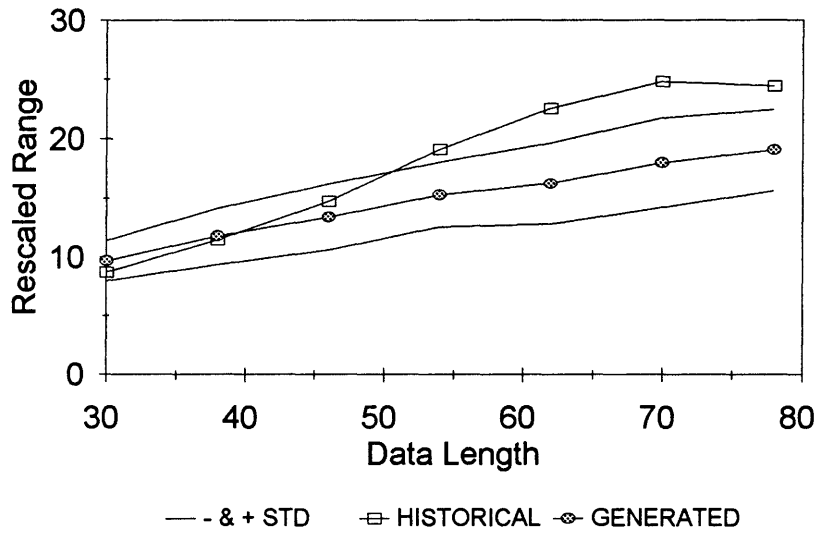
— - & + STD   -□- HISTORICAL -⊗- GENERATED
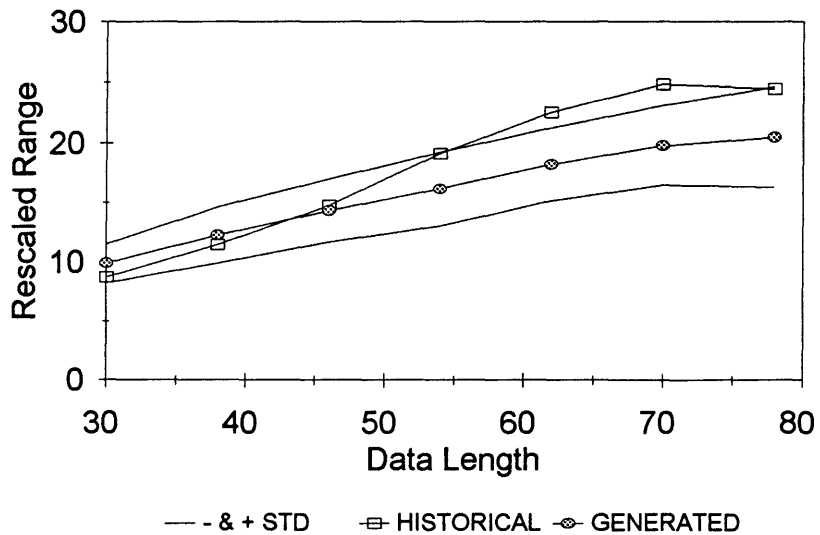
Fig 4-25   Historic and generated rescaled range for different data lengths for
Malakal based on ARMA(2,1) Model and least squares estimation
method.

151
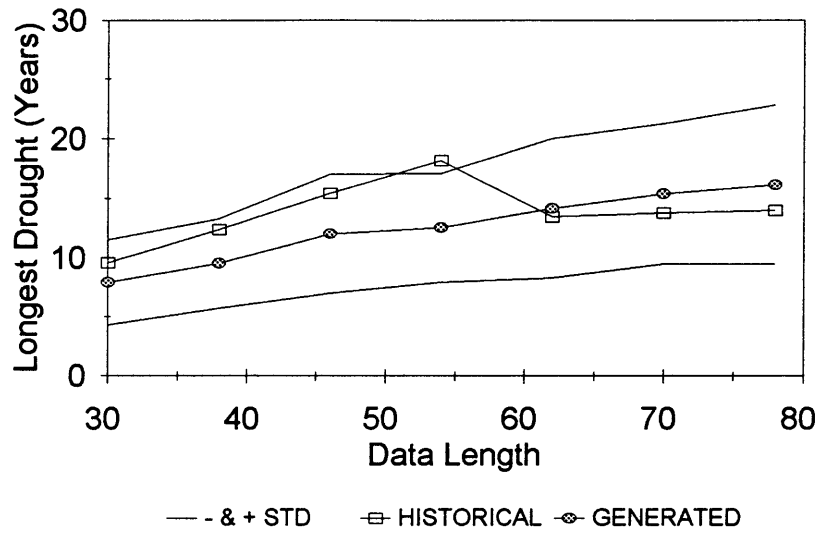
Fig 4-26   Historic and generated longest drought for different data lengths for
            Malakal based on GNN Model and regression analysis estimation
            method with i=1, L=30.

Fig 4-27   Historic and generated longest drought for different data lengths for
            Malakal based on AR(1) Model and least squares estimation method.

Fig 4-28   Historic and generated longest drought for different data lengths for
            Malakal based on ARMA(1,1) Model and least squares estimation method

Fig 4-29   Historic and generated longest drought for different data lengths for
            Malakal based on ARMA(2,1) Model and least squares estimation method

Fig 4-30   Historic and generated maximum deficit for different data lengths for
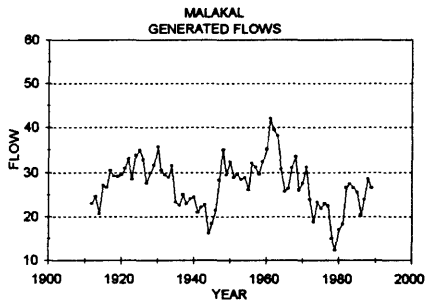   Malakal based on GNN Model and regression analysis estimation
            method with i=1, L=30.

Fig 4-31   Historic and generated maximum deficit for different data lengths for
            Malakal based on AR(1) Model and least squares estimation method.

Fig 4-32　Historic and generated maximum deficit for different data lengths for
　　　　　Malakal based on ARMA(1,1) Model and least squares estimation method

Fig 4-33　Historic and generated maximum deficit for different data lengths for
　　　　　Malakal based on ARMA(2,1) Model and least squares estimation method

155

Fig 4-34   Historic and generated surplus for different data lengths for
            Malakal based on GNN Model and regression analysis estimation
            method with i=1, L=30.

Fig 4-35   Historic and generated surplus for different data lengths for
            Malakal based on AR(1) Model and least squares estimation method.

156

Fig 4-36   Historic and generated surplus for different data lengths for Malakal
based on ARMA(1,1) Model and least squares estimation method.

Fig 4-37   Historic and generated surplus for different data lengths for Malakal
based on ARMA(2,1) Model and least squares estimation method.

157

Fig 4-38   Historic and generated autocorrelogram for Lake St.Clair based on
the GNN model and MOM estimation method with i=1,k=2.

Fig 4-39   Historic and generated autocorrelogram for Lake St.Clair based on
the GNN model and MOM estimation method with i=1,k=4.

Fig 4-40 Historic and generated autocorrelogram for Lake St.Clair based on
the GNN model and MOM estimation method with i=2,k=3.
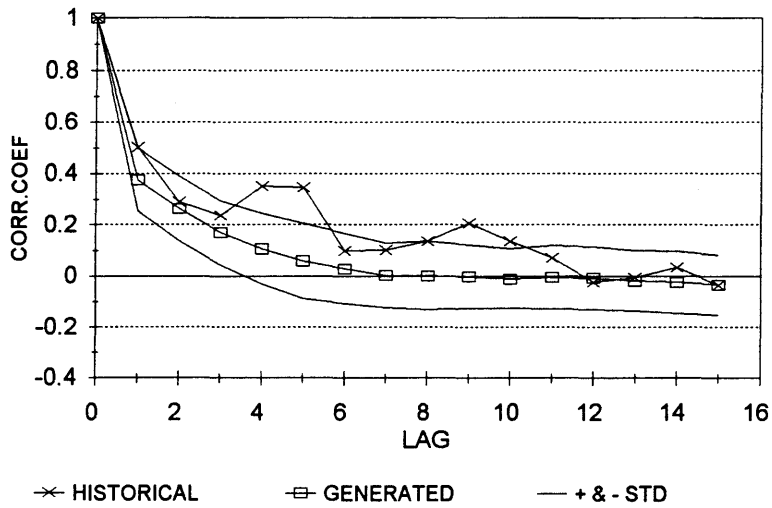
Fig 4-41 Historic and generated autocorrelogram for Lake St.Clair based on
the GNN model and MOM estimation method with i=3,k=6.

159
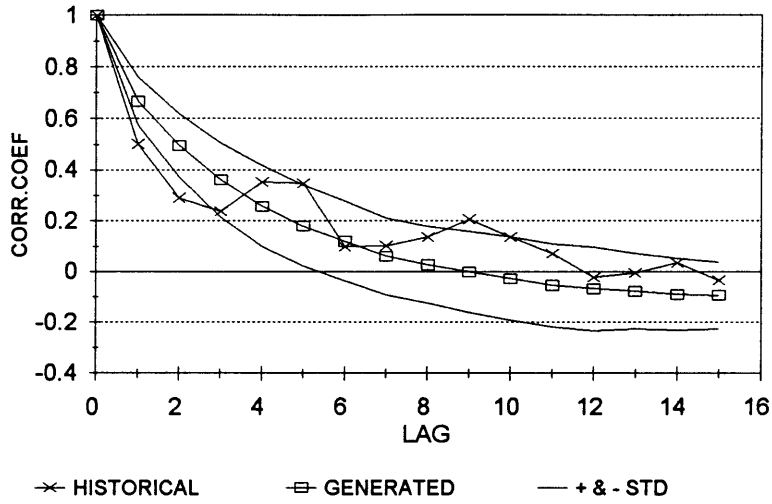
Fig 4-42　Historic and generated autocorrelogram for Lake St.Clair based on
the GNN model and MOM estimation method with i=4,k=8.
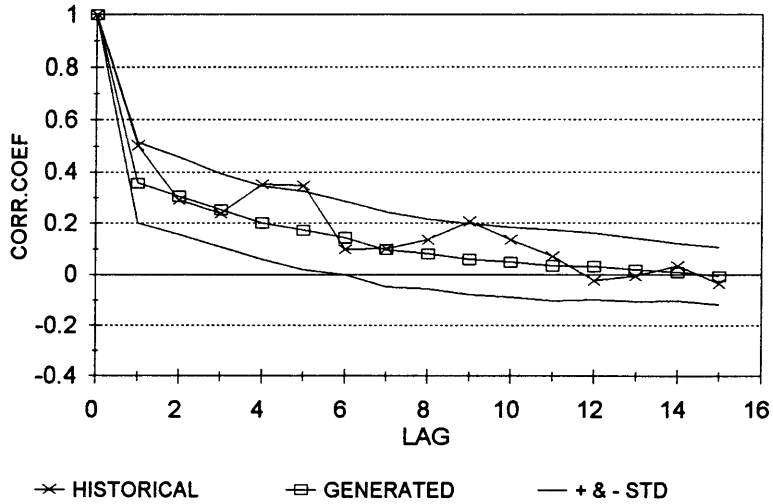
Fig 4-43　Historic and generated autocorrelogram for Lake St.Clair based on
the GNN model and MOM estimation method with i=4,k=9.

Fig 4-44   Historic and generated autocorrelogram for Lake St.Clair based on
the GNN model and regression analysis estimation method with
i=1, L=30.
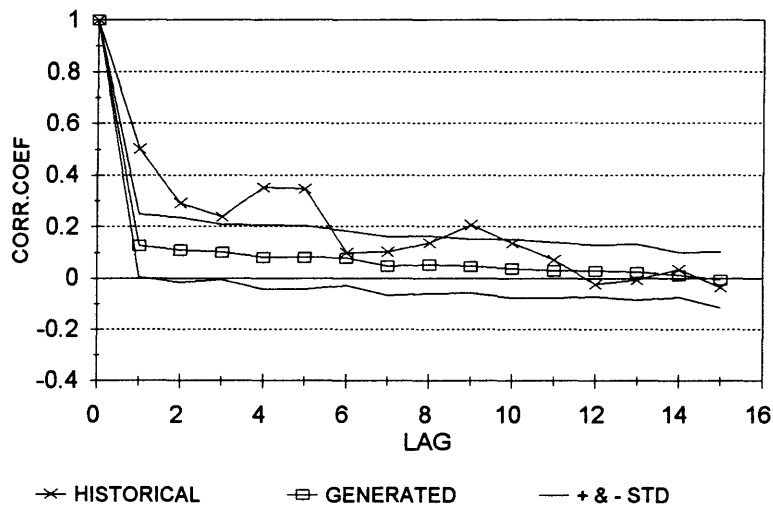
Fig 4-45   Historic and generated autocorrelogram for Lake St.Clair based on
the GNN model and fitting the ACF estimation method with L=30.
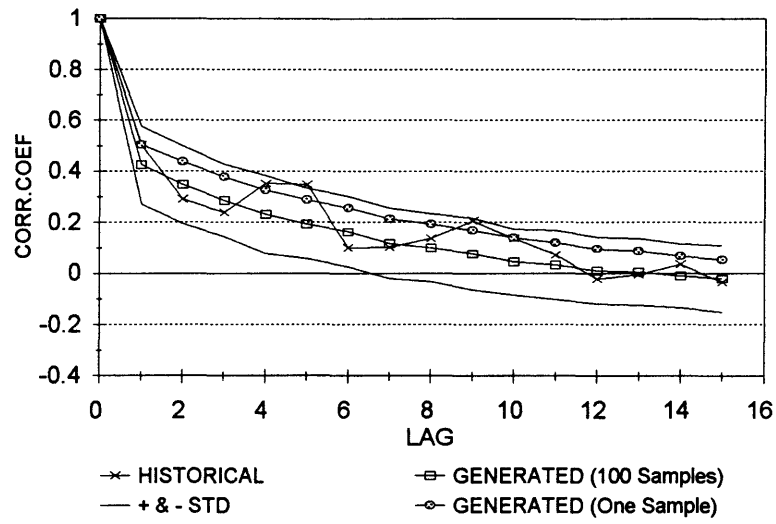
Fig 4-46   Historic and generated autocorrelogram for Lake St.Clair based on
the GNN model and using range properties estimation method with
i=1, L=30.
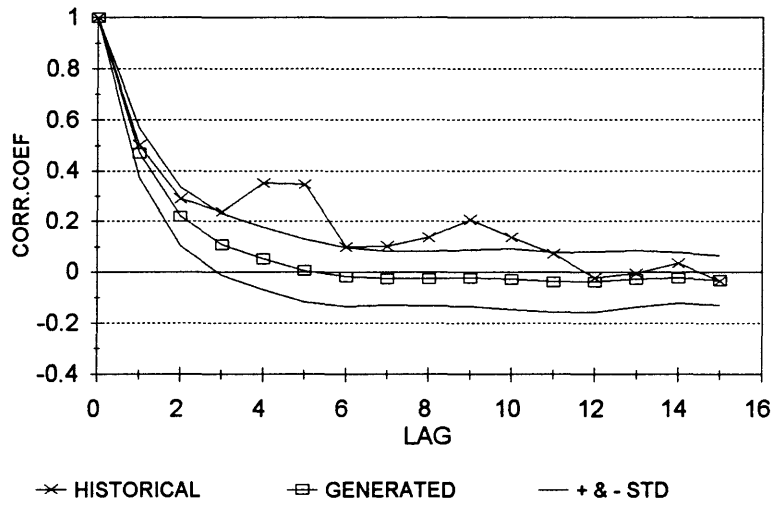
Fig 4-47   Historic and generated autocorrelogram for Lake St.Clair based on
the AR(1) model and MOM estimation method.
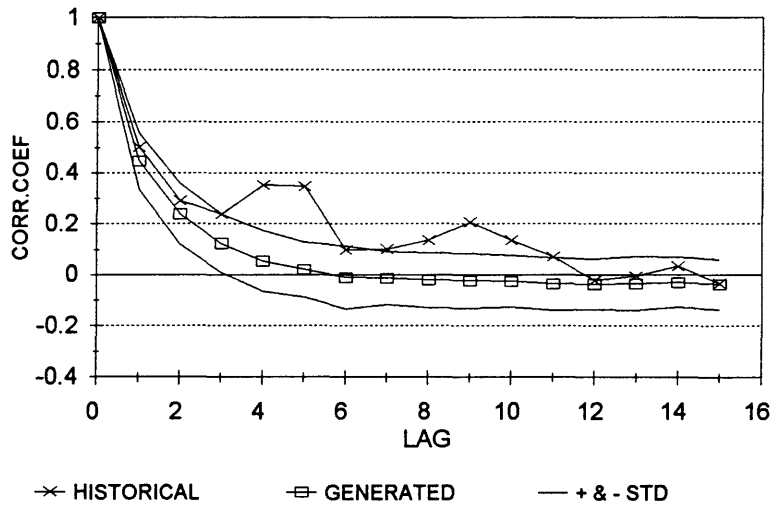
Fig 4-48   Historic and generated autocorrelogram for Lake St.Clair based on
the AR(2) model and MOM estimation method.

163
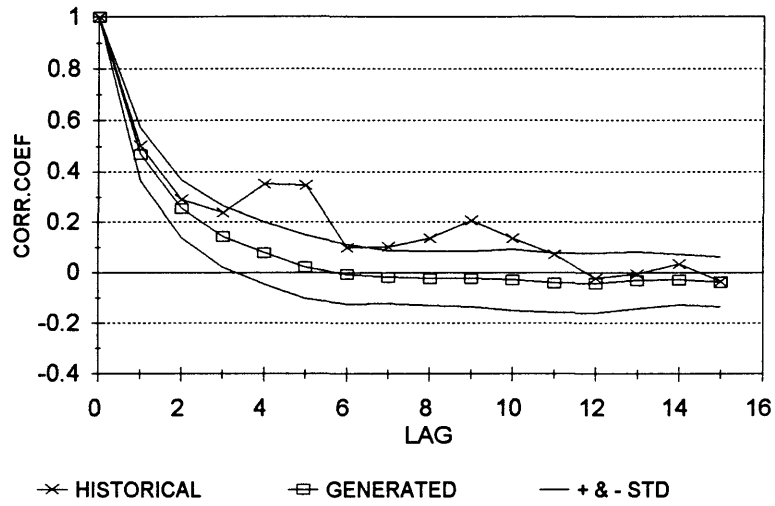
Fig 4-49   Historic and generated autocorrelogram for Lake St.Clair based on
the ARMA(1,1) model and MOM estimation method.
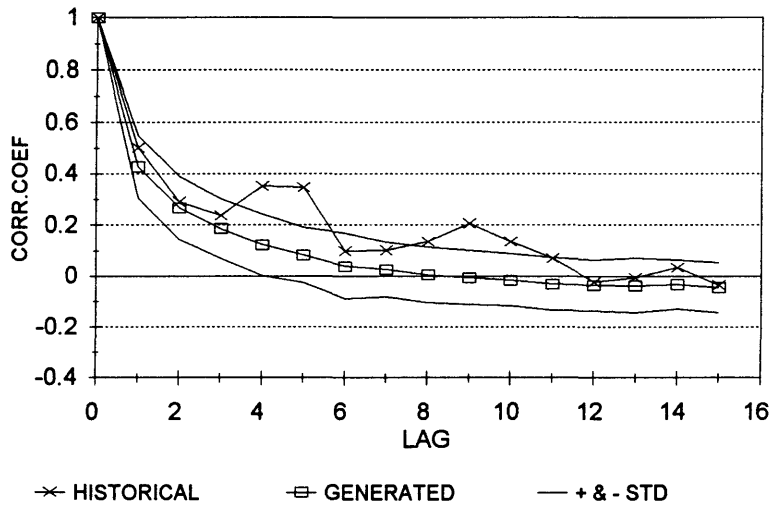
Fig 4-50   Historic and generated autocorrelogram for Lake St.Clair based on
the ARMA(2,1) model and least squares estimation method.

164
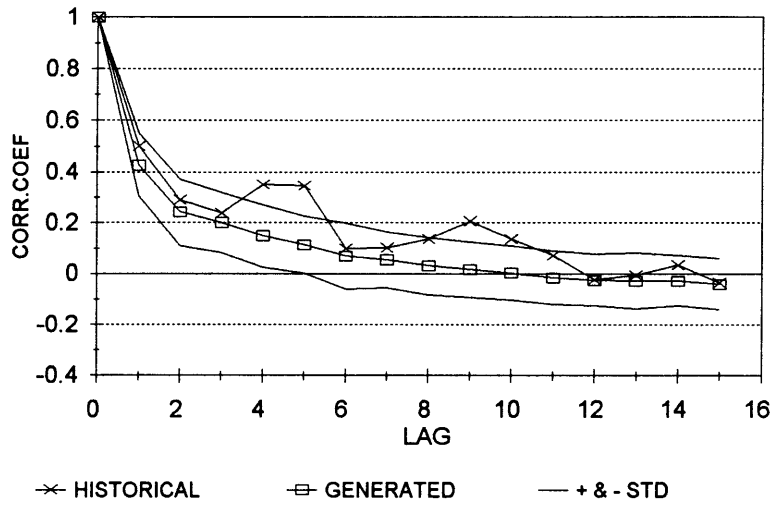
Fig 4-51   Historic and generated autocorrelogram for Lake St.Clair based on
the ARMA(2,2) model and least squares estimation method.
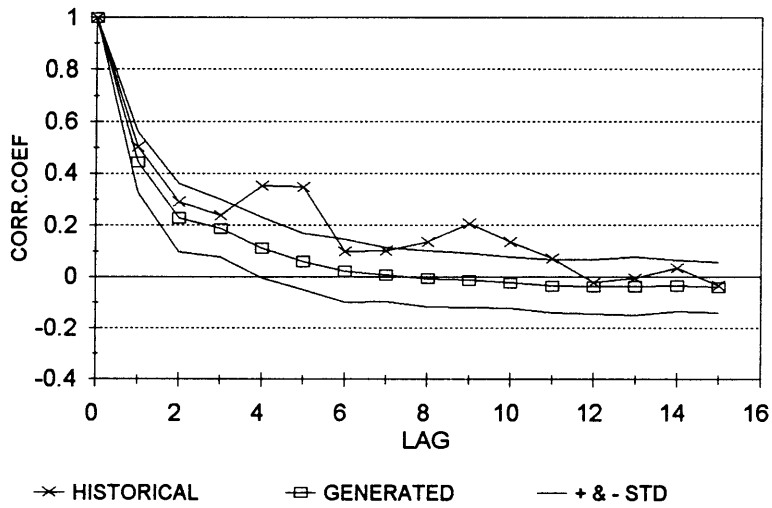
Fig 4-52   Historic and generated autocorrelogram for Lake St.Clair based on
the AR(3) model and MOM estimation method.
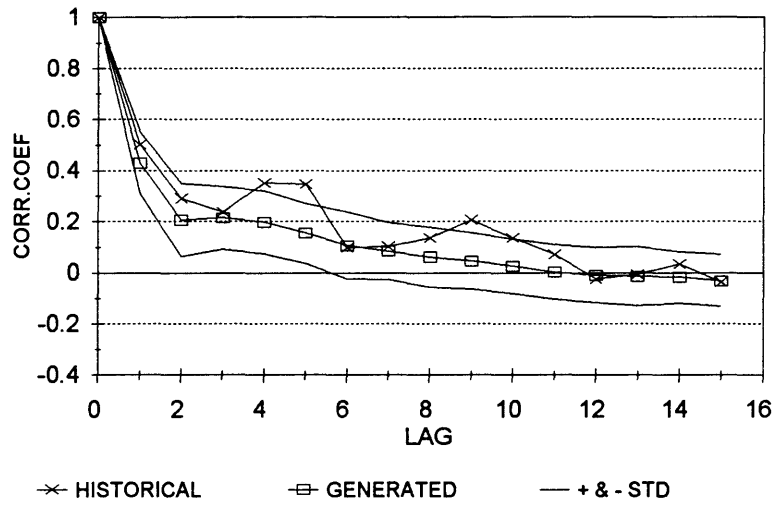
Fig 4-53   Historic and generated autocorrelogram for Lake St.Clair based on
the ARMA(3,1) model and MOM estimation method.
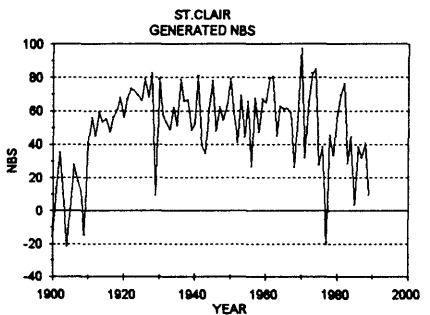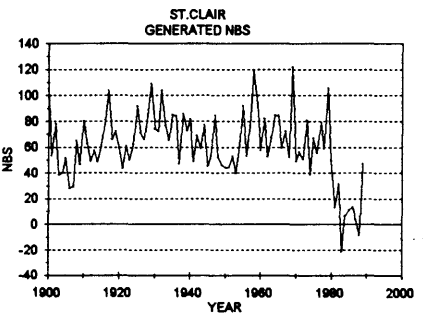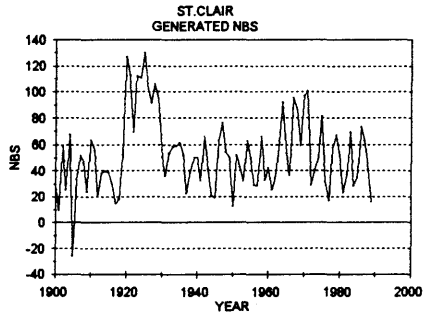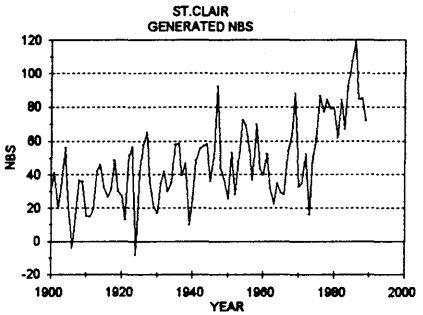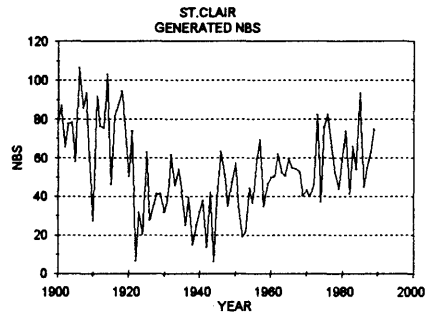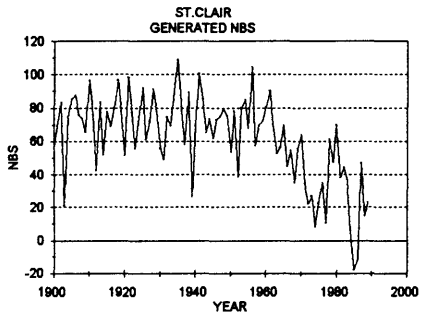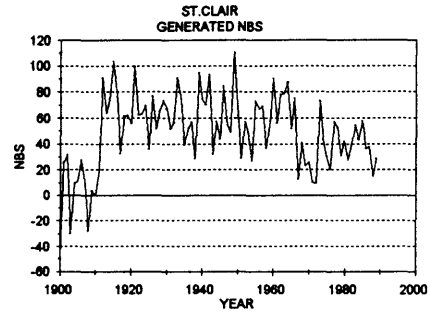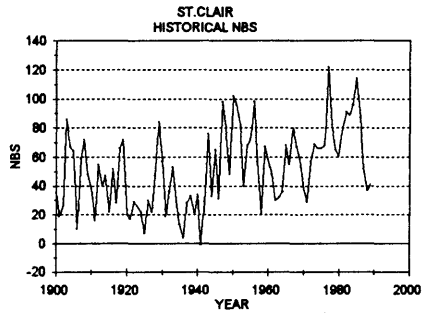
Fig 4-54   Historical and Generated NBS for lake St.Clair based on GNN model
and regression analysis estimation method.

167
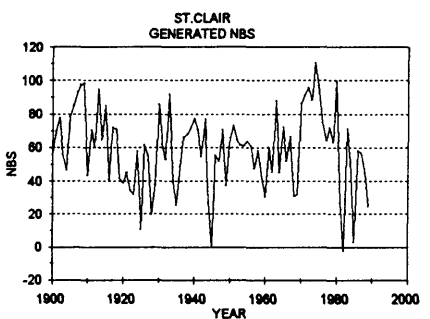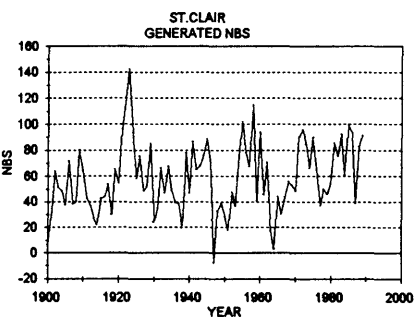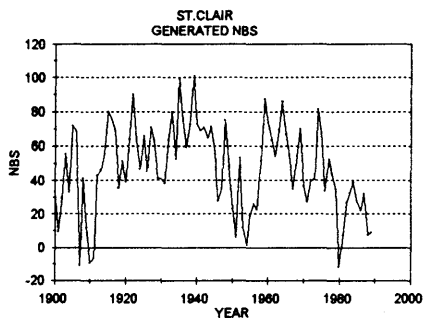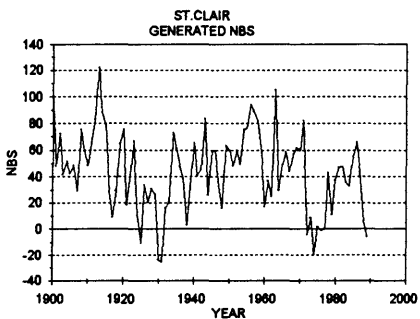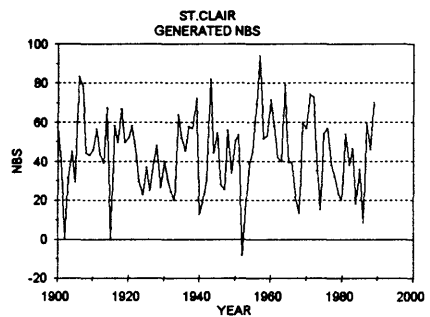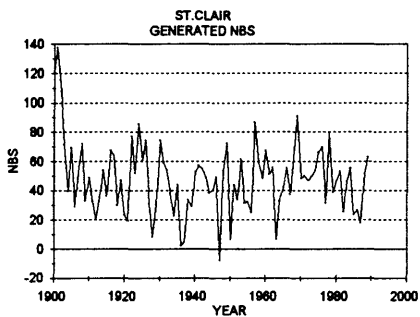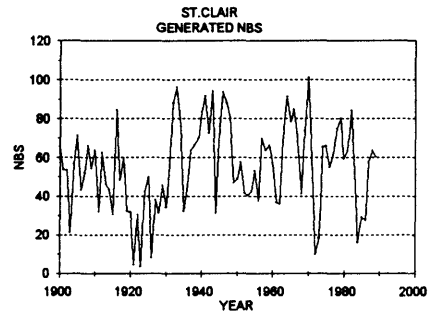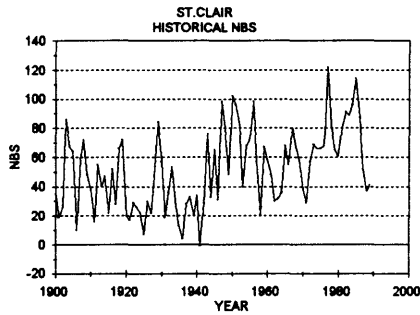
Fig 4-55   Historical and Generated NBS for lake St.Clair based on ARMA(1,1) model
and MOM estimation method.

168
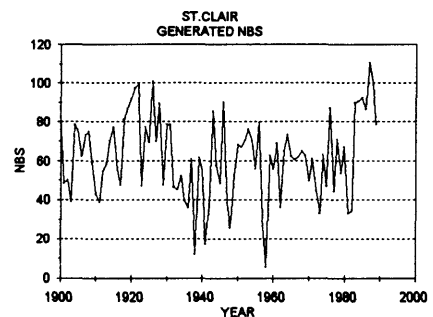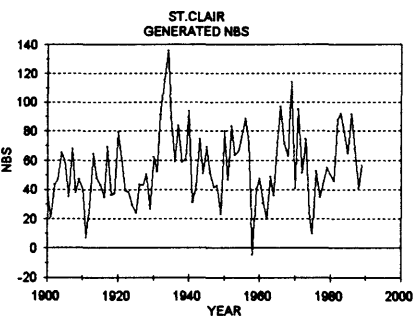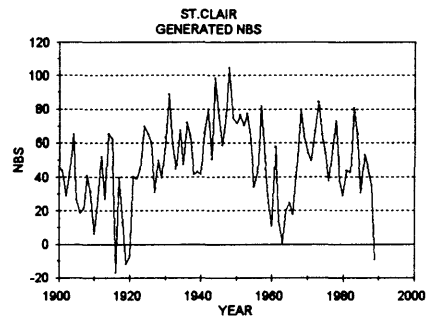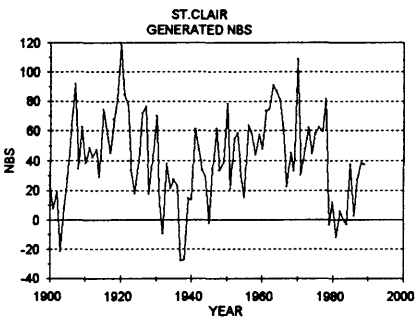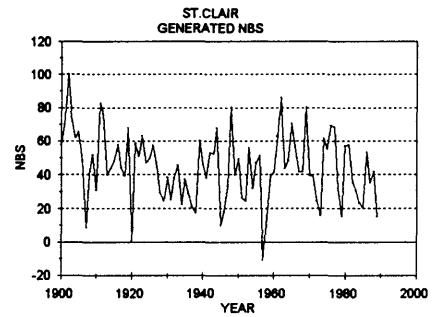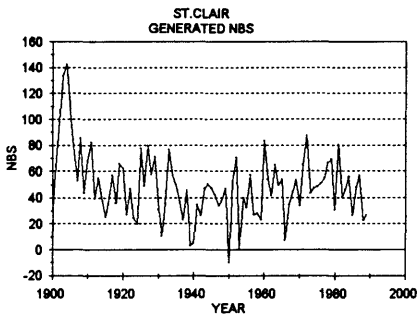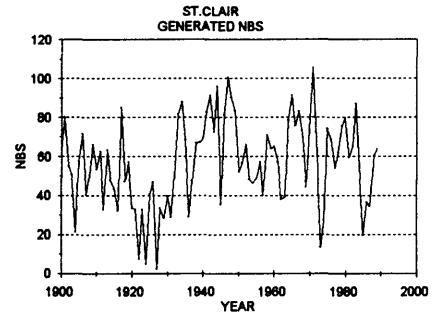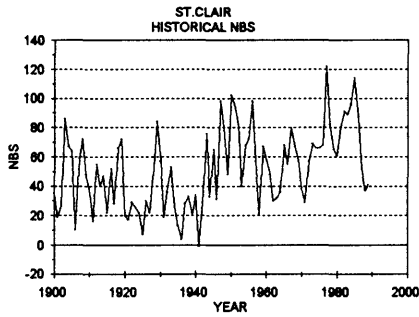
Fig 4-56   Historical and Generated NBS for lake St.Clair based on ARMA(2,2) model
and least squares estimation method.

169
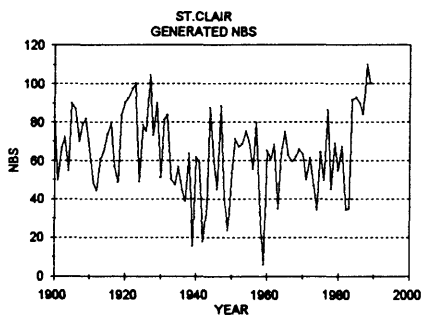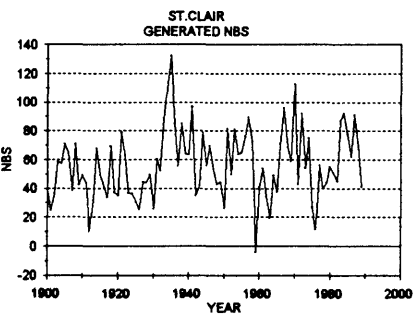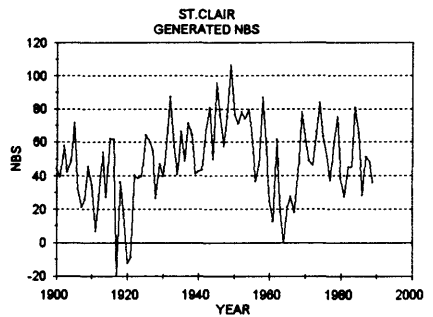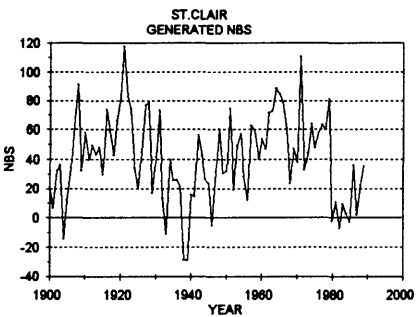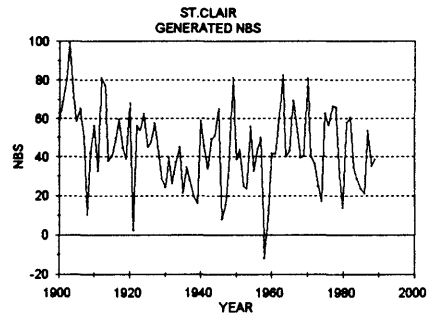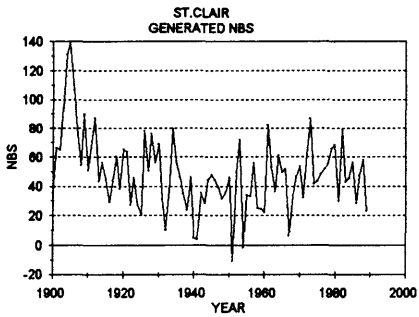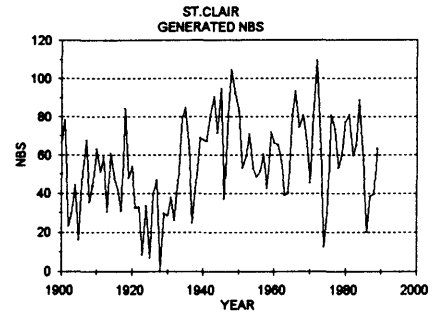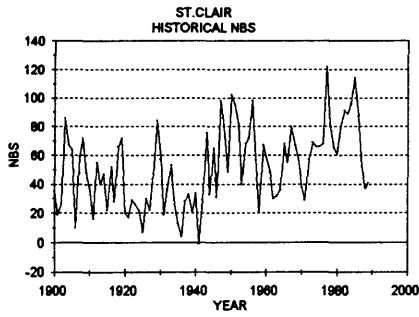
Fig 4-57 Historical and Generated NBS for lake St.Clair based on ARMA(3,1) model and MOM estimation method.

170
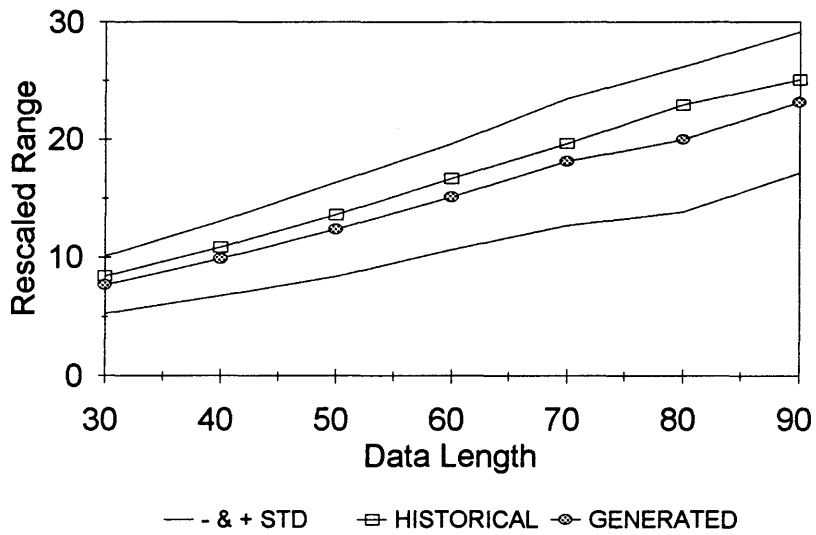
Fig 4-58   Historic and generated rescaled range for different data lengths for
lake St.Clair based on GNN Model and regression analysis estimation
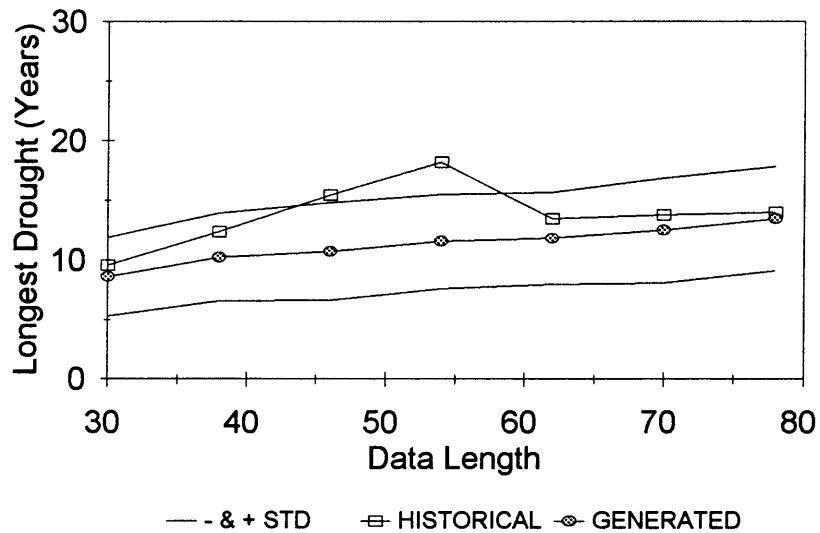method with i=1, L=30.
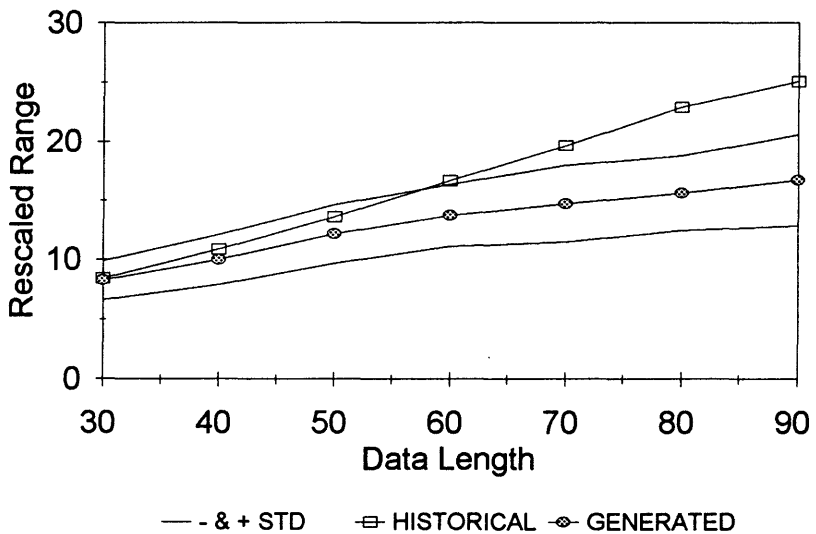
Fig 4-59   Historic and generated rescaled range for different data lengths for
lake St.Clair based on AR(1) Model and MOM estimation method.
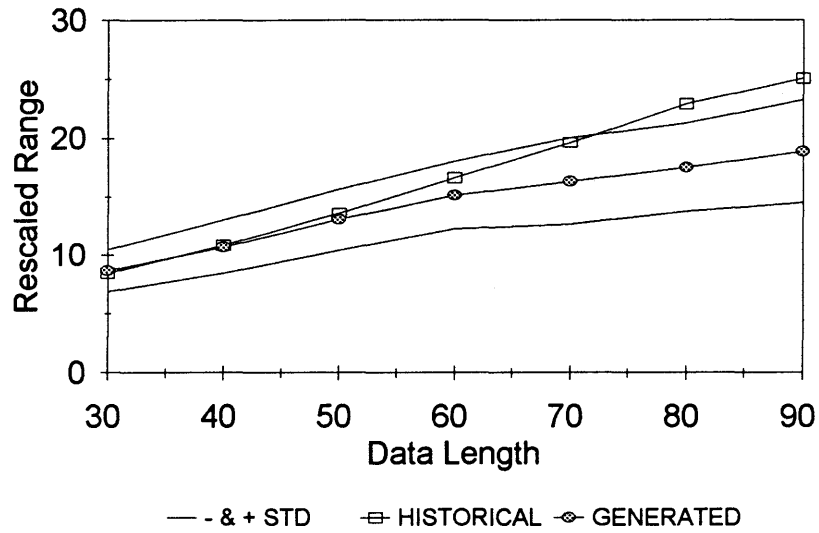
Fig 4-60   Historic and generated rescaled range for different data lengths for
            lake St.Clair based on ARMA(1,1) Model and MOM estimation method.
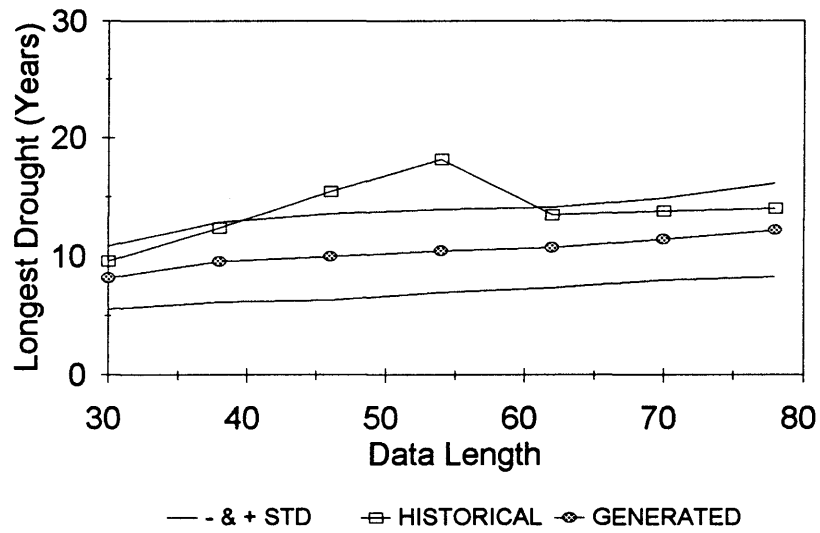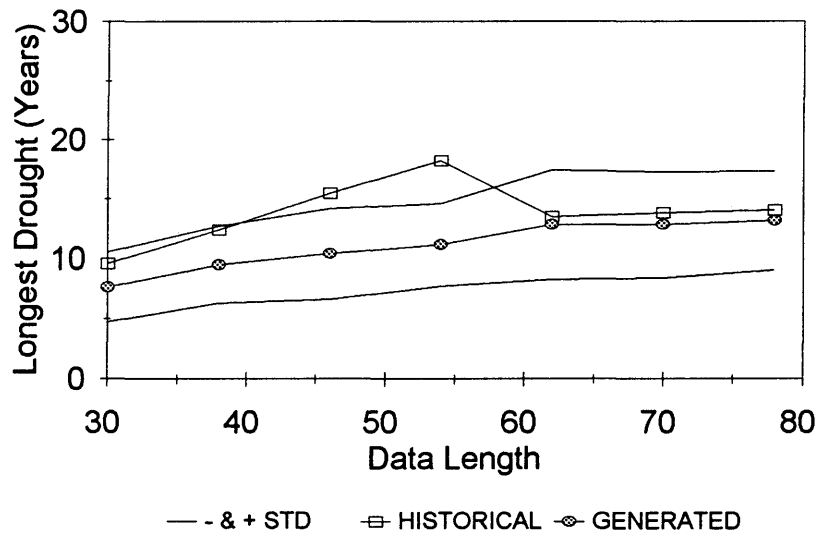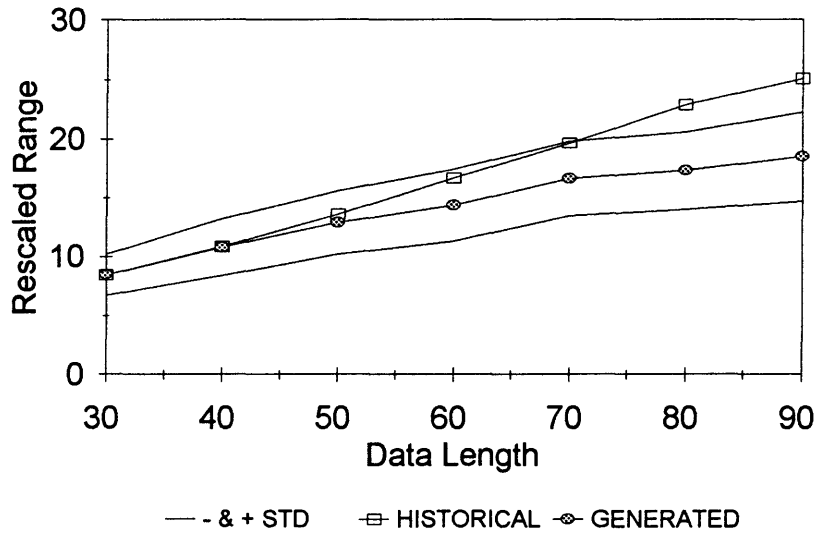

Fig 4-61   Historic and generated rescaled range for different data lengths for
            lake St.Clair based on ARMA(2,1) Model and least squares estimation
            method.

172
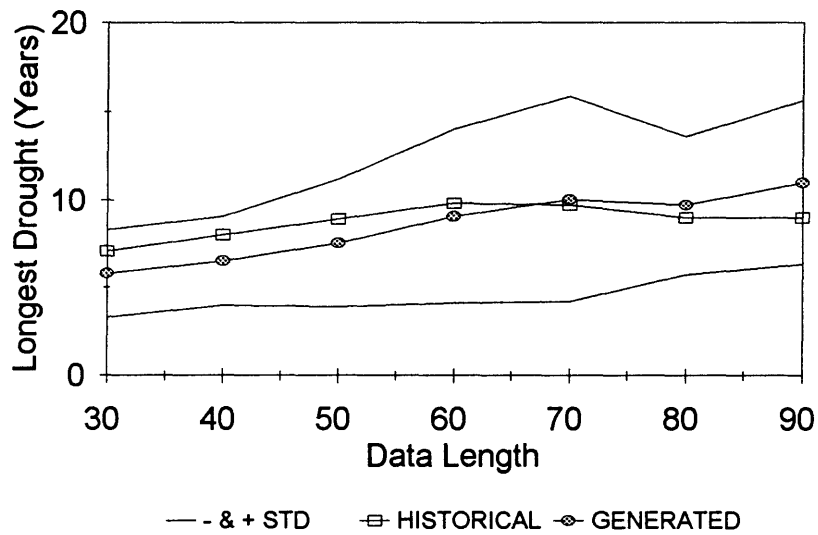
Fig 4-62   Historic and generated longest drought for different data lengths for
           lake St.Clair based on GNN Model and regression analysis estimation
           method with i=1, L=30.
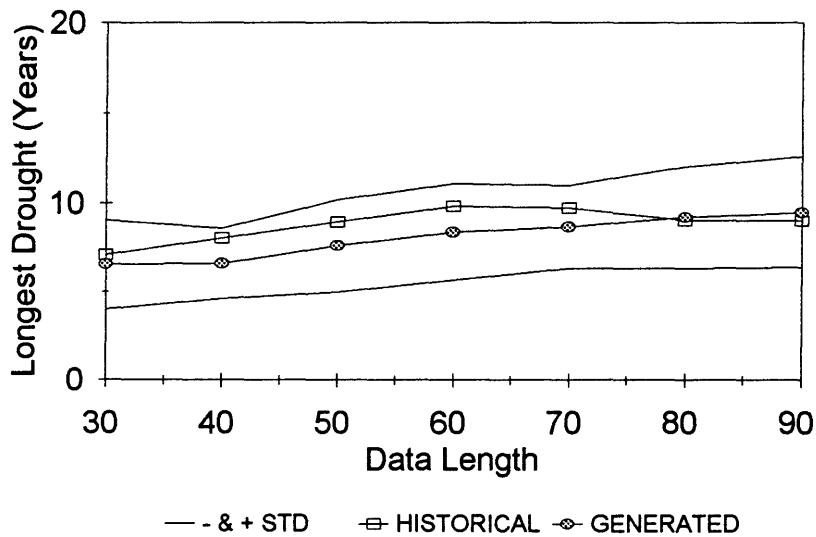
Fig 4-63   Historic and generated longest drought for different data lengths for
           lake St.Clair based on AR(1) Model and MOM estimation method.

173
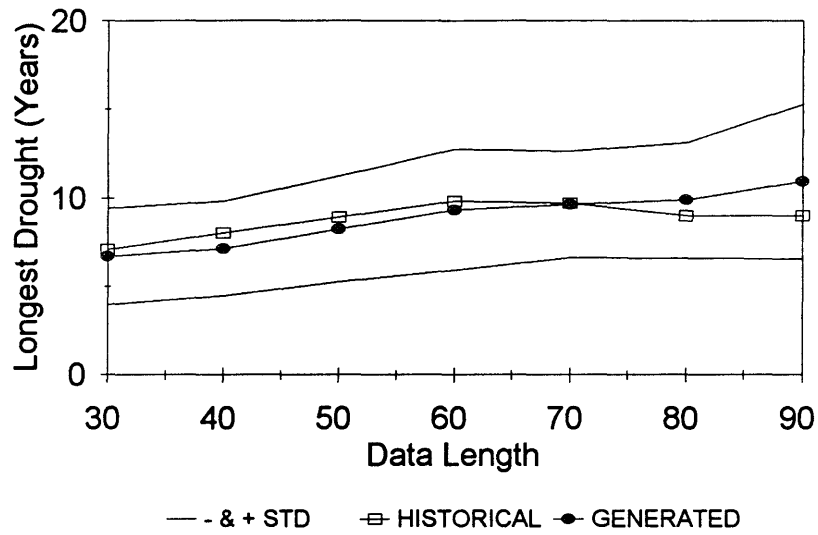
Fig 4-64   Historic and generated longest drought for different data lengths for
          lake St.Clair based on ARMA(1,1) Model and MOM estimation method.
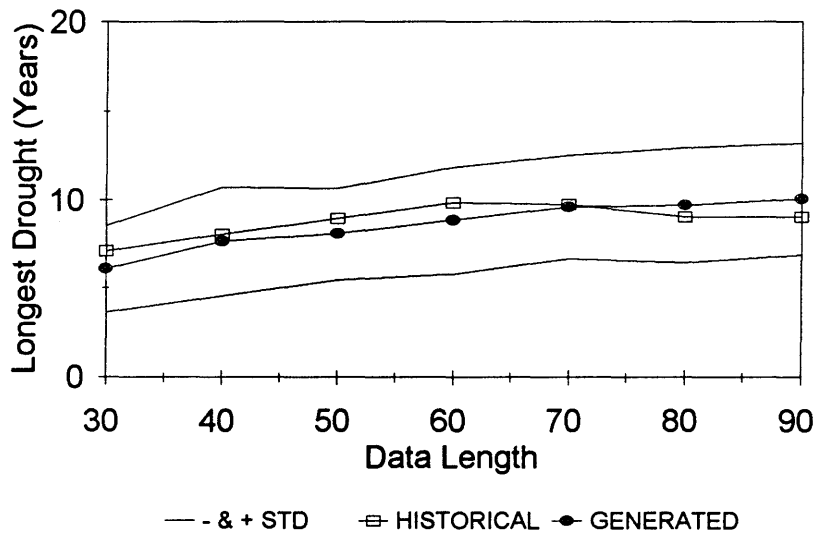
Fig 4-65   Historic and generated longest drought for different data lengths for
          lake St.Clair based on ARMA(2,1) Model and least squares estimation
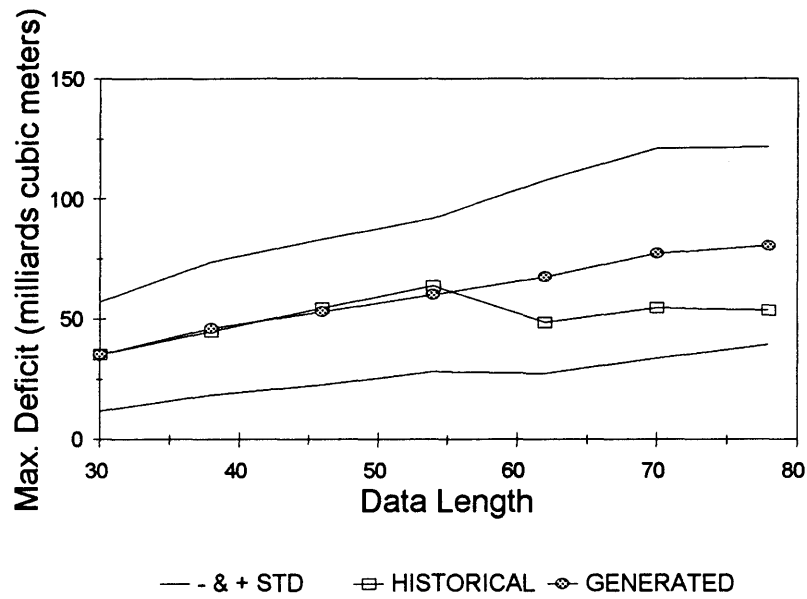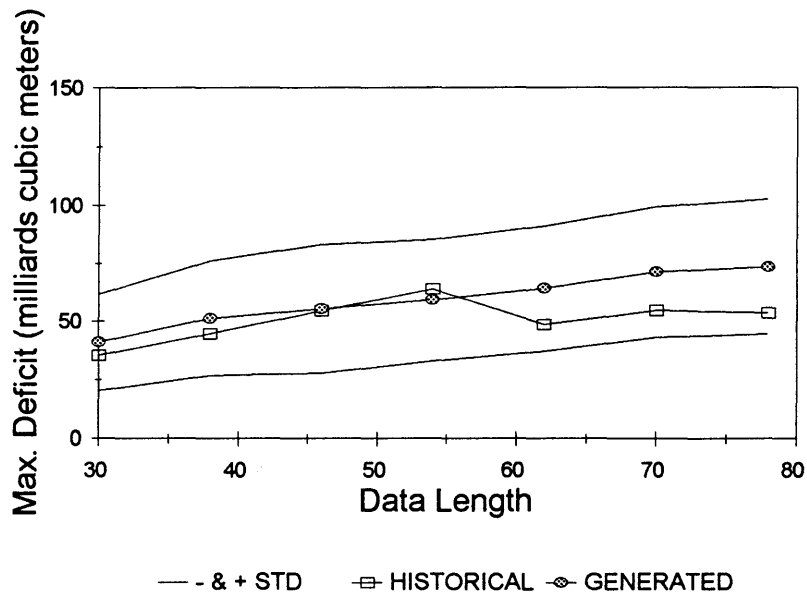   method.

174
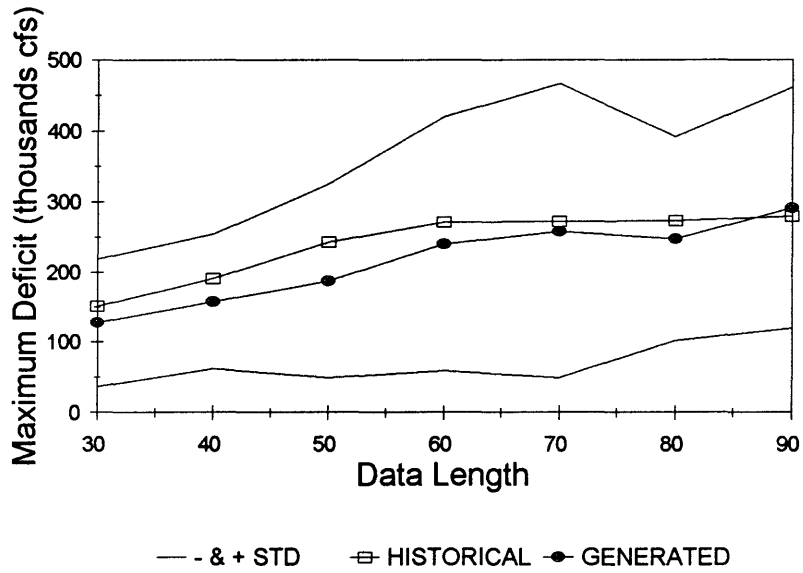
Fig 4-66   Historic and generated maximum deficit for different data lengths for
            lake St.Clair based on GNN Model and regression analysis estimation
            method with i=1, L=30.
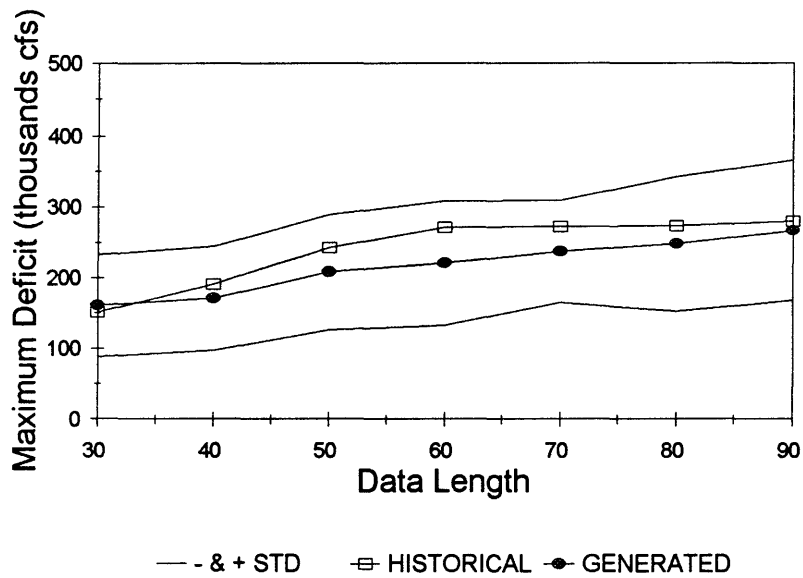
Fig 4-67   Historic and generated maximum deficit for different data lengths for
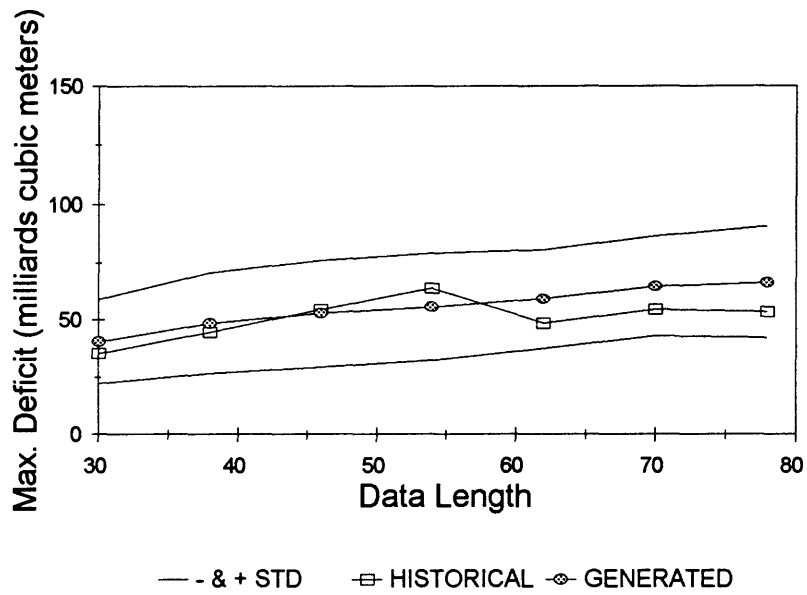   lake St.Clair based on AR(1) Model and MOM estimation method.
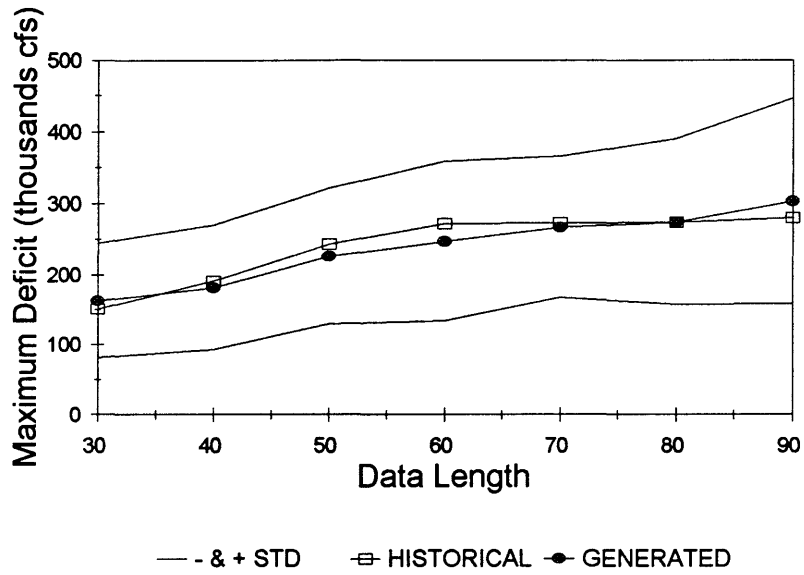
Fig 4-68   Historic and generated maximum deficit for different data lengths for
lake St.Clair based on ARMA(1,1) Model and MOM estimation method.
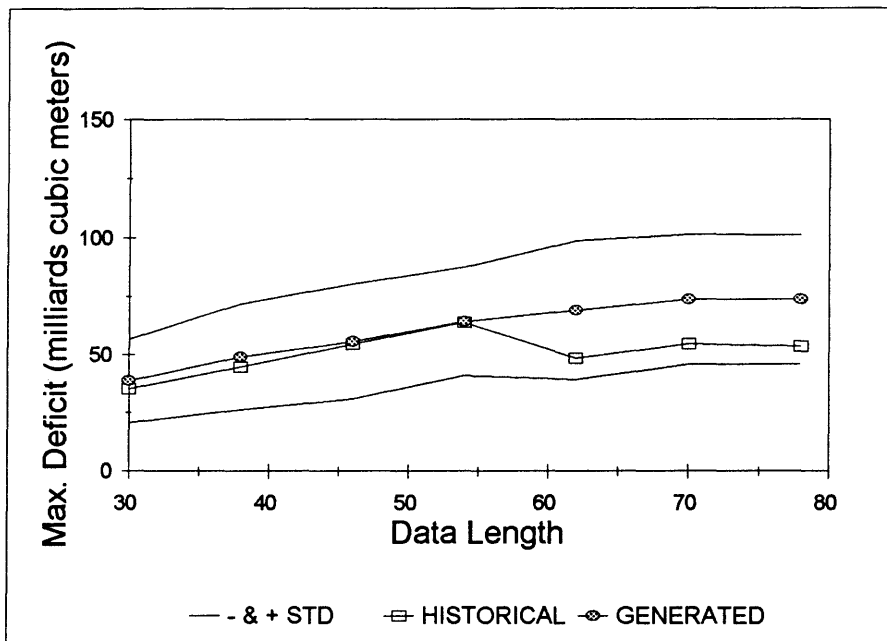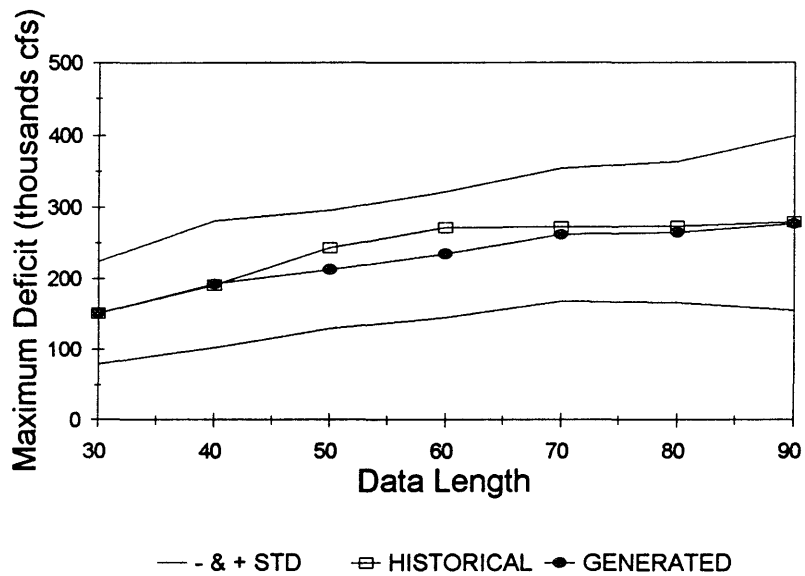

Fig 4-69   Historic and generated smaximum deficit for different data lengths
for lake St.Clair based on ARMA(2,1) Model and least squares
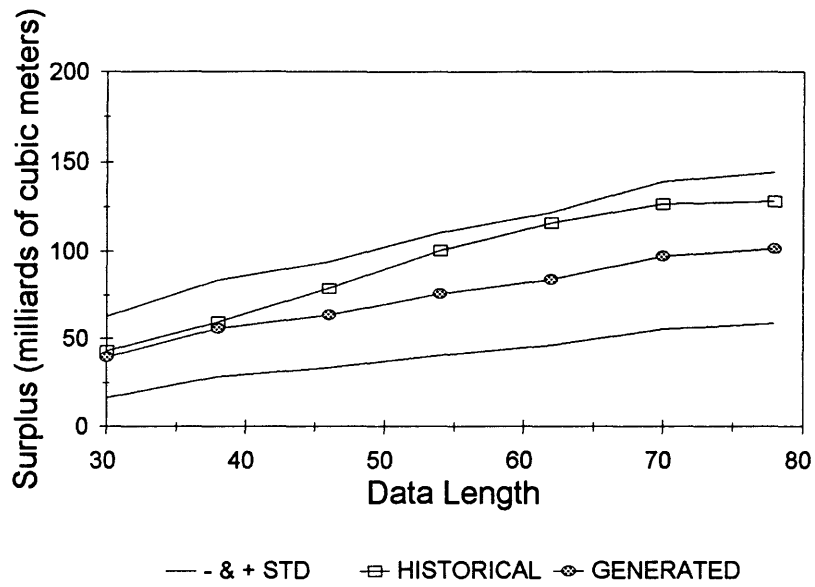estimation method.
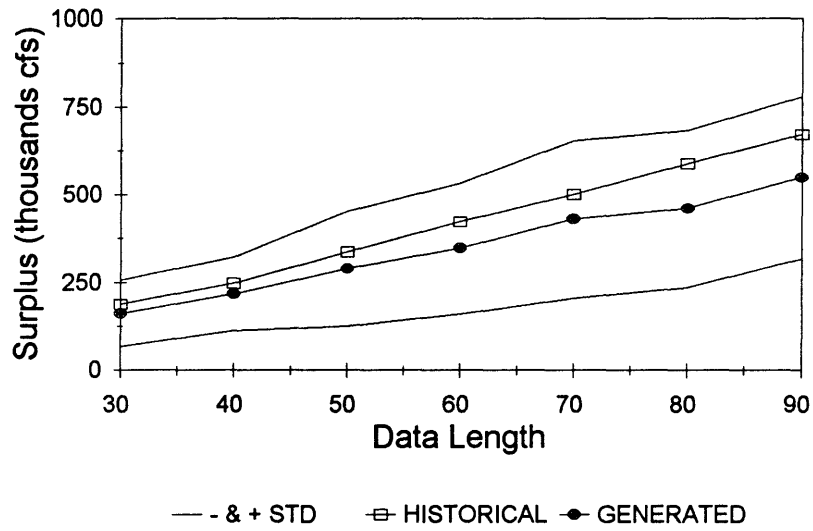
Fig 4-70   Historic and generated surplus for different data lengths for lake
            St.Clair based on GNN Model and regression analysis estimation
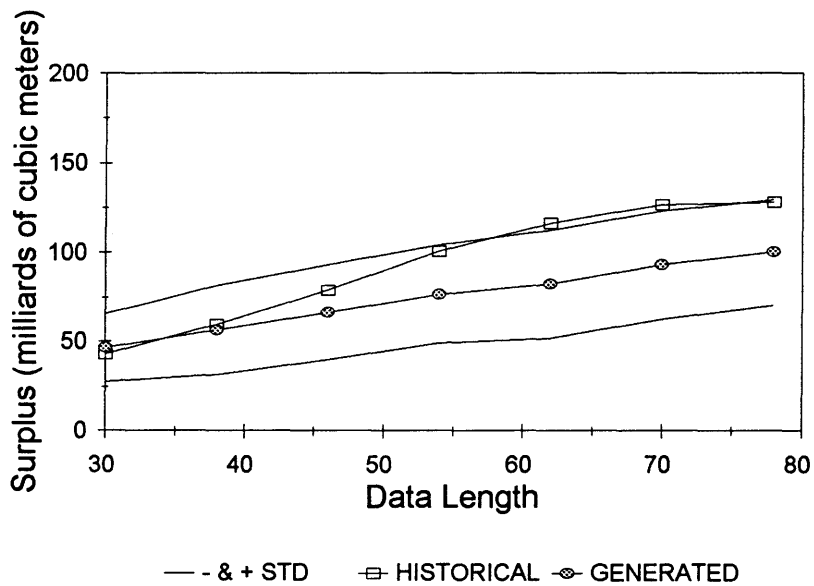   method with i=1, L=30.
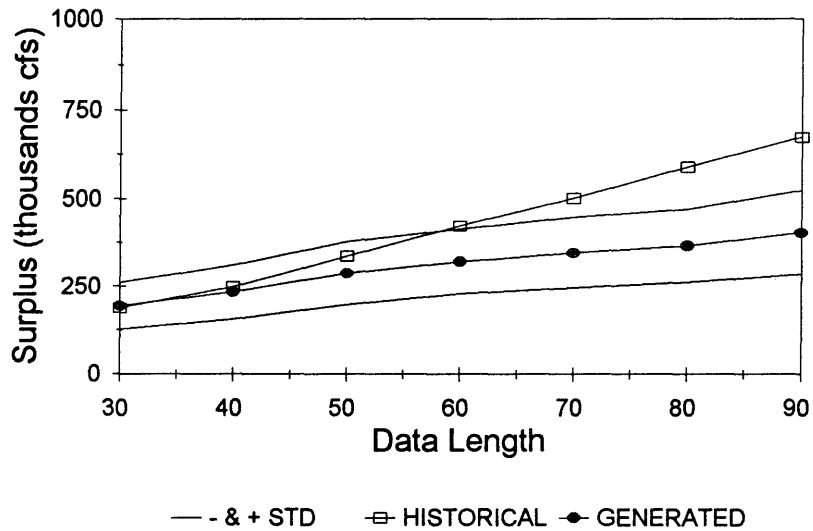
Fig 4-71   Historic and generated surplus for different data lengths for lake
            St.Clair based on AR(1) Model and MOM estimation method.

177
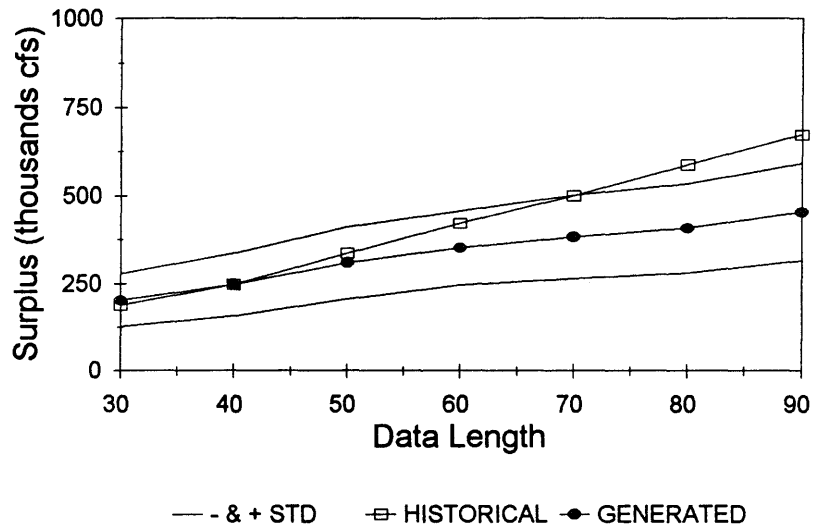
Fig 4-72 Historic and generated surplus for different data lengths for lake
          St.Clair based on ARMA(1,1) Model and MOM estimation method.
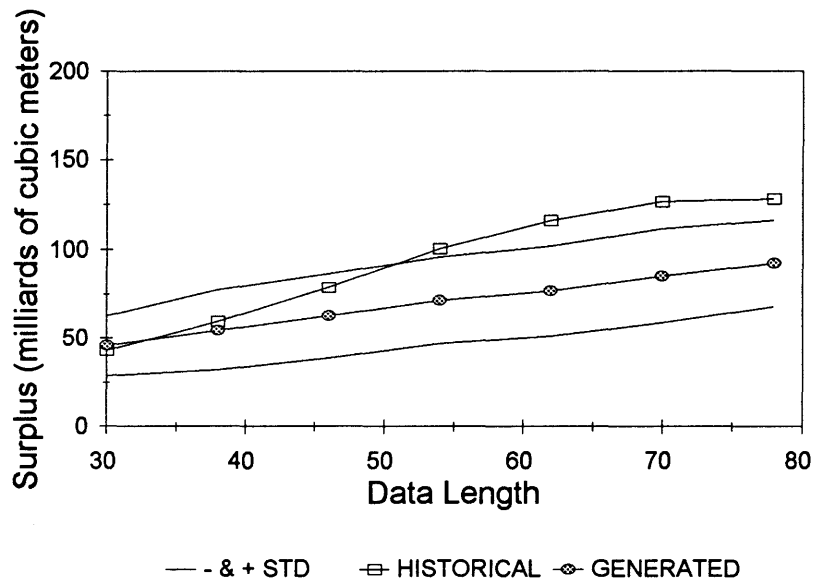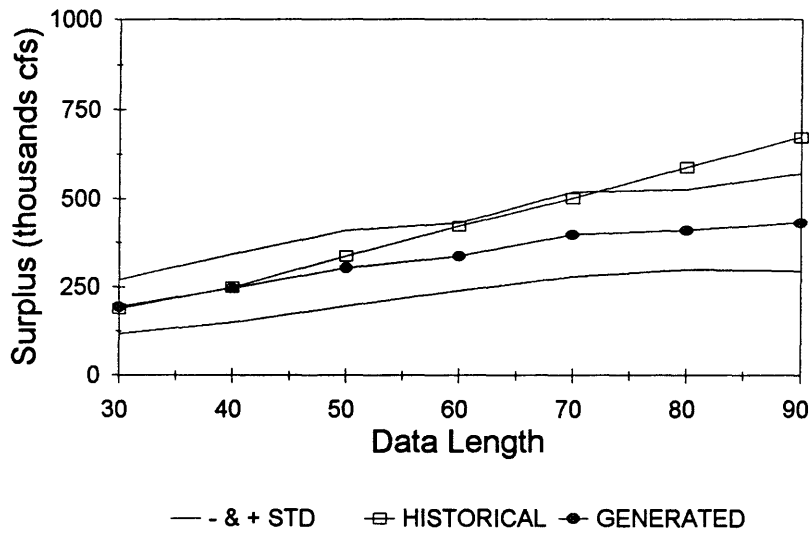

Fig 4-73 Historic and generated surplus for different data lengths for lake
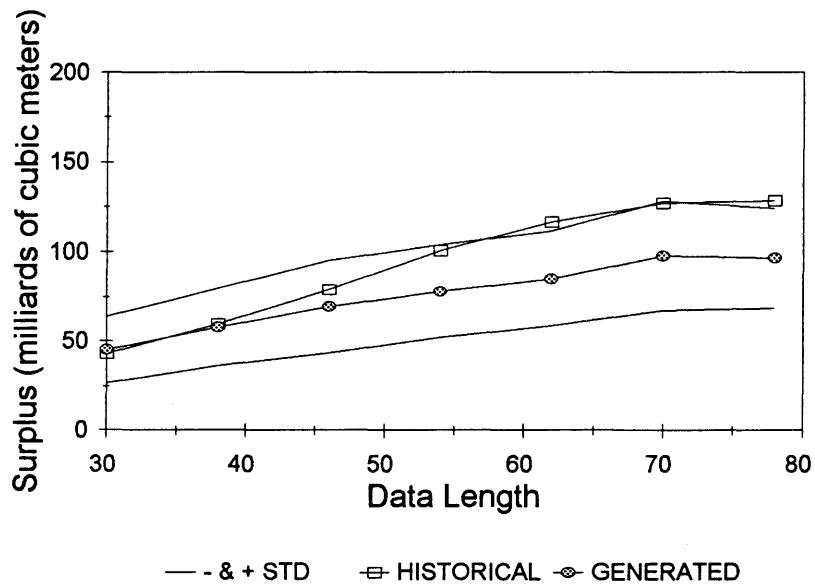   St.Clair based on ARMA(2,1) Model and least squares estimation
          method.

178
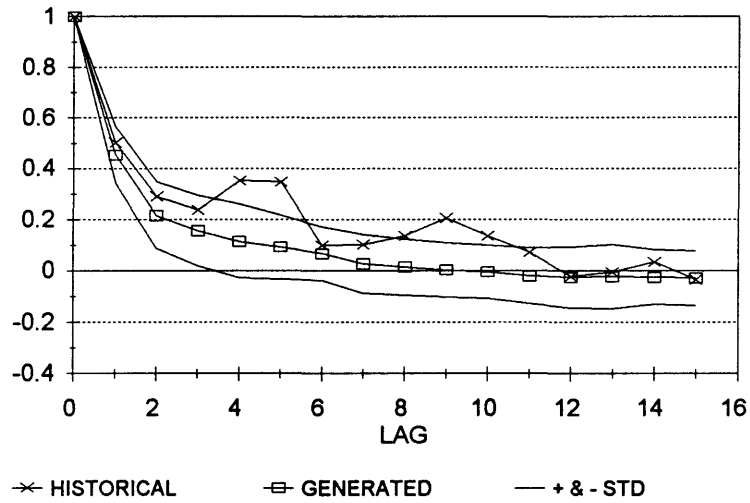
Fig 4-74 Historic and generated autocorrelogram for Lake St.Clair based on
the GNN-1 model and MOM estimation method with i=2,k=3.
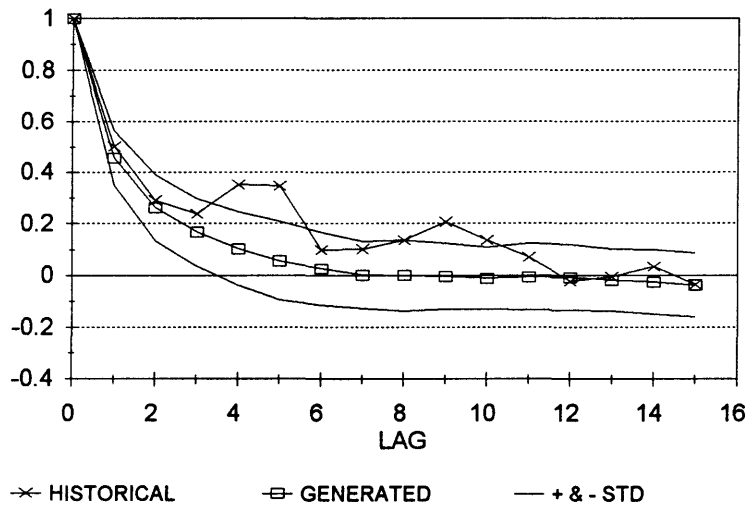
Fig 4-75 Historic and generated autocorrelogram for Lake St.Clair based on
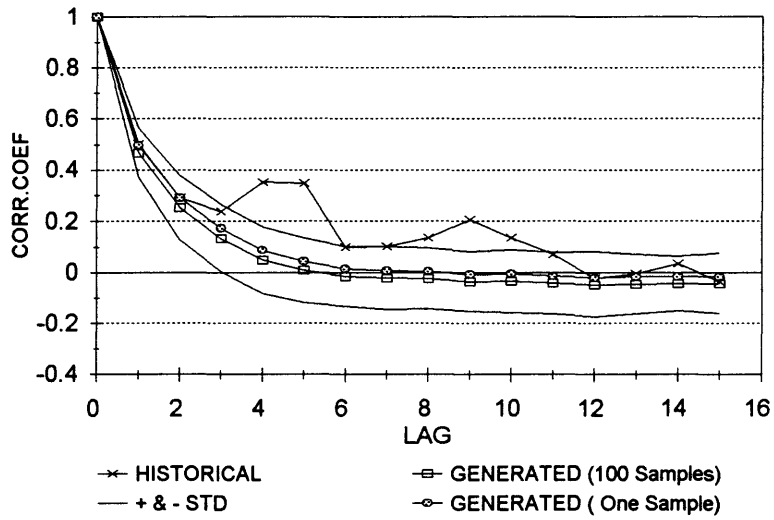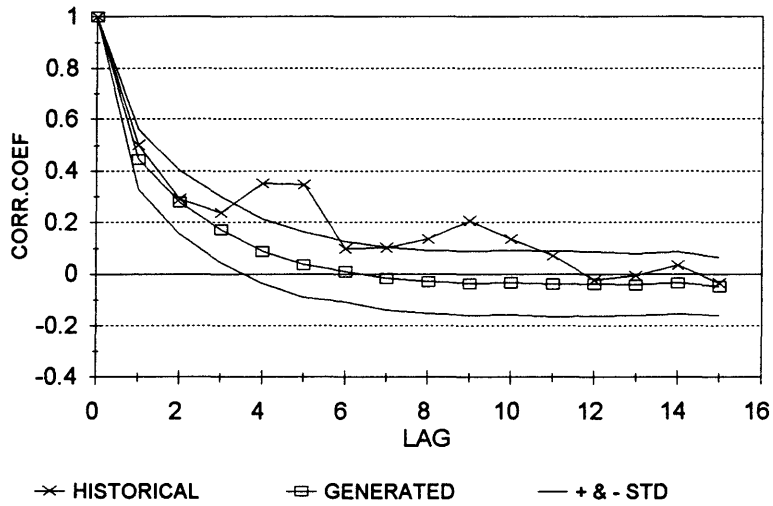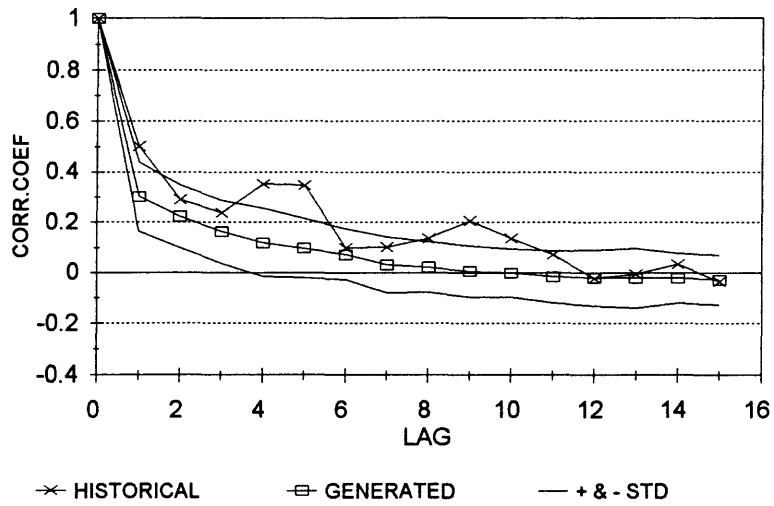the GNN-1 model and MOM estimation method with i=3,k=6.

179
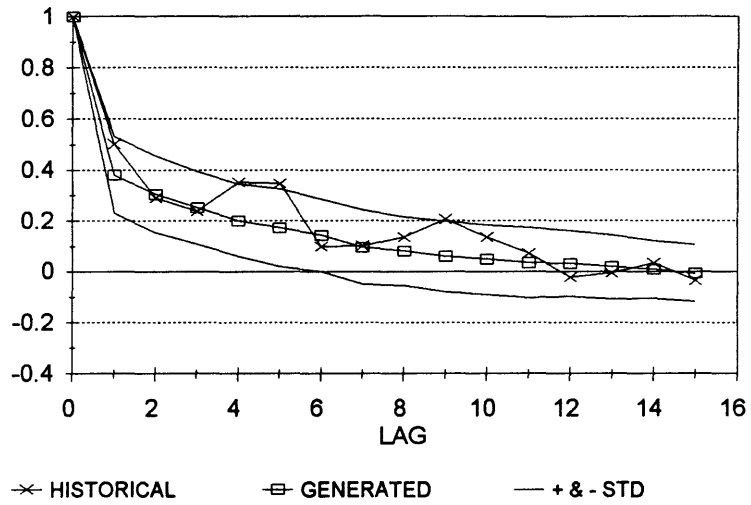
Fig 4-76   Historic and generated autocorrelogram for Lake St.Clair based on
            the GNN-1 model and MOM estimation method with i=4,k=9.
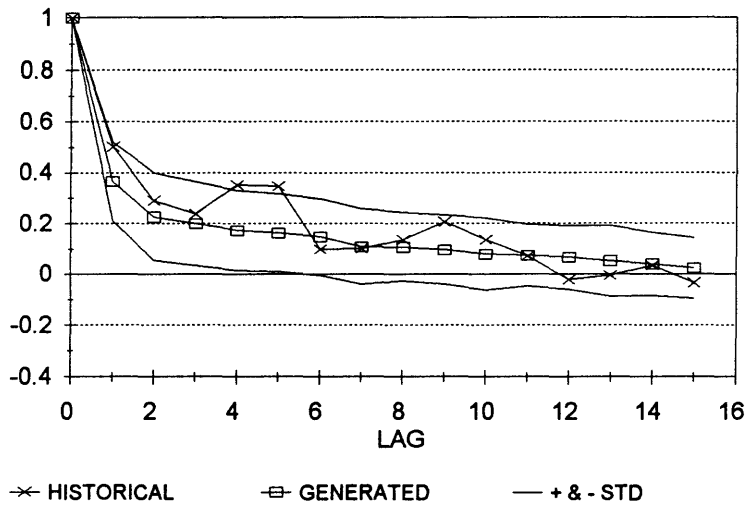
Fig 4-77   Historic and generated autocorrelogram for Lake St.Clair based on
            the GNN-1model and regression analysis estimation method with
            i=4, L=30.

180
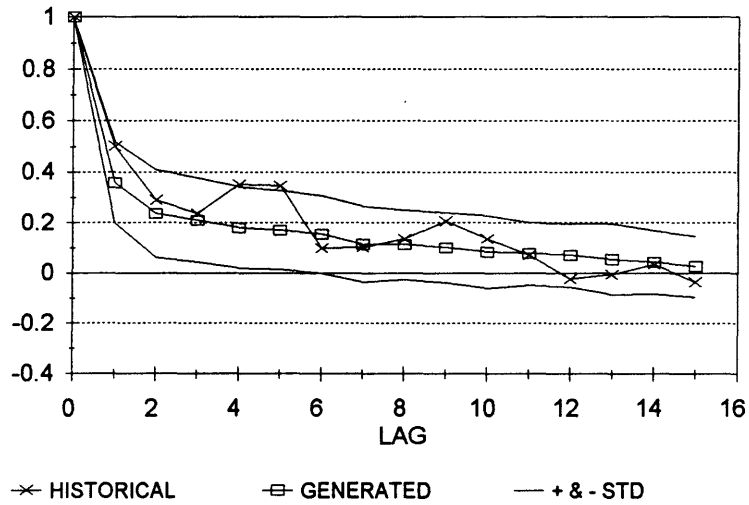
Fig 4-78   Historic and generated autocorrelogram for Lake St.Clair based on
the GNN-1model and regression analysis estimation method with
i=5, L=30.
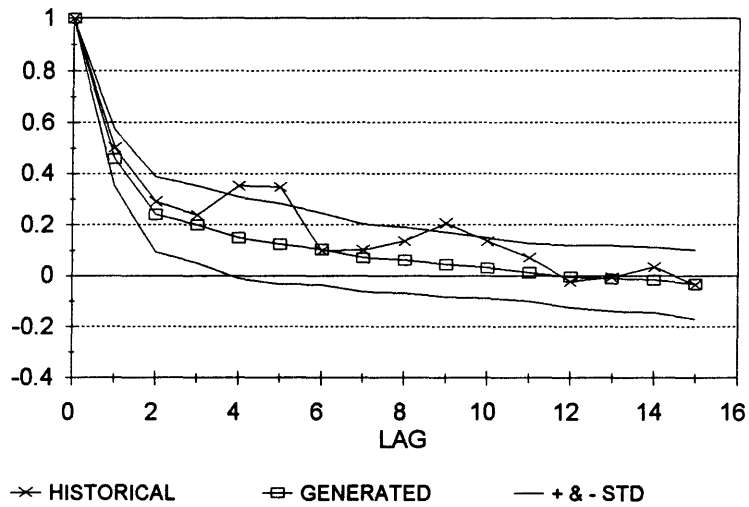
Fig 4-79   Historic and generated autocorrelogram for Lake St.Clair based on
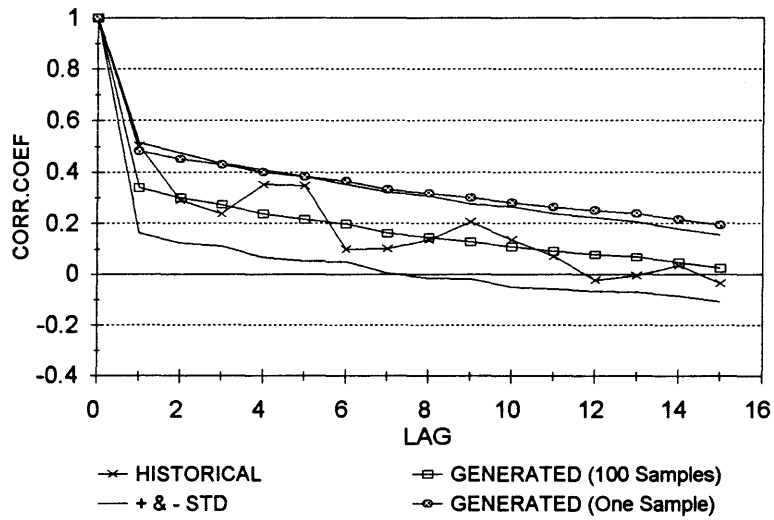the GNN-1model and using range properties estimation method with
i=2, L=30.

181
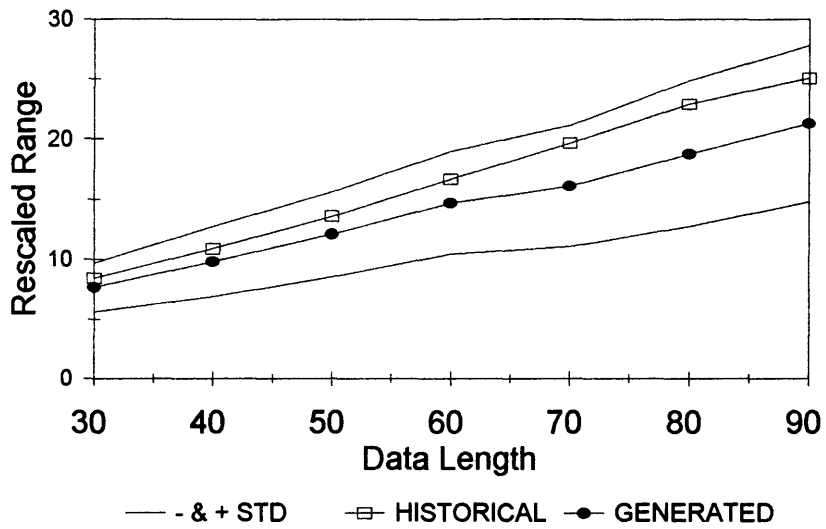
Fig 4-80   Historic and generated rescaled range for Lake St.Clair based on
the GNN-1 model and regression analysis estimation method with
i=4, L=30.
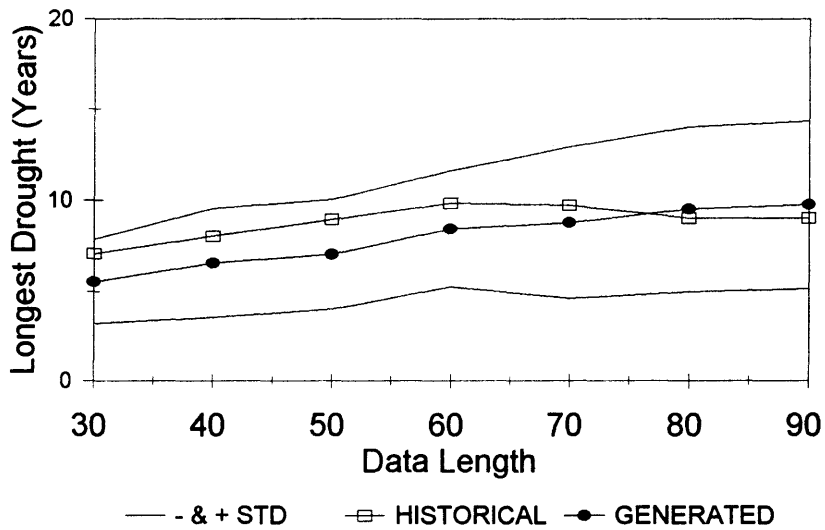
Fig 4-81   Historic and generated longest drought for Lake St.Clair based on
the GNN-1 model and regression analysis estimation method with
i=4, L=30.

CHAPTER V

MULTIVARIATE SHIFTING MEAN PROCESS

5.1 General

Planning and management of water resources systems generally require the analysis of many of hydrological processes such as streamflows at several sites in a given region. Multivariate analysis and modeling has been developed to be applied in such cases. In single site modeling, the purpose of the model is to simulate the temporal dependance for that site. On the other hand, multivariate models are designed to simulate both the temporal (within each site) dependance as well as the spatial (between sites) dependance. Although the concepts to develop a multivariate model are the same as for the univariate models, the mathematics is more complex in the multivariate case especially for full vector multivariate models. This difficulty motivated the use of simpler models known as contemporaneous model. In contemporaneous models, the parameter matrices are assumed to be diagonal and as such the mathematics involved in such models is simpler. In this chapter a contemporaneous GNN model will be developed and tested to check whether the model is capable of simulating the temporal and spatial dependance that exists in water resources systems.

5.2 The Contemporaneous GNN model (CGNN)

5.2.1 General

A contemporaneous model that can be used to simulate processes that exhibit shifts in their mean will be developed in this section. Although this model can be used for

183

multisites, the developed equations will be presented for the bivariate case for illustration purposes.

5.2.2 CGNN Model Formulation and Moment Equations

The contemporaneous GNN process can be written as:

$$\underline{X}_t = \underline{M}^*_t + \underline{Z}_t \qquad\qquad (5\text{-}1)$$

where $\underline{X}_t$ is an (n x 1) vector with elements X_t^i, i=1,...n , n is the number of sites and each element represent the observed process for site i ; $\underline{M}_t$ is an (n x 1) vector of iid normally distributed variables; $\underline{M}^*_t$ is an (n x 1) vector with elements $\underline{M}^{*(i)}_t$, each element is defined for each site i in the same manner as for the univariate case discussed in chapter 3; $\underline{Z}_t$ is an (n x 1) vector of iid normally distributed variables representing the noise terms. Further, the assumptions for this model are:

1. $\underline{M}_t$ is spatially dependent only at lag-zero, that is:

$$cov\ (M_t^{(i)},\ M_{t-k}^{(j)}) = 0 \qquad for\ k \neq 0$$

$$cov\ (M_t^{(i)},\ M_{t-k}^{(j)}) = \rho_m^{ij}\ (0)\ \sigma_m^i\ \sigma_m^j \qquad for\ k = 0$$

2. $\underline{Z}_t$ process is also dependent only at lag-zero, that is:

$$cov\ (Z_t^{(i)},\ Z_{t-k}^{(j)}) = 0 \qquad\qquad for\ k \neq 0$$

$$cov\ (Z_t^{(i)},\ Z_{t-k}^{(j)}) = \rho_z^{ij}(0)\ \sigma_z^{(i)}\ \sigma_z^{(j)} \qquad for\ k = 0$$

where $\rho_m^{ij}(0)$ and $\rho_z^{ij}(0)$ are the lag-zero cross correlations for the {M} series and the {Z} series for sites i and j, respectively. Also define the following terms:

$$E\left(\underline{X}_t \, \underline{X}_t^T\right) = \underline{S}$$

$$E\left(\underline{M}_t^* \, \underline{M}_t^{*T}\right) = \underline{C}^*$$

$$E\left(\underline{M}_t \, \underline{M}_t^T\right) = \underline{C}$$

$$E\left(\underline{Z}_t \, \underline{Z}_t^T\right) = \underline{G}$$

where $\underline{S}$ is the lag-0 variance-covariance matrix of the $\underline{X}_t$ process, $\underline{C}^*$ is the lag-0 variance-covariance matrix of the $\underline{M}_t^*$ process, $\underline{C}$ is the lag-0 variance-covariance matrix of the $\underline{M}_t$ process, and $\underline{G}$ is the lag-0 variance-covariance matrix of the $\underline{Z}_t$ process. The ij elements of the above matrices can be written as:

S^{ij} = lag-0 covariance between sites i and j of the $\underline{X}_t$ process.

C^{*ij} = lag-0 covariance between sites i and j of the $\underline{M}_t^*$ process.

C^{ij} = lag-0 covariance between sites i and j of the $\underline{M}_t$ process.

G^{ij} = lag-0 covariance between sites i and j of the $\underline{Z}_t$ process.

Notice that in chapter 3, it was shown that for each site

$$var\ (M_t^*) = var\ (M_t) = \sigma_m^2 \ .$$

As a result the ii element of the $\underline{C}^*$ matrix can be written as

$$C^{*ii} = C^{ii}$$

In the bivariate case, for example, one can write the above matrices as:

$$\underline{S} = \begin{bmatrix} \sigma_x^{2(1)} & cov\ (X_t^{(1)},\ X_t^{(2)}) \\ cov\ (X_t^{(1)},\ X_t^{(2)}) & \sigma_x^{2(2)} \end{bmatrix}$$

$$\underline{C}^* = \begin{bmatrix} \sigma_m^{2(1)} & cov\ (M_t^{*(1)},\ M_t^{*(2)}) \\ cov\ (M_t^{*(2)},\ M_t^{*(1)}) & \sigma_m^{2(2)} \end{bmatrix}$$

$$
\underline{C} = \begin{bmatrix} \sigma_m^{2(1)} & cov\ (M_t^{(1)},\ M_t^{(2)}) \\ & \\ cov\ (M_t^{(2)},\ M_t^{(1)}) & \sigma_m^{2(2)} \end{bmatrix}
$$

$$
\underline{G} = \begin{bmatrix} \sigma_z^{2(1)} & cov\ (Z_t^{(1)},\ Z_t^{(2)}) \\ & \\ cov\ (Z_t^{(1)},\ Z_t^{(2)}) & \sigma_z^{2(2)} \end{bmatrix}
$$

The model in Eq. (5-1) can be decoupled into univariate models for each site. For example, for the bivariate case

$$
X_t^{(1)} = M_t^{*(1)} + Z_t^{(1)} \tag{5-2}
$$

$$
X_t^{(2)} = M_t^{*(2)} + Z_t^{(2)} \tag{5-3}
$$

Without loss of generality, let us assume that the mean of the process {X} is set to zero at each site, that is

$$
E(X_t^{(i)}) = 0
$$

Multiplying both sides of Eq. (5-1) by $\underline{X}_t^T$

187

$$\underline{X}_t \underline{X}_t^T = \underline{M}_t^* \underline{X}_t^T + \underline{Z}_t \underline{X}_t^T \tag{5-4}$$

$$= \underline{M}_t^* \underline{M}_t^{*T} + \underline{M}_t^* \underline{Z}_t^T + \underline{Z}_t \underline{M}_t^{*T} + \underline{Z}_t \underline{Z}_t^T$$

and taking expectation

$$E(\underline{X}_t \underline{X}_t^T) = E(\underline{M}_t^* \underline{M}_t^{*T}) + E(\underline{M}_t^* \underline{Z}_t^T) + E(\underline{Z}_t \underline{M}_t^{*T}) + E(\underline{Z}_t \underline{Z}_t^T) \tag{5-5}$$

and because of the independence of {M} and {Z} it may be shown that

$$E(\underline{M}_t^* \underline{Z}_t^T) = E(\underline{Z}_t \underline{M}_t^{*T}) = 0 \tag{5-6}$$

As a result, Eq. (5-5) gives

$$E(\underline{X}_t \underline{X}_t^T) = E(\underline{M}_t^* \underline{M}_t^{*T}) + E(\underline{Z}_t \underline{Z}_t^T) \tag{5-7}$$

or

$$\underline{S} = \underline{C}^* + \underline{G} \tag{5-8}$$

Then the ij element of matrix S can be written as:

$$S^{ij} = C^{*ij} + G^{ij} \tag{5-9}$$

188

Based on the assumptions of the model, two types of the model are identified:

(1) A model where the lengths of the shifts are not the same at all stations (i.e. different geometric distribution parameter p for each station).

(2) A model where the lengths of the shifts are the same at all stations (i.e. same parameter p).

It can be shown that the first model is not stationary and thus will not be considered here. The proof of the non-stationarity of this model is shown in Appendix B. Further, for the stationary model, it can be shown that:

$$cov\ (M_t^{*(i)},\ M_{t-k}^{*(j)})\ =\ cov\ (M_t^{(i)},\ M_{t-k}^{(j)})\ =\ \rho_m^{ij}(0)\ \sigma_m^{(i)}\ \sigma_m^{(j)} \qquad k=0 \qquad (5\text{-}10)$$

As a result, Eq. (5-9) gives

$$S^{ij}\ =\ C^{ij}\ +\ G^{ij} \qquad\qquad (5\text{-}11)$$

or

$$\sigma_x^{2(i)}\ =\ \sigma_m^{2(i)}\ +\ \sigma_z^{2(i)} \qquad\qquad (5\text{-}12)$$

$$\rho_x^{ij}(0)\ \sigma_x^{(i)}\ \sigma_x^{(j)}\ =\ \rho_m^{ij}(0)\ \sigma_m^{(i)}\ \sigma_m^{(j)}\ +\ \rho_z^{ij}(0)\ \sigma_z^{(i)}\ \sigma_z^{(j)} \qquad (5\text{-}13)$$

189

For the bivariate model, Eq. (5-11) gives

$$
\begin{bmatrix}
\sigma_x^{2^{(1)}} & \rho_x^{12}(0)\ \sigma_x^{(1)}\ \sigma_x^{(2)} \\
\\
\rho_x^{12}(0)\ \sigma_x^{(1)}\ \sigma_x^{(2)} & \sigma_x^{2^{(2)}}
\end{bmatrix}
=
\begin{bmatrix}
\sigma_m^{2^{(1)}} & \rho_m^{12}(0)\ \sigma_m^{(1)}\ \sigma_m^{(2)} \\
\\
\rho_m^{12}(0)\ \sigma_m^{(1)}\ \sigma_m^{(2)} & \sigma_m^{2^{(2)}}
\end{bmatrix}
$$

$$
+
\begin{bmatrix}
\sigma_z^{2^{(1)}} & \rho_z^{12}(0)\ \sigma_z^{(1)}\ \sigma_z^{(2)} \\
\\
\rho_z^{12}(0)\ \sigma_z^{(1)}\ \sigma_z^{(2)} & \sigma_z^{2^{(2)}}
\end{bmatrix}
$$

(5-14)

The spatially correlated {Z} process can be modeled as

$$
\underline{Z}_t = \underline{B}\ \underline{\varepsilon}_t
$$

(5-15)

where $\underline{\varepsilon}_t$ is an (n x 1) vector of standardized normal variables independent in both time and space and $\underline{B}$ is an (n x n) parameter matrix. Also, the spatially correlated {M} process can be modeled as

$$
\underline{M}_t = \underline{F}\ \underline{\zeta}_t
$$

(5-16)

where $\underline{\zeta}_t$ is an (n x 1) vector of standardized normal variables independent in both time

190

and space and E is an (n x n) parameter matrix. Also in this case, it can be shown that:

$$cov\ (M_t^{*(i)}, M_{t-1}^{*(j)}) = \rho_m^{ij}(0)\ \sigma_m^{(i)}\ \sigma_m^{(j)}\ (1 - \bar{p})$$

(5-17)

where $\bar{p}$ is the average of the estimated parameter p for all sites.

Based on the assumptions of this model, two special cases are identified here:

Case 1:

1. $\underline{M}_t$ is spatially independent, that is:

$$cov\ (M_t^{(i)}, M_{t-k}^{(j)}) = 0 \quad for\ k \ge 0$$

As a result, it can be shown that

$$cov\ (M_t^{*(i)}, M_{t-k}^{*(j)}) = 0 \quad for\ k \ge 0$$

2. $\underline{Z}_t$ is spatially dependent only at lag-zero, that is:

$$cov\ (Z_t^{(i)}, Z_{t-k}^{(j)}) = 0 \qquad\qquad for\ k \ne 0$$

$$cov\ (Z_t^{(i)}, Z_{t-k}^{(j)}) = \rho_z^{ij}(0)\ \sigma_z^{(i)}\ \sigma_z^{(j)} \qquad for\ k = 0$$

and as a result Eq. (5-9) gives

$$S^{ii} = C^{ii} + G^{ii}$$

(5-18)

$$S^{ij} = G^{ij} \qquad for\ i \ne j$$

(5-19)

191

and for the bivariate model

$$
\begin{bmatrix}
\sigma_x^{2^{(1)}} & \rho_x^{12}(0)\ \sigma_x^{(1)}\ \sigma_x^{(2)} \\
\\
\\
\rho_x^{12}(0)\ \sigma_x^{(1)}\ \sigma_x^{(2)} & \sigma_x^{2^{(2)}}
\end{bmatrix}
=
\begin{bmatrix}
\sigma_m^{2^{(1)}} & 0 \\
\\
\\
0 & \sigma_m^{2^{(2)}}
\end{bmatrix}
$$

(5-20)

$$
+
\begin{bmatrix}
\sigma_z^{2^{(1)}} & \rho_z^{12}(0)\ \sigma_z^{(1)}\ \sigma_z^{(2)} \\
\\
\\
\rho_z^{12}(0)\ \sigma_z^{(1)}\ \sigma_z^{(2)} & \sigma_z^{2^{(2)}}
\end{bmatrix}
$$

Case 2

1. $\underline{M}_t$ is spatially dependent only at lag-zero, that is:

$$cov\ (M_t^{(i)},\ M_{t-k}^{(j)}) = 0 \qquad for\ k \neq 0$$

$$cov\ (M_t^{(i)},\ M_{t-k}^{(j)}) = \rho_m^{ij}\ (0)\ \sigma_m^i\ \sigma_m^j \qquad for\ k = 0$$

2. $\underline{Z}_t$ is spatially independent, that is:

$$cov\ (Z_t^{(i)},\ Z_{t-k}^{(j)}) = 0 \qquad for\ k \geq 0$$

As a result, Eq. (5-9) gives

$$S^{ii} = C^{ii} + G^{ii}$$

(5-21)

192

$$S^{ij} = C^{ij} \qquad \text{for} \quad i \neq j \qquad\qquad (5\text{-}22)$$

and for the bivariate model

$$
\begin{bmatrix}
\sigma_x^{2^{(1)}} & \rho_x^{12}(0)\ \sigma_x^{(1)}\ \sigma_x^{(2)} \\[2em]
\rho_x^{12}(0)\ \sigma_x^{(1)}\ \sigma_x^{(2)} & \sigma_x^{2^{(2)}}
\end{bmatrix}
=
\begin{bmatrix}
\sigma_m^{2^{(1)}} & \rho_m^{12}(0)\ \sigma_m^{(1)}\ \sigma_m^{(2)} \\[2em]
\rho_m^{12}(0)\ \sigma_m^{(1)}\ \sigma_m^{(2)} & \sigma_m^{2^{(2)}}
\end{bmatrix}
$$

$$
+
\begin{bmatrix}
\sigma_z^{2^{(1)}} & 0 \\[2em]
0 & \sigma_z^{2^{(2)}}
\end{bmatrix}
$$

$$\qquad\qquad (5\text{-}23)$$

5.2.3 Parameter Estimation of the CGNN Model

The parameter p at each station can be estimated using the univariate model at each station. Note that for this case the assumption is that all stations have the same parameter p. It is clear that we might obtain different estimated values of p at each station, hence an average value $\bar{p}$ can be found as:

$$\bar{p} = \frac{1}{n}\sum_{i=1}^{n} \hat{p}^{(i)} \qquad\qquad (5\text{-}24)$$

193

where $\hat{p}^{(i)}$ is the estimated parameter based on the univariate model at station (i). $\sigma_m^{2^{(i)}}$ and $\sigma_z^{2^{(i)}}$ can be estimated from the univariate model at each site. The parameters $\rho_m^{ij}(0)$ and $\rho_z^{ij}(0)$ can be estimated as follows: Multiplying both sides of Eq. (5-1) by $\underline{X}_{t-1}^T$ and taking expectation and using Eq. (5-17), it can be shown that :

$$\rho_x^{ij}(1) \; \sigma_x^{(i)} \; \sigma_x^{(j)} = \rho_m^{ij}(0) \; \sigma_m^{(i)} \; \sigma_m^{(j)} \; (1 - \overline{p}) \tag{5-25}$$

and using Eq. (5-25), $\rho_m^{ij}(0)$ can be estimated as:

$$\hat{\rho}_m^{ij}(0) = \frac{\hat{\rho}_x^{ij}(1) \; \hat{\sigma}_x^{(i)} \; \hat{\sigma}_x^{(j)}}{\hat{\sigma}_m^{(i)} \; \hat{\sigma}_m^{(j)} \; (1 - \overline{p})} \tag{5-26}$$

Also, Eq. (5-13) can be used to estimate $\rho_z^{ij}(0)$ as follows:

$$\hat{\rho}_z^{ij}(0) = \frac{\hat{\rho}_x^{ij}(0) \; \hat{\sigma}_x^{(i)} \; \hat{\sigma}_x^{(j)} - \hat{\rho}_m^{ij}(0) \; \hat{\sigma}_m^{(i)} \; \hat{\sigma}_m^{(j)}}{\hat{\sigma}_z^{(i)} \; \hat{\sigma}_z^{(j)}} \tag{5-27}$$

It remains to estimate the parameter matrix $\underline{B}$. In the bivariate model, Eq. (5-15) can be written as:

$$\begin{bmatrix} Z_t^{(1)} \\ Z_t^{(2)} \end{bmatrix} = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} \begin{bmatrix} \varepsilon_t^{(1)} \\ \varepsilon_t^{(2)} \end{bmatrix} \tag{5-28}$$

194

where $\varepsilon_t^{(1)}$ *and* $\varepsilon_t^{(2)}$ are spatially uncorrelated white noise series with mean$=0$ and variances

$$\sigma_\varepsilon^{2^{(1)}} = \sigma_\varepsilon^{2^{(1)}} = 1.0 \text{ and } b_{ij} = \text{Coefficients to be estimated.}$$

Multiplying both sides of Eq. (5-15) by $\underline{Z}_t^T$ and taking expectation

$$E\,(\underline{Z}_t\,\underline{Z}_t^T) = E\,(\underline{B}\,\underline{\varepsilon}_t\,\underline{Z}_t^T)$$

$$= E\,(\underline{B}\,\underline{\varepsilon}_t\,\underline{\varepsilon}_t^T\,\underline{B}_t^T)$$

$$\underline{G} = \underline{B}\,\underline{B}^T \tag{5-29}$$

The above equation can be used to estimate the parameter matrix B. If matrix G is positive definite then, the B matrix can be assumed to be lower triangular (i.e. $b_{12} = 0$ in Eq. (5-28)). In this case the coefficients of the matrix B can be found directly. For example for the bivariate case

$$b_{11} = \sqrt{\sigma_z^{2^{(1)}}}$$

$$b_{21} = \frac{\rho_z^{12}(0)\,\sigma_z^{(1)}\,\sigma_z^{(2)}}{b_{11}}$$

$$b_{22} = \sqrt{\sigma_z^{2^{(2)}} - (b_{21})^2}$$

If matrix G is positive or semi positive definite, the Singular Value Decomposition (SVD) method can be used to estimate B . In this method, matrix B is assumed to be full

(i.e. $b_{12} \neq 0$ for the bivariate case). The matrix F of Eq. (5-16) can be estimated in the same manner as was done for matrix B. For the special cases mentioned in the above section, the parameters are estimated as follows:

Case 1:

The parameters p, σ_m^2 and σ_z^2 at each site can be estimated by one of the methods mentioned before based on the univariate model for each site. ρ_z^{ij} (0) can be estimated from

Eq. (5-13) as

$$\hat{\rho}_z^{ij}(0) = \frac{\hat{\rho}_x^{ij}(0) \; \hat{\sigma}_x^{(i)} \; \hat{\sigma}_x^{(j)}}{\hat{\sigma}_z^{(i)} \; \hat{\sigma}_z^{(j)}} \tag{5-30}$$

Case 2

In this case, the parameter p at each station can be estimated using the univariate model at each station. Parameter $\bar{p}$ can be estimated from Eq.(5-24). ρ_m^{ij} (0) can be estimated using Eq. (5-13) as:

$$\hat{\rho}_m^{ij}(0) = \frac{\hat{\rho}_x^{ij}(0) \; \hat{\sigma}_x^{(i)} \; \hat{\sigma}_x^{(j)}}{\hat{\sigma}_m^{(i)} \; \hat{\sigma}_m^{(j)}} \tag{5-31}$$

5.2.4 Application of the CGNN Model to Observed Hydrologic Data

Nile River Basin

The CGNN model was used to model the annual streamflows for **Aswan**, Khartoum, Roseires, and Malakal based on the data for the period 1914-1983. The model

196

parameters at each site namely $p^{(i)}$, $\sigma_m^{2^{(i)}}$, and $\sigma_z^{2^{(i)}}$ were estimated by using the regression method. The parameters $\rho_m^{ij}(0)$ and $\rho_z^{ij}(0)$ were estimated by using Eqs. (5-26) and (5-27) respectively. The estimated parameters are:

For Aswan: p = 0.105, σ_m^2 = 24.55, σ_z^2 = 78.94

For Khartoum: p = 0.065, σ_m^2 = 12.15, σ_z^2 = 69.13

For Roseires: p = 0.068, σ_m^2 = 12.86, σ_z^2 = 54.28

For Malakal: p = 0.084, σ_m^2 = 27.03, σ_z^2 = 4.67

MATRIX-G $\longrightarrow C_y^{(0)}$

78.936	69.091	61.445	2.993
69.091	69.129	58.412	1.066
61.445	58.412	54.280	1.117
2.993	1.066	1.117	4.668

For generating annual flows, the following procedure was followed:

1. Two vectors of standard normal numbers $\varepsilon_t^{(i)}$ and $\zeta_t^{(i)}$, i=1,...,n were generated where n is the number of sites (four in this case). Then Eq. (5-15) was used to generate a vector of cross correlated normal random numbers $Z_t^{(i)}$, i=1,...,n. In a similar manner Eq. (5-16) was used for generating a vector of cross correlated normal random numbers $M_t^{(i)}$, i=1,...,n.

2. Based on the estimated parameters $p^{(i)}$ at each site i, the average parameter $\bar{p}$ was obtained by using Eq. (5-24). Consequently, a vector of geometric random numbers was generated. These numbers represent the random lengths of common means { N_l }, l=1,2,3,4,...

3. Using the generated numbers $M_l^{(i)}$, $i = 1, ..., n$ and the lengths $\{ N_l \}$, $l = 1, 2, 3, 4, ...$ a vector of autocorrelated and cross correlated normal numbers $M_l^{*(i)}$, $i = 1, ..., n$ were generated.

4. Subsequently, Eq.(5-1) was used for generating the vector of autocorrelated and cross- correlated numbers $X_l^{(i)}$.

100 multivariate samples of length equal to the historical samples were generated using the above procedure. The generated statistics were compared to the observed ones to check weather the CGNN model was capable of preserving some important and basic statistics of the observed data. The results of the generation experiment are shown in Tables 5-1 and 5-2. It is shown in these tables that the mean and the standard deviation are well preserved for all the four stations. In general, the CGNN model was also capable of preserving the longest drought except for Aswan as shown in Table 5-1. The rescaled range and Hurst coefficient are also well preserved for each site as shown in Table 5-1. The Maximum deficit was relatively well preserved for Aswan, Khartoum, and Roseires. However, there was relatively significant bias in Malakal case. The Surplus was relatively well preserved for Aswan and Roseires and some bias exists for Khartoum and Malakal. Table 5-2 show the historic and the generated lag-0 cross correlation between the stations which shows that the CGNN model was able to preserve the spatial correlation structure that exists between the studied stations. Figures 5-1 through 5-4 show the generated autocorrelograms for the CGNN model. These figures suggest that the CGNN performed well in preserving the correlogram. From the above discussion, one can conclude that the CGNN is doing a relatively good job in preserving the basic short term and long term

statistics on the univariate level as well as preserving the spatial correlation structure that exists between the stations. It should be mentioned here that simulation experiments were also done based on cases 1 and 2 of CGNN model mentioned in section 5.2.2. The performance in preserving the short and long term statistics was similar to the CGNN model above. However, the lag-0 cross correlations were not preserved well in both cases especially the high cross correlations such as the ones between Khartoum and Aswan, Roseires and Aswan, and Roseires and Khartoum.

Similar generation experiments where done with contemporaneous multivariate CARMA(1,1) and CARMA(2,1) models. The parameters of the models were estimated using the least squares (LS) method. These models performed well in preserving the mean, standard deviation, and the lag-0 cross correlations between the stations as shown in Tables 5-3 through 5-6. The models performed well in preserving the longest drought for all station except for Aswan. The rescaled range for Malakal was not well preserved by the CARMA models as shown in Tables 5-3 and 5-5. The CGNN model did a better job than CARMA models in this case. For the maximum deficit the CGNN model performed relatively better than CARMA models for Khartoum and Roseires. The opposite is true for Aswan and Malakal. Regarding the surplus, the biases were smaller for the CGNN model than the CARMA models as shown in Tables 5-1, 5-3, and 5-5.

Great Lakes System

The CGNN model was fitted to the annual (NBS) for lakes Superior, Michigan-Huron, St.Clair, Erie and Ontario. The data used was for the period 1900 - 1989. The

parameters of the fitted model were estimated at each site using the regression method. The estimated parameters were as follows:

For Superior: $p = 0.050$, $\sigma_m^2 = 7689.69$, $\sigma_z^2 = 33477.49$

For Mich-Huron: $p = 0.015$, $\sigma_m^2 = 20293.69$, $\sigma_z^2 = 7645.66$

For St.Clair: $p = 0.141$, $\sigma_m^2 = 412.75$, $\sigma_z^2 = 309.35$

For Erie: $p = 0.146$, $\sigma_m^2 = 2692.07$, $\sigma_z^2 = 9095.68$

For Ontario: $p = 0.244$, $\sigma_m^2 = 3114.85$, $\sigma_z^2 = 6336.37$

MATRIX-G

33477.500	25080.900	486.393	4167.745	1839.909
25080.900	76045.660	2688.940	11738.520	12877.730
486.393	2688.940	309.346	692.805	781.848
4167.745	11738.520	692.805	9095.684	4075.028
1839.909	12877.730	781.848	4075.028	6336.374

Following the same procedure as for the Nile River System 100 multivariate samples of length equal to the historical samples were generated by CGNN model . The results are shown in Tables 5-7 and 5-8. The CGNN model performed well in preserving the mean and the standard deviation for all lakes. The longest drought was relatively well preserved for all sites. Similarly, the rescaled range and Hurst coefficient were also well preserved. The surplus was also relatively well preserved although there was some significant bias for lakes Superior and Erie. Also, the CGNN was also able to preserve the spatial correlation structure that exists in the Great Lakes system as shown in Table 5-8. The CGNN model also performed well in preserving the autocorrelograms for each site as shown in Figs. 5-

13 through 5-17.

As was done for the Nile basin, simulation experiments where done with CARMA(1,1) and CARMA(2,1) models. The parameters of these models were estimated using the least squares method. It should be noted here that for lake Erie an ARMA(1,0) was used to estimate the univarite model parameters. The use of a univariate ARMA(1,1) and ARMA(2,1) resulted in unreasonable estimates of the variance-covariance matrix of the residuals G which caused the CARMA models to be unable to preserve the historical statistics upon generation. The CARMA models performed well in preserving the mean, standard deviation, and the lag-0 cross correlations between the stations as shown in Tables 5-9 through 5-12 . The longest drought was well preserved for all lakes. However, the rescaled range was not preserved for lakes St.Clair and Erie and was significantly underestimated. Remember that the CGNN model did preserve the rescaled range at these sites as discussed above. The performance of the CARMA models was relatively better than the CGNN model in preserving the maximum deficit except for lake Superior. The opposite is completely true for the surplus as shown in Tables 5-7, 5-9, and 5-11.

From the results of the generation experiments one can conclude that there is no clear winner between CGNN model and CARMA models. In some aspects, the CGNN model performed better than the CARMA models and in others CARMA performed better. This raises the question which is can we somehow combine the two models (i.e. CARMA and CGNN) and use CARMA for those sites that are best modeled by it and CGNN for those sites in which CGNN performed better than CARMA. The answer is yes and the model is GNN-CARMA.

5.3 The GNN-CARMA Model

5.3.1 General

In the CGNN model, the assumption was made that each site be modeled by a GNN model univariatly and then these models can be combined to form the CGNN. In the model to be developed here, the sites will be modeled by using a combination of GNN and ARMA models. This model can be used in a region for which some sites exhibit shifts and the other sites do not appear to have the shifting phenomenon. The univariate models for each site will be combined to form the contemporaneous multivariate model which will be called GNN-CARMA(p,q) model. In this section, the GNN-CARMA (p,q) model will be developed and its moment equations will be derived. For illustration purposes, the derivation will be done for the GNN-CARMA(1,1) model. Finally, the equations developed will be generalized to the GNN-CARMA(p,q) model.

5.3.2 Model Formulation and Moment Equations of the GNN-CARMA Model

The GNN-CARMA(p,q) model can be written as:

$$\underline{\Phi}(B) \ \underline{X}_t = \underline{\Psi} \ \underline{M}_t^* + \underline{\Theta}(B) \ \underline{Z}_t \qquad (5\text{-}32)$$

where $\underline{X}_t$ is an (n x 1) column vector with elements X_t^i , i=1,...n , n is the number of sites and each element represent the observed process for site i ; $\underline{M}_t$ is an (n x 1) vector of iid normally distributed variables; $\underline{M}_t^*$ is an (n x 1) vector with elements $\underline{M}_t^{*(i)}$, each element is defined for each site i in the same manner as for the univariate case discussed

202

in chapter 3; $\underline{Z}_t$ is an (n x 1) vector of iid normally distributed variables representing the noise terms; $\underline{\Phi}(B)$ and $\Theta(B)$ are square matrices of polynomials in B which are respectively defined as

$$\underline{\Phi}(B) = I - \underline{\Phi}_1 B - \underline{\Phi}_1 B^2 - ... - \underline{\Phi}_p B^p$$

$$\underline{\Theta}(B) = I - \underline{\Theta}_1 B - \underline{\Theta}_1 B^2 - ... - \underline{\Theta}_q B^q$$

in which I is an (nxn) identity matrix; $\underline{\Phi}_j, j = 1,...,p$, $\underline{\Theta}_j, j = 1,...,q$, and $\underline{\Psi}$ are (nxn) appropriate matrices. For example, for the bivariate model, if site 1 is modeled by an ARMA (1,1) model and site 2 is modeled by a GNN model, the parameter matrices $\underline{\Phi}$, $\underline{\Theta}$, and $\underline{\Psi}$ can be written as:

$$\underline{\Phi} = \begin{bmatrix} \varphi & 0 \\ & \\ 0 & 0 \end{bmatrix} \tag{5-33}$$

and

$$\underline{\theta} = \begin{bmatrix} \theta & 0 \\ & \\ 0 & 0 \end{bmatrix} \tag{5-34}$$

$$\underline{\Psi} = \begin{bmatrix} 0 & 0 \\ & \\ 0 & 1 \end{bmatrix}$$

(5-35)

B^j is a scaler difference operator such that $B^j X_t = X_{t-1}$.

The model in Eq. (5-32) can be decoupled into univariate models for each site. For example, for the bivariate case if site 1 is modeled by an ARMA(1,1) and site with a GNN model then Eq. (5-32) can be written for site 1 as follows

$$X_t^{(1)} = \varphi X_{t-1}^{(1)} + Z_t^{(1)} - \theta Z_{t-1}^{(1)}$$

(5-36)

and for site 2

$$X_t^{(2)} = M_t^{*^{(2)}} + Z_t^{(2)}$$

(5-37)

As a result, the bivariate model can be written as :

$$\begin{bmatrix} X_t^{(1)} \\ \\ X_t^{(2)} \end{bmatrix} = \begin{bmatrix} \varphi & 0 \\ \\ 0 & 0 \end{bmatrix} \begin{bmatrix} X_{t-1}^{(1)} \\ \\ X_{t-1}^{(2)} \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ \\ 0 & 1 \end{bmatrix} \begin{bmatrix} M_t^{*^{(1)}} \\ \\ M_t^{*^{(2)}} \end{bmatrix} + \begin{bmatrix} Z_t^{(1)} \\ \\ Z_t^{(2)} \end{bmatrix} - \begin{bmatrix} \theta & 0 \\ \\ 0 & 0 \end{bmatrix} \begin{bmatrix} Z_{t-1}^{(1)} \\ \\ Z_{t-1}^{(2)} \end{bmatrix}$$

(5-38)

In deriving the moment equations for the GNN-CARMA(p,q) we will show the

204

derivation for the GNN-CARMA(1,1) model and then generalize the results for other

ARMA(p,q) models.

Multiplying both sides of Eq. (5-32) with $\underline{X}_t^T$ and taking expectation

$$E(\underline{X}_t\,\underline{X}_t^T) = \underline{\Phi}\,E(\underline{X}_{t-1}\,\underline{X}_t^T) + \underline{\Psi}\,E(\underline{M}_t^*\,\underline{X}_t^T) + E(\underline{Z}_t\,\underline{X}_t^T) - \underline{\Theta}\,E(\underline{Z}_{t-1}\,\underline{X}_t^T) \quad (5\text{-}39)$$

In terms of matrices S, C, and G defined in section 5.2.2. the right hand side of the above

equation are

$$E(\underline{X}_{t-1}\,\underline{X}_t^T) = S\,\underline{\Phi}^T - G\,\underline{\Theta}^T$$

$$E(\underline{M}_t^*\,\underline{X}_t^T) = C\,\underline{\Psi}^T$$

$$E(\underline{Z}_t\,\underline{X}_t^T) = G$$

$$E(\underline{Z}_{t-1}\,\underline{X}_t^T) = G\,\underline{\Phi}^T - G\,\underline{\Theta}^T$$

As a result Eq. (5-39) gives

$$S = \underline{\Phi}\,S\,\underline{\Phi}^T - \underline{\Phi}\,G\,\underline{\Theta}^T + \underline{\Psi}\,C\,\underline{\Psi}^T + G - \underline{\Theta}\,G\,\underline{\Phi}^T + \underline{\Theta}\,G\,\underline{\Theta}^T$$

The ij element in the S matrix can be written as:

$$S^{ij} = \varphi^i \, \varphi^j \, S^{ij} - \varphi^i \, \theta^j \, G^{ij} + \psi^i \, \psi^j \, C^{ij} + G^{ij} - \theta^i \, \varphi^j \, G^{ij} + \theta^i \, \theta^j \, G^{ij} \quad \text{(5-40)}$$

Note that the above equation applies for all sites in the model. For example for our bivariate model above, the above equation reduces to (keeping in mind that $\psi^1 = 0$ and $\varphi^2 = 0$, $\theta^2 = 0$, *and* $\psi^2 = 1$). Then

$$S^{11} = \varphi^1 \, \varphi^1 \, S^{11} - \varphi^1 \, \theta^1 \, G^{11} + G^{11} - \theta^1 \, \varphi^1 \, G^{11} + \theta^1 \, \theta^1 \, G^{11}$$

$$S^{22} = C^{22} + G^{22}$$

$$S^{12} = G^{12}$$

For the GNN-CARMA(1,0) model it can be shown that the moment equation is

$$S^{ij} = \varphi^i \, \varphi^j \, S^{ij} + \psi^i \, \psi^j \, C^{ij} + G^{ij} \quad \text{(5-41)}$$

Again considering our bivariate model and assuming that site 1 is modeled by an ARMA(1,0) and site 2 is modeled by a GNN model, then

$$S^{11} = \varphi^1 \, \varphi^1 \, S^{11} + G^{11}$$

$$S^{22} = C^{22} + G^{22}$$

$$S^{12} = G^{12}$$

Remember that Eq. (5-40) works for any number of sites in which each site is modeled univariately by an ARMA(1,1) or a GNN model and Eq. (5-41) works for the ARMA(1,0) case. Lets look at a more complicated example and see the corresponding equations. If 5 sites are modeled by the GNN-CARMA (1,1) where the first three sites are modeled by CARMA(1,1) and the last two are modeled by CGNN, the elements of the S matrix can be written as

$$S^{11} = \varphi^1 \varphi^1 S^{11} - \varphi^1 \theta^1 G^{11} + G^{11} - \theta^1 \varphi^1 G^{11} + \theta^1 \theta^1 G^{11}$$

$$S^{13} = \varphi^1 \varphi^3 S^{13} - \varphi^1 \theta^3 G^{13} + G^{13} - \theta^1 \varphi^3 G^{13} + \theta^1 \theta^3 G^{13}$$

$$S^{23} = \varphi^2 \varphi^3 S^{23} - \varphi^2 \theta^3 G^{23} + G^{23} - \theta^2 \varphi^3 G^{23} + \theta^2 \theta^3 G^{23}$$

$$S^{14} = G^{14}$$

$$S^{25} = G^{25}$$

$$S^{44} = C^{44} + G^{44}$$

$$S^{45} = C^{45} + G^{45}$$

From the above discussion, it can be concluded that the lag-0 variance-covariance matrix S of the GNN-CARMA (p,q) model is similar to the one obtained for the CARMA (p,q) model. The difference is only in the elements which correspond to the sites modeled univariately by the GNN model. The elements of the S matrix which correspond to the sites modeled univariately by ARMA (p,q) are the same for both the CARMA (p,q) model and the GNN-CARMA (p,q) models. The elements of the S matrix which correspond to the sites modeled by the CGNN model (i.e. univariately modeled by GNN) can be obtained by using the developed equations for the CGNN model.

Finally, the spatially correlated $\{Z\}$ process can be modeled by

$$\underline{Z}_t = \underline{B}\,\underline{\varepsilon}_t \tag{5-42}$$

where $\underline{\varepsilon}_t$ is an (n x 1) vector of standardized normal variables independent in both time and space and $\underline{B}$ is an (n x n) parameter matrix.

5.3.3 Parameter Estimation of the GNN-CARMA Model

The parameters for the GNN-CARMA (p,q) model can be estimated as follows:

1. For the sites modeled univariately by the ARMA (p,q) model, the parameters φ^i and θ^i can be estimated using the univarite models at each site i.

2. For the sites modeled by CGNN model, the parameters are estimated using the procedures presented in section 5.2.3.

3. The elements of the G matrix can be estimated by the moment equations of the model such as the one derived in Eq. (5-40) for the GNN-CARMA(1,1) model.

5.3.4 Application of the GNN-CARMA model to Observed Hydrologic Data

Nile River Basin

The GNN-CARMA (1,1) model was used to model the annual streamflows for Aswan, Khartoum, Roseires, and Malakal based on the data for the period 1914-1983. Malakal record shows a significant shift in its record which suggests that a GNN model may be useful in this case. Also it was shown in section 5.2.4 that the CARMA(1,1) and CARMA(2,1) did not perform well in preserving the rescaled range and Hurst coefficient for Malakal whereas GNN results were relatively better. As a result the GNN-CARMA(1,1) was used in which Aswan, Roseires, Khartoum flows were modeled by a CARMA (1,1) and the Malakal flows were modeled by GNN. The parameters of the CARMA(1,1) model were estimated by the LS Method and the parameters of GNN model for Malakal were estimated by using the regression estimation method. The estimated parameters are:

For Aswan: φ_1 = -0.26, θ_1 = -0.467

For Khartoum: φ_1 = -0.466, θ_1 = -0.675

For Roseires: φ_1 = -0.425, θ_1 = -0.686

For Malakal: p = 0.084, σ_m^2 = 27.13, σ_z^2 = 4.56

MATRIX-G

96.62	85.69	79.11	22.55
62.10	76.98	62.63	-2.06
60.60	66.82	62.01	-0.35
22.55	-2.06	-0.35	4.59

For generating annual flows for the sites, the following procedure was followed:

1. A vector of standard normal numbers $\varepsilon_t^{(i)}$, i=1,4 were generated. Then Eq. (5-42) was used to generate a vector of cross correlated normal random numbers $Z_t^{(i)}$, i=1,4.

2. Based on the estimated parameters Eq. (5-32) was used to generate a vector of autocorrelated and cross-correlated numbers $X_t^{(i)}$, i=1,4.

100 multivariate samples of length equal to the historical samples were generated using the above procedure. The results of the generation experiment are shown in Tables 5-13 and 5-14. The mean and the standard deviation are well preserved for all the four stations. In general, the GNN-CARMA (1,1) model was also capable of preserving the longest drought for all stations except for Aswan. The model also performed well in preserving the rescaled range and Hurst coefficient and did a better job than CARMA models in preserving the rescaled range for Malakal. Similar to CARMA(1,1) and CARMA(2,1) models, the GNN-CARMA (1,1) did perform well in preserving the maximum deficit except may be for Roseires. The GNN-CARMA(1,1) also performed relatively better than CARMA in preserving the Surplus as shown in Tables 5-3, 5-5, and 5-13. Table 5-14 show the historic and the generated lag-0 cross correlation between the sites studied which shows that the GNN-CARMA (1,1) model is able, in general, to preserve the spatial correlation structure that exists between the stations. Also, Figs. 5-28 through 5-31 show the generated Autocorrelograms for the GNN-CARMA(1,1) model. These figures suggest that the model performed well in preserving the correlogram. In summary, one can argue that the GNN-CARMA (1,1) is doing a good job in preserving

210

the basic short term statistics and the long term statistics on the univariate level as well as preserving the spatial correlation structure that exists between the studied stations.

By comparing the results of the GNN-CARMA(1,1) and CARMA(1,1) and CARMA(2,1), one can conclude that the GNN-CARMA model did a better job than the CARMA models in preserving the short and long term observed statistics for the studied sites. The strongest evidence of that is the ability of the GNN-CARMA model to preserve the rescaled range for St.Clair compared with the poor performance of the CARMA models in this regard. Remember that the GNN-CARMA(1,1) model has the same number of parameters as the CARMA(1,1) model and one less parameter than CARMA(2,1).

Great Lakes System

The GNN-CARMA (1,1) model was fitted to the annual (NBS) for the great lakes system for the period of 1900 - 1990. CARMA(1,1) model was used in modeling the data for lakes Superior, Michigan, Erie, and Ontario. St.Clair was modeled by a GNN model because lake St.Clair record shows a significant shift in its record which suggests that a GNN model may be useful. Also it was shown in section 5.2.4 that CARMA(1,1) and CARMA(2,1) did not perform well in preserving the rescaled range and Hurst coefficient for lake St.Clair and may be a GNN model will do a better job. The parameters of the CARMA (1,1) model were estimated by using the LS method and the parameters of the GNN model for St.Clair were estimated by using the regression method. The estimated parameters are:

For Superior: $\varphi_1 = 0.312$, $\theta_1 = 0.144$

For Mich-Huron: $\varphi_1 = 0.238$, $\theta_1 = 0.049$

For Erie: $\varphi_1 = 0.201$

For Ontario: $\varphi_1 = 0.794$, $\theta_1 = 0.559$

For St.Clair: $p = 0.053$, $\sigma_m^2 = 383.81$, $\sigma_z^2 = 338.30$

MATRIX-G

39926.80	32991.14	6615.02	7603.88	1355.48
33487.98	92816.79	16281.19	28031.06	3783.39
6725.51	16310.86	11309.09	10550.12	1607.92
4607.82	17289.59	6609.04	8223.74	1492.57
1355.48	3783.39	1607.92	1492.57	338.30

Following the same procedure as for the Nile River System 100 multivariate samples of length equal to the historical samples were generated. Tables 5-15 and 5-16 show the results of the generation experiment. As in the Nile example, the GNN-CARMA (1,1) model performed very well in preserving the mean and the standard deviation for each site. The Model did a good job in preserving the rescaled range at each site except may be for lake Erie. The GNN-CARMA (1,1) model did also perform well in preserving the longest drought at each site as shown in Table 5-15. Also, the GNN-CARMA (1,1) was also able to preserve the spatial correlation structure that exists in the Great Lakes system as shown in Table 5-28.

As mentioned above, the results for the rescaled range for lake Erie were not as good as other lakes. Alternatively a GNN-CARMA(1,1) was used to model the lakes

system in which lakes St.Clair and lake Erie are modeled by a CGNN model and lakes Superior, Michigan, and Ontario are modeled by CARMA(1,1) model. Tables 5-17 and 5-18 show the results of the generation experiment and Figs. 5-32 through 5-36 show the generated Autocorrelograms. The model performance is very good. The model was able to preserve the rescaled range for Lake Erie as well as performing well in preserving the other statistics. This confirms the finding in the Nile example which suggests that the GNN-CARMA should be the model to use instead of CARMA(1,1) or CARMA(2,1) in the Nile and the Great Lakes basins.

Table 5-1 :   Results of generation experiment for the Nile River System based on
               CGNN model.

SITE	HISTORICAL	GEN-STD	GENERATED	GEN+STD
MEAN (milliards of cubic meters)				
ASWAN	84.95	82.10	84.98	87.86
KHARTOUM	52.16	50.25	52.21	54.18
ROSEIRES	49.92	47.93	50.00	51.99
MALAKAL	30.01	27.32	30.09	32.86
STANDARD DEVIATION (milliards of cubic meters)				
ASWAN	10.17	8.83	9.90	10.97
KHARTOUM	9.02	7.97	8.86	9.75
ROSEIRES	8.19	7.21	8.03	8.85
MALAKAL	5.63	3.37	4.87	6.38
LONGEST DROUGHT (Yeras)				
ASWAN	11.00	4.88	7.64	10.40
KHARTOUM	7.00	4.43	6.52	8.61
ROSEIRES	6.00	4.68	6.69	8.70
MALAKAL	14.00	8.28	15.75	23.22
RESCALED RANGE				
ASWAN	11.70	9.97	13.97	17.97
KHARTOUM	13.26	9.07	12.35	15.64
ROSEIRES	12.53	9.70	13.13	16.57
MALAKAL	25.21	17.61	22.93	28.25
HURST COEFICIENT				
ASWAN	0.692	0.645	0.730	0.813
KHARTOUM	0.727	0.619	0.697	0.774
ROSEIRES	0.711	0.636	0.714	0.792
MALAKAL	0.908	0.801	0.873	0.944

Table 5-1 :   Cont'd.

SITE	HISTORICAL	GEN-STD	GENERATED	GEN+STD
MAXIMUM DEFICIT (milliards of cubic meters)				
ASWAN	65.19	37.39	74.70	112.02
KHARTOUM	52.75	36.30	54.59	72.88
ROSEIRES	61.98	33.08	52.74	72.40
MALAKAL	55.96	28.08	87.47	146.87
SURPLUS (milliards of cubic meters)				
ASWAN	118.99	79.07	128.87	178.66
KHARTOUM	119.58	64.43	100.16	135.89
ROSEIRES	102.64	61.97	96.26	130.56
MALAKAL	124.05	54.97	108.53	162.09

Table 5-2: Historical and generated lag-0 cross correlation coeficients for the Nile River System based on CGNN model. Note that Values in paranthesis are the generated ones.

	ASWAN	KHARTOUM	ROSEIRS	MALAKAL
ASWAN	1.000 (1.000)	0.809 (0.841)	0.863 (0.876)	0.394 (0.328)
KHARTOUM		1.000 (1.000)	0.935 (0.939)	-0.041 (-0.027)
ROSEIRES			1.000 (1.000)	-0.008 (0.004)
MALAKAL				1.000 (1.000)

Table 5-3 :   Results of generation experiment for the Nile River System based on
               CARMA(1,1) model.

SITE	HISTORICAL	GEN-STD	GENERATED	GEN+STD
MEAN (milliards of cubic meters)				
ASWAN	84.95	83.32	84.88	86.44
KHARTOUM	52.16	50.99	52.10	53.22
ROSEIRES	49.92	48.77	49.86	50.95
MALAKAL	30.01	28.22	29.97	31.72
STANDARD DEVIATION (milliards of cubic meters)				
ASWAN	10.17	9.98	10.89	11.79
KHARTOUM	9.02	7.72	8.52	9.33
ROSEIRES	8.19	7.27	8.01	8.75
MALAKAL	5.63	5.21	6.02	6.83
LONGEST DROUGHT (Yeras)				
ASWAN	11.00	4.45	6.42	8.39
KHARTOUM	7.00	4.18	6.04	7.90
ROSEIRES	6.00	4.12	6.12	8.12
MALAKAL	14.00	7.99	12.16	16.33
RESCALED RANGE				
ASWAN	11.70	8.42	10.79	13.17
KHARTOUM	13.26	7.87	10.17	12.48
ROSEIRES	12.53	7.83	10.23	12.63
MALAKAL	25.21	13.19	17.49	21.79
HURST COEFICIENT				
ASWAN	0.692	0.600	0.662	0.725
KHARTOUM	0.727	0.583	0.645	0.708
ROSEIRES	0.711	0.580	0.646	0.712
MALAKAL	0.908	0.727	0.796	0.866

Table 5-3 :   Cont'd.

SITE	HISTORICAL	GEN-STD	GENERATED	GEN+STD
MAXIMUM DEFICIT (milliards of cubic meters)				
ASWAN	65.19	46.09	65.87	85.66
KHARTOUM	52.75	32.49	47.21	61.94
ROSEIRES	61.98	31.74	47.05	62.36
MALAKAL	55.96	44.84	76.23	107.61
SURPLUS (milliards of cubic meters)				
ASWAN	118.99	76.84	106.25	135.67
KHARTOUM	119.58	57.59	78.84	100.08
ROSEIRES	102.64	54.75	75.62	96.50
MALAKAL	124.05	60.66	93.71	126.75

Table 5-4: Historical and generated lag-0 cross correlation coeficients for the Nile River System based on CARMA(1,1) model. Note that Values in paranthesis are the generated ones.

	ASWAN	KHARTOUM	ROSEIRS	MALAKAL
ASWAN	1.000 (1.000)	0.809 (0.829)	0.863 (0.843)	0.394 (0.267)
KHARTOUM		1.000 (1.000)	0.935 (0.931)	-0.041 (-0.012)
ROSEIRES			1.000 (1.000)	-0.008 (0.014)
MALAKAL				1.000 (1.000)

Table 5-5:   Results of generation experiment for the Nile River System based on
              CARMA(2,1) model.

SITE	HISTORICAL	GEN-STD	GENERATED	GEN+STD
MEAN (milliards of cubic meters)				
ASWAN	84.95	83.37	84.87	86.37
KHARTOUM	52.16	51.04	52.11	53.21
ROSEIRES	49.92	48.85	49.87	50.89
MALAKAL	30.01	28.18	29.98	31.78
STANDARD DEVIATION (milliards of cubic meters)				
ASWAN	10.17	9.29	10.21	11.13
KHARTOUM	9.02	7.94	8.75	9.56
ROSEIRES	8.19	7.46	8.25	9.03
MALAKAL	5.63	4.89	5.70	6.52
LONGEST DROUGHT (Yeras)				
ASWAN	11.00	4.14	6.05	7.97
KHARTOUM	7.00	4.18	5.84	7.50
ROSEIRES	6.00	4.23	5.97	7.71
MALAKAL	14.00	7.85	12.07	16.29
RESCALED RANGE				
ASWAN	11.70	8.45	10.65	12.85
KHARTOUM	13.26	7.63	9.88	12.14
ROSEIRES	12.53	7.74	10.01	12.27
MALAKAL	25.21	13.57	17.98	22.40
HURST COEFICIENT				
ASWAN	0.692	0.601	0.659	0.718
KHARTOUM	0.727	0.575	0.637	0.700
ROSEIRES	0.711	0.577	0.641	0.704
MALAKAL	0.908	0.736	0.804	0.873

Table 5-5 :   Cont'd.

SITE	HISTORICAL	GEN-STD	GENERATED	GEN+STD
MAXIMUM DEFICIT (milliards of cubic meters)				
ASWAN	65.19	42.91	61.74	80.57
KHARTOUM	52.75	34.32	47.86	61.40
ROSEIRES	61.98	32.00	45.00	57.99
MALAKAL	55.96	40.56	74.20	107.84
SURPLUS (milliards of cubic meters)				
ASWAN	118.99	71.48	98.68	125.88
KHARTOUM	119.58	57.98	79.18	100.39
ROSEIRES	102.64	54.25	74.31	94.36
MALAKAL	124.05	58.31	91.94	125.57

Table 5-6: Historical and generated lag-0 cross correlation coeficients for the Nile River System based on CARMA(2,1) model. Note that Values in paranthesis are the generated ones.

	ASWAN	KHARTOUM	ROSEIRS	MALAKAL
ASWAN	1.000 (1.000)	0.809 (0.801)	0.863 (0.834)	0.394 (0.318)
KHARTOUM		1.000 (1.000)	0.935 (0.934)	-0.041 (-0.053)
ROSEIRES			1.000 (1.000)	-0.008 (-0.005)
MALAKAL				1.000 (1.000)

Table 5-7:   Results of generation experiment for the Great Lakes System based on CGNN model.

SITE	HISTORICAL	GEN-STD	GENERATED	GEN+STD
MEAN (thousands cfs)				
SUPERIOR	870.14	855.43	871.71	907.99
MICH-HURON	1344.26	1278.61	1338.81	1399.01
ST. CLAIR	51.71	49.94	52.97	59.99
ERIE	241.77	221.81	243.06	264.31
ONTARIO	432.56	412.57	434.82	457.06
STANDARD DEVIATION (thousands cfs)				
SUPERIOR	202.90	181.17	195.92	210.67
MICH-HURON	310.38	276.66	302.55	328.43
ST. CLAIR	26.87	20.53	24.62	28.72
ERIE	108.57	97.49	106.61	115.73
ONTARIO	97.22	85.12	95.50	105.88
LONGEST DROUGHT (Years)				
SUPERIOR	5.00	4.32	7.34	10.36
MICH-HURON	8.00	4.86	7.30	9.74
ST. CLAIR	9.00	6.98	11.29	15.60
ERIE	8.00	4.52	7.27	10.02
ONTARIO	10.00	4.92	8.37	11.82
RESCALED RANGE				
SUPERIOR	11.75	11.24	15.29	19.34
MICH-HURON	16.14	11.07	15.12	19.18
ST. CLAIR	25.04	15.93	21.49	27.05
ERIE	18.30	12.05	16.26	20.48
ONTARIO	17.49	13.08	18.00	22.92
HURST COEFICIENT				
SUPERIOR	0.647	0.634	0.707	0.779
MICH-HURON	0.731	0.629	0.704	0.778
ST. CLAIR	0.846	0.724	0.796	0.869
ERIE	0.764	0.653	0.723	0.794
ONTARIO	0.752	0.676	0.749	0.823

Table 5-7:   Cont'd.

SITE	HISTORICAL	GEN-STD	GENERATED	GEN+STD
MAXIMUM DEFICIT (thousands cfs)				
SUPERIOR	1653.72	759.20	1426.63	2094.06
MICH-HURON	2540.05	1350.72	2215.99	3081.27
ST. CLAIR	279.40	134.96	313.99	493.03
ERIE	681.07	428.56	786.15	1143.74
ONTARIO	777.56	439.12	853.08	1267.03
SURPLUS (thousands cfs)				
SUPERIOR	2283.72	1848.96	2800.33	3751.69
MICH-HURON	3950.38	2842.44	4169.62	5469.80
ST. CLAIR	672.84	299.44	489.75	680.06
ERIE	1986.67	1072.62	1574.45	2076.28
ONTARIO	1700.00	985.77	1565.79	2145.82

Table 5-8: Historical and generated lag-0 cross correlation coeficients for the Great Lakes System based on CGNN model. Note that Values in paranthesis are the generated ones.

	SUPERIOR	MICH-HURON	ST.CLAIR	ERIE	ONTARIO
SUPERIOR	1.000 (1.000)	0.544 (0.523)	0.249 (0.232)	0.311 (0.312)	0.267 (0.249)
MICH-HURON		1.000 (1.000)	0.454 (0.454)	0.502 (0.485)	0.624 (0.623)
ST.CLAIR			1.000 (1.000)	0.551 (0.496)	0.571 (0.564)
ERIE				1.000 (1.000)	0.661 (0.614)
ONTARIO					1.000 (1.000)

Table 5-9:   Results of generation experiment for the Great Lakes System based on CARMA(1,1) model.

SITE	HISTORICAL	GEN-STD	GENERATED	GEN+STD
MEAN (thousands cfs)				
SUPERIOR	870.14	843.06	868.56	894.05
MICH-HURON	1344.26	1292.55	1338.71	1384.87
ST. CLAIR	51.71	46.12	52.44	58.77
ERIE	241.77	226.14	241.13	256.12
ONTARIO	432.56	412.62	431.13	449.65
STANDARD DEVIATION (thousands cfs)				
SUPERIOR	202.90	185.52	201.95	218.39
MICH-HURON	310.38	302.02	328.20	354.38
ST. CLAIR	26.87	22.93	25.47	28.01
ERIE	108.57	104.48	114.47	124.45
ONTARIO	97.22	78.12	86.11	94.10
LONGEST DROUGHT (Years)				
SUPERIOR	5.00	4.60	6.20	7.80
MICH-HURON	8.00	5.10	7.10	9.10
ST. CLAIR	9.00	6.82	9.84	12.86
ERIE	8.00	4.79	6.87	8.95
ONTARIO	10.00	6.00	8.71	11.42
RESCALED RANGE				
SUPERIOR	11.75	9.75	12.50	15.25
MICH-HURON	16.14	9.39	12.09	14.79
ST. CLAIR	25.04	14.53	18.69	22.85
ERIE	18.30	9.84	12.73	15.62
ONTARIO	17.49	12.52	16.52	20.52
HURST COEFICIENT				
SUPERIOR	0.647	0.601	0.658	0.714
MICH-HURON	0.731	0.591	0.648	0.706
ST. CLAIR	0.846	0.704	0.763	0.821
ERIE	0.764	0.600	0.661	0.723
ONTARIO	0.752	0.663	0.729	0.795

Table 5-9:   Cont'd.

SITE	HISTORICAL	GEN-STD	GENERATED	GEN+STD
MAXIMUM DEFICIT (thousands cfs)				
SUPERIOR	1653.72	903.75	1240.20	1576.65
MICH-HURON	2540.05	1545.72	2306.98	3068.47
ST. CLAIR	279.40	169.54	263.60	357.66
ERIE	681.07	473.10	752.18	1031.26
ONTARIO	777.56	460.34	773.25	1086.16
SURPLUS (thousands cfs)				
SUPERIOR	2283.72	1644.75	2278.40	2912.06
MICH-HURON	3950.38	2686.87	3561.71	4436.56
ST. CLAIR	672.84	285.15	427.18	569.22
ERIE	1986.67	914.36	1307.56	1700.75
ONTARIO	1700.00	906.57	1282.29	1658.01

Table 5-10: Historical and generated lag-0 cross correlation coeficients for the Great Lakes System based on CARMA(1,1) model. Note that Values in paranthesis are the generated ones.

	SUPERIOR	MICH-HURON	ST.CLAIR	ERIE	ONTARIO
SUPERIOR	1.000 (1.000)	0.544 (0.552)	0.249 (0.239)	0.311 (0.309)	0.267 (0.307)
MICH-HURON		1.000 (1.000)	0.454 (0.477)	0.502 (0.553)	0.624 (0.768)
ST.CLAIR			1.000 (1.000)	0.551 (0.615)	0.571 (0.535)
ERIE				1.000 (1.000)	0.661 (0.822)
ONTARIO					1.000 (1.000)

Table 5-11:   Results of generation experiment for the Great Lakes System based on CARMA(2,1) model.

SITE	HISTORICAL	GEN-STD	GENERATED	GEN+STD
MEAN (thousands cfs)				
SUPERIOR	870.14	840.72	868.59	896.42
MICH-HURON	1344.26	1294.27	1338.12	1381.98
ST. CLAIR	51.71	46.13	51.94	57.76
ERIE	241.77	226.38	242.03	257.68
ONTARIO	432.56	411.15	431.27	451.39
STANDARD DEVIATION (thousands cfs)				
SUPERIOR	202.90	186.65	203.00	219.34
MICH-HURON	310.38	295.73	323.44	351.14
ST. CLAIR	26.87	22.25	25.02	27.78
ERIE	108.57	108.32	117.87	127.43
ONTARIO	97.22	78.61	86.87	95.14
LONGEST DROUGHT (Years)				
SUPERIOR	5.00	4.64	6.56	8.49
MICH-HURON	8.00	4.95	6.98	9.01
ST. CLAIR	9.00	6.38	9.82	13.26
ERIE	8.00	4.65	6.49	8.33
ONTARIO	10.00	5.51	8.25	10.99
RESCALED RANGE				
SUPERIOR	11.75	10.02	12.80	15.59
MICH-HURON	16.14	9.73	12.60	15.48
ST. CLAIR	25.04	13.80	18.04	22.28
ERIE	18.30	9.09	12.26	15.43
ONTARIO	17.49	12.47	16.62	20.77
HURST COEFICIENT				
SUPERIOR	0.647	0.607	0.664	0.720
MICH-HURON	0.731	0.599	0.659	0.719
ST. CLAIR	0.846	0.689	0.752	0.816
ERIE	0.764	0.582	0.650	0.717
ONTARIO	0.752	0.664	0.730	0.796

Table 5-11:   Cont'd.

SITE	HISTORICAL	GEN-STD	GENERATED	GEN+STD
MAXIMUM DEFICIT (thousands cfs)				
SUPERIOR	1653.72	947.38	1372.35	1797.32
MICH-HURON	2540.05	1458.28	2162.17	2866.06
ST. CLAIR	279.40	140.22	259.29	378.36
ERIE	681.07	547.14	773.00	998.86
ONTARIO	777.56	470.58	761.55	1052.52
SURPLUS (thousands cfs)				
SUPERIOR	2283.72	1712.29	2344.14	2976.00
MICH-HURON	3950.38	2565.87	3638.38	4711.45
ST. CLAIR	672.84	273.58	404.49	535.41
ERIE	1986.67	889.80	1288.50	1687.21
ONTARIO	1700.00	866.10	1310.76	1755.41

Table 5-12: Historical and generated lag-0 cross correlation coeficients for the Great Lakes System based on CARMA(2,1) model. Note that Values in paranthesis are the generated ones.

	SUPERIOR	MICH-HURON	ST.CLAIR	ERIE	ONTARIO
SUPERIOR	1.000 (1.000)	0.544 (0.545)	0.249 (0.334)	0.311 (0.318)	0.267 (0.350)
MICH-HURON		1.000 (1.000)	0.454 (0.544)	0.502 (0.567)	0.624 (0.819)
ST.CLAIR			1.000 (1.000)	0.551 (0.636)	0.571 (0.509)
ERIE				1.000 (1.000)	0.661 (0.819)
ONTARIO					1.000 (1.000)

231

Table 5-13: Results of generation experiment for the Nile River System based on GNN-CARMA(1,1) model.

SITE	HISTORICAL	GEN-STD	GENERATED	GEN+STD
MEAN (milliards of cubic meters)				
ASWAN	84.95	83.08	84.88	86.69
KHARTOUM	52.16	50.93	52.10	53.28
ROSEIRES	49.92	48.75	49.86	50.97
MALAKAL	30.01	27.47	30.09	32.71
STANDARD DEVIATION (milliards of cubic meters)				
ASWAN	10.17	11.59	12.58	13.57
KHARTOUM	9.02	8.11	8.95	9.78
ROSEIRES	8.19	7.50	8.25	9.00
MALAKAL	5.63	3.43	4.67	5.92
LONGEST DROUGHT (Yeras)				
ASWAN	11.00	4.32	6.05	7.78
KHARTOUM	7.00	4.45	6.05	7.65
ROSEIRES	6.00	4.38	6.15	7.92
MALAKAL	14.00	8.43	14.49	20.86
RESCALED RANGE				
ASWAN	11.70	8.44	10.83	13.22
KHARTOUM	13.26	7.70	10.05	12.41
ROSEIRES	12.53	7.67	10.19	12.70
MALAKAL	25.21	17.49	22.17	26.86
HURST COEFICIENT				
ASWAN	0.692	0.600	0.663	0.726
KHARTOUM	0.727	0.577	0.642	0.706
ROSEIRES	0.711	0.576	0.644	0.713
MALAKAL	0.908	0.800	0.865	0.929

Table 5-13:   Cont'd.

SITE	HISTORICAL	GEN-STD	GENERATED	GEN+STD
MAXIMUM DEFICIT (milliards of cubic meters)				
ASWAN	65.19	52.27	72.99	93.72
KHARTOUM	52.75	36.21	49.12	62.02
ROSEIRES	61.98	33.52	46.99	60.46
MALAKAL	55.96	32.02	75.85	119.68
SURPLUS (milliards of cubic meters)				
ASWAN	118.99	90.50	123.32	156.14
KHARTOUM	119.58	59.94	82.56	105.17
ROSEIRES	102.64	55.72	77.78	99.84
MALAKAL	124.05	57.63	101.54	145.44

Table 5-14: Historical and generated lag-0 cross correlation coeficients for the Nile River System based on GNN-CARMA(1,1) model. Note that Values in paranthesis are the generated ones.

	ASWAN	KHARTOUM	ROSEIRS	MALAKAL
ASWAN	1.000 (1.000)	0.809 (0.540)	0.863 (0.596)	0.394 (0.228)
KHARTOUM		1.000 (1.000)	0.935 (0.935)	-0.041 (-0.027)
ROSEIRES			1.000 (1.000)	-0.008 (0.008)
MALAKAL				1.000 (1.000)

Table 5-15: Results of generation experiment for the Great Lakes System based on GNN-CARMA(1,1) model in which Lakes Superior, Michigan-Huron, and Ontario are modeled univariately by ARMA(1,1), lake Erie by ARMA(1,0), and lake St. Clair by GNN.

SITE	HISTORICAL	GEN-STD	GENERATED	GEN+STD
MEAN (thousands cfs)				
SUPERIOR	870.14	842.48	868.84	896.20
MICH-HURON	1344.26	1292.08	1336.99	1381.80
ST. CLAIR	51.71	42.29	52.76	63.24
ERIE	241.77	226.36	241.29	256.22
ONTARIO	432.56	412.27	432.50	452.74
STANDARD DEVIATION (thousands cfs)				
SUPERIOR	202.90	188.36	203.93	219.49
MICH-HURON	310.38	295.57	321.99	348.41
ST. CLAIR	26.87	19.86	24.07	28.28
ERIE	108.57	104.83	114.29	123.74
ONTARIO	97.22	80.98	89.37	97.76
LONGEST DROUGHT (Years)				
SUPERIOR	5.00	4.75	6.88	9.01
MICH-HURON	8.00	4.95	6.98	9.01
ST. CLAIR	9.00	5.90	10.81	15.72
ERIE	8.00	5.27	6.92	8.57
ONTARIO	10.00	5.24	8.14	11.04
RESCALED RANGE				
SUPERIOR	11.75	9.28	12.32	15.35
MICH-HURON	16.14	9.73	12.60	15.48
ST. CLAIR	25.04	15.97	22.59	29.20
ERIE	18.30	9.48	12.49	15.51
ONTARIO	17.49	12.25	16.36	20.46
HURST COEFICIENT				
SUPERIOR	0.647	0.603	0.660	0.717
MICH-HURON	0.731	0.588	0.652	0.716
ST. CLAIR	0.846	0.721	0.806	0.891
ERIE	0.764	0.594	0.656	0.718
ONTARIO	0.752	0.659	0.726	0.792

Table 5-16: Historical and generated lag-0 cross correlation coeficients for the Great Lakes System based on GNN-CARMA(1,1) model in which Lakes Superior, Michigan-Huron, and Ontario are modeled univariately by ARMA(1,1), lake Erie by ARMA(1,0), and lake St. Clair by GNN. Note that Values in paranthesis are the generated ones.

	SUPERIOR	MICH-HURON	ST.CLAIR	ERIE	ONTARIO
SUPERIOR	1.000 (1.000)	0.544 (0.532)	0.249 (0.297)	0.311 (0.315)	0.267 (0.279)
MICH-HURON		1.000 (1.000)	0.454 (0.492)	0.502 (0.549)	0.624 (0.697)
ST.CLAIR			1.000 (1.000)	0.551 (0.610)	0.571 (0.621)
ERIE				1.000 (1.000)	0.661 (0.772)
ONTARIO					1.000 (1.000)

Table 5-17:   Results of generation experiment for the Great Lakes System based on GNN-CARMA(1,1) model in which Lakes Superior, Michigan-Huron, and Ontario are modeled univariately by ARMA(1,1), lakes Erie and and St. Clair by GNN.

SITE	HISTORICAL	GEN-STD	GENERATED	GEN+STD
MEAN (thousands cfs)				
SUPERIOR	870.14	840.59	868.74	892.89
MICH-HURON	1344.26	1299.24	1343.76	1388.27
ST. CLAIR	51.71	40.85	53.16	65.47
ERIE	241.77	215.68	249.10	282.52
ONTARIO	432.56	411.16	431.60	452.04
STANDARD DEVIATION (thousands cfs)				
SUPERIOR	202.90	185.23	201.98	218.73
MICH-HURON	310.38	304.02	329.81	355.59
ST. CLAIR	26.87	18.94	23.72	28.50
ERIE	108.57	92.81	103.47	114.13
ONTARIO	97.22	84.96	94.41	103.86
LONGEST DROUGHT (Years)				
SUPERIOR	5.00	4.77	6.73	8.68
MICH-HURON	8.00	4.81	6.73	8.65
ST. CLAIR	9.00	5.06	10.88	16.71
ERIE	8.00	4.44	7.71	10.98
ONTARIO	10.00	4.87	8.62	12.37
RESCALED RANGE				
SUPERIOR	11.75	9.50	12.34	15.17
MICH-HURON	16.14	9.77	12.63	15.49
ST. CLAIR	25.04	15.11	22.27	29.43
ERIE	18.30	10.81	16.19	21.57
ONTARIO	17.49	12.43	16.76	21.08
HURST COEFICIENT				
SUPERIOR	0.647	0.598	0.654	0.710
MICH-HURON	0.731	0.599	0.659	0.720
ST. CLAIR	0.846	0.705	0.800	0.894
ERIE	0.764	0.633	0.718	0.802
ONTARIO	0.752	0.662	0.732	0.801

Table 5-17:   Cont'd.

SITE	HISTORICAL	GEN-STD	GENERATED	GEN+STD
MAXIMUM DEFICIT (thousands cfs)				
SUPERIOR	1653.72	861.26	1334.60	1807.94
MICH-HURON	2540.05	1439.85	2217.12	2994.39
ST. CLAIR	279.40	64.21	289.09	513.97
ERIE	681.07	305.46	757.99	1210.52
ONTARIO	777.56	450.71	843.01	1235.32
SURPLUS (thousands cfs)				
SUPERIOR	2283.72	1641.56	2265.98	2890.40
MICH-HURON	3950.38	2724.90	3638.67	4642.44
ST. CLAIR	672.84	262.38	524.20	786.02
ERIE	1986.67	868.76	1599.96	2331.16
ONTARIO	1700.00	931.83	1437.74	1943.65

Table 5-18: Historical and generated lag-0 cross correlation coeficients for the Great Lakes System based on GNN-CARMA(1,1) model in which Lakes Superior, Michigan-Huron, and Ontario are modeled univariately by ARMA(1,1), lakes Erie and and St. Clair by GNN. Note that Values in paranthesis are the generated ones.

	SUPERIOR	MICH-HURON	ST.CLAIR	ERIE	ONTARIO
SUPERIOR	1.000 (1.000)	0.544 (0.519)	0.249 (0.310)	0.311 (0.324)	0.267 (0.257)
MICH-HURON		1.000 (1.000)	0.454 (0.565)	0.502 (0.552)	0.624 (0.667)
ST.CLAIR			1.000 (1.000)	0.551 (0.511)	0.571 (0.622)
ERIE				1.000 (1.000)	0.661 (0.659)
ONTARIO					1.000 (1.000)

239
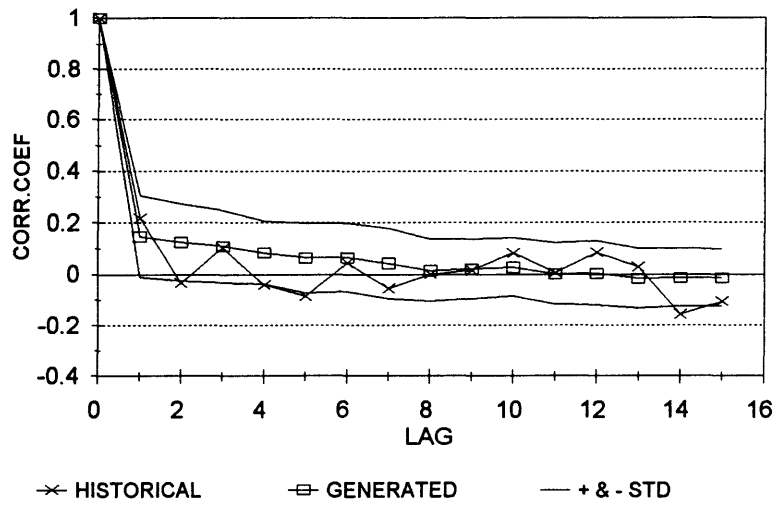
Fig 5-1   Historic and generated autocorrelogram for Aswan based on CGNN Model.
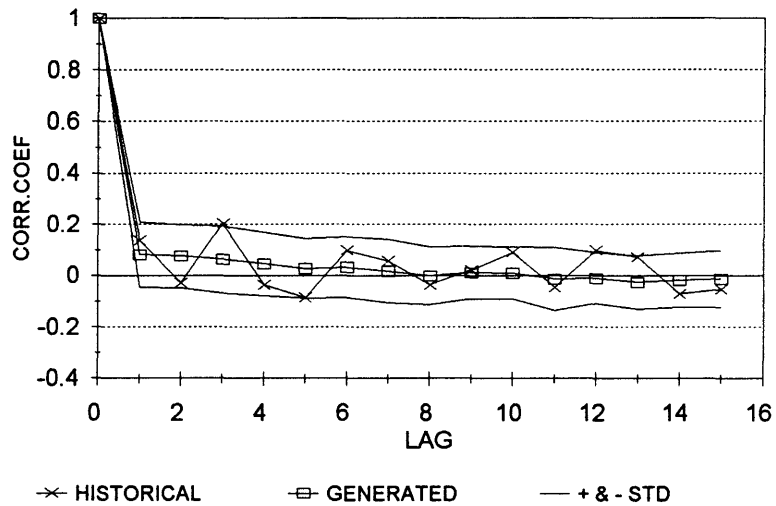
Fig 5-2   Historic and generated autocorrelogram for Khartoum based on CGNN Model.

240
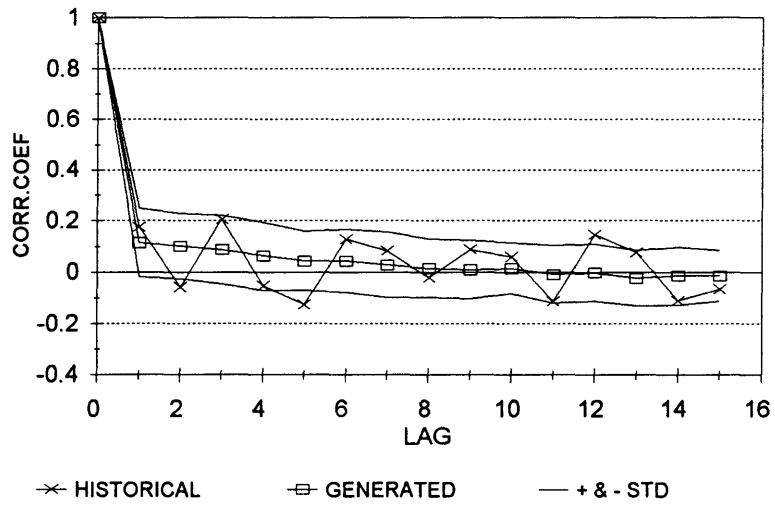
Fig 5-3   Historic and generated autocorrelogram for Roseires based on CGNN
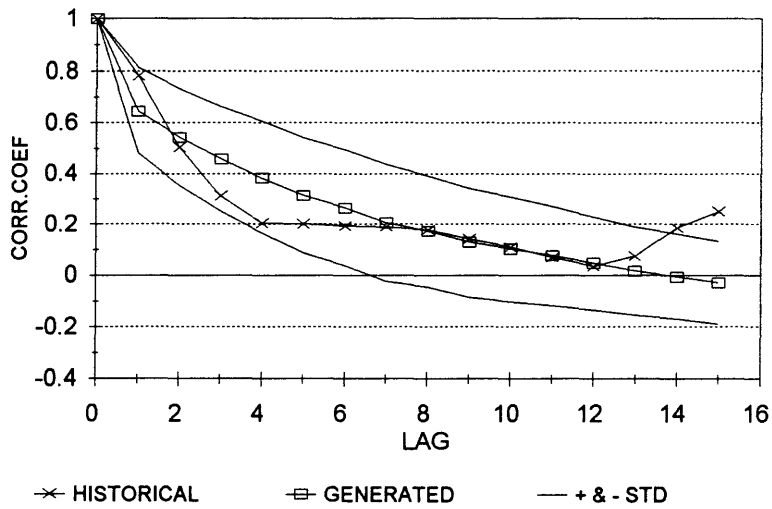   Model.

Fig 5-4   Historic and generated autocorrelogram for Malakal based on CGNN
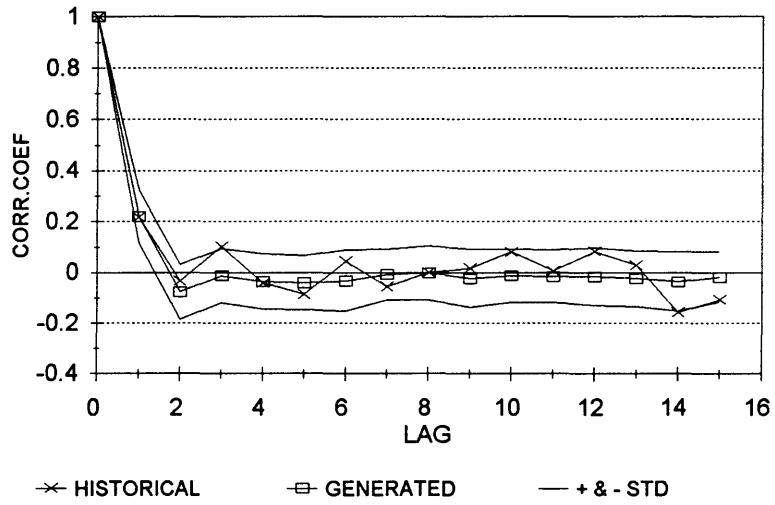   Model.

Fig 5-5   Historic and generated autocorrelogram for Aswan based on
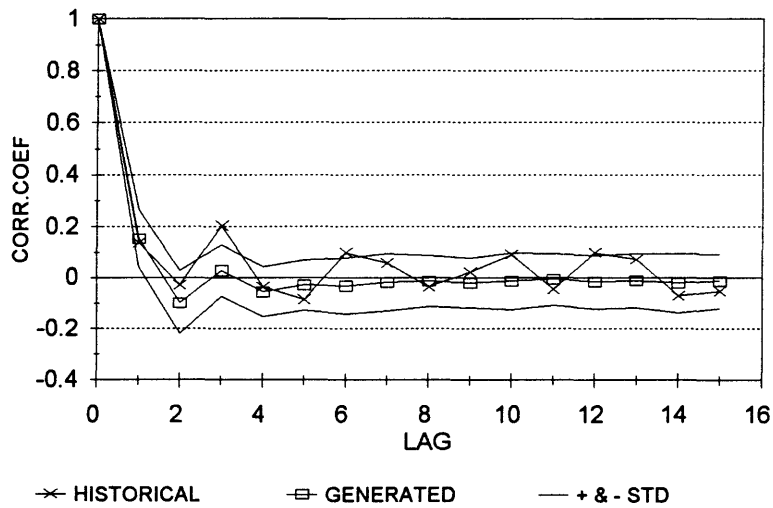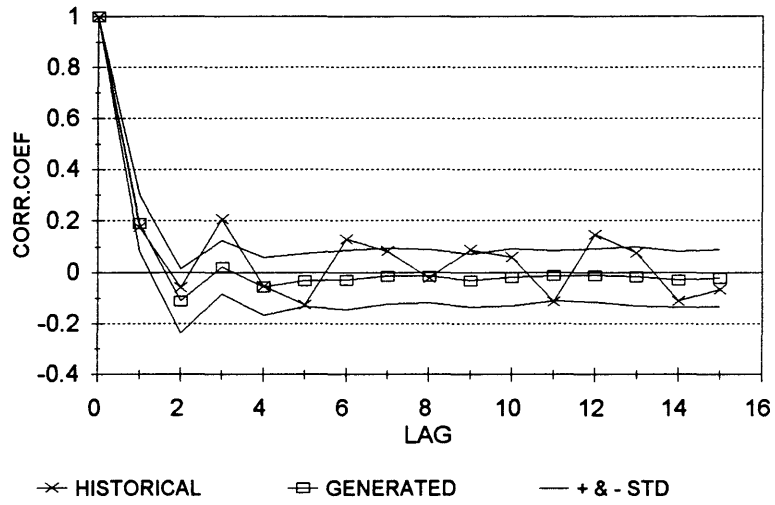   CARMA(1,1) Model.

Fig 5-6   Historic and generated autocorrelogram for Khartoum based on
           CARMA(1,1) Model.

242

Fig 5-7   Historic and generated autocorrelogram for Roseires based on
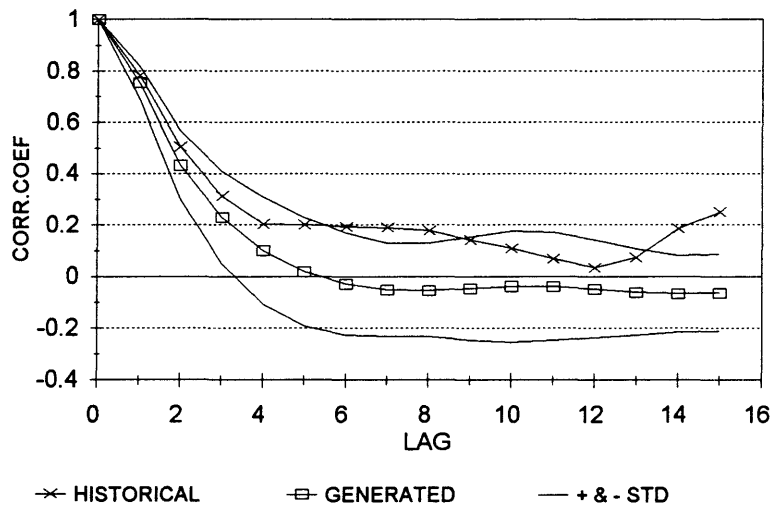CARMA(1,1) Model.

Fig 5-8   Historic and generated autocorrelogram for Malakal based on
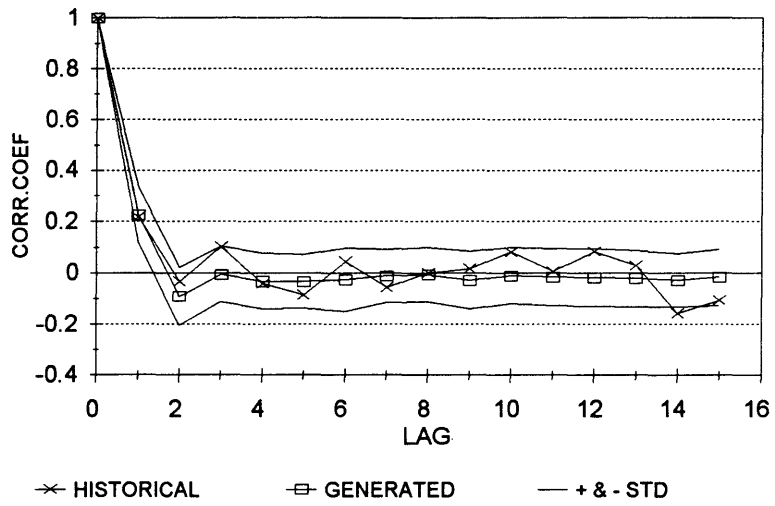CARMA(1,1) Model.

Fig 5-9   Historic and generated autocorrelogram for Aswan based on
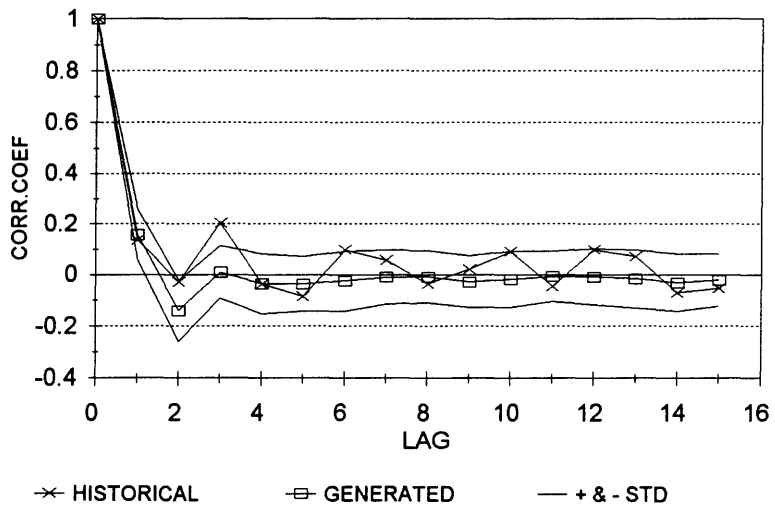   CARMA(2,1) Model.

Fig 5-10   Historic and generated autocorrelogram for Khartoum based on
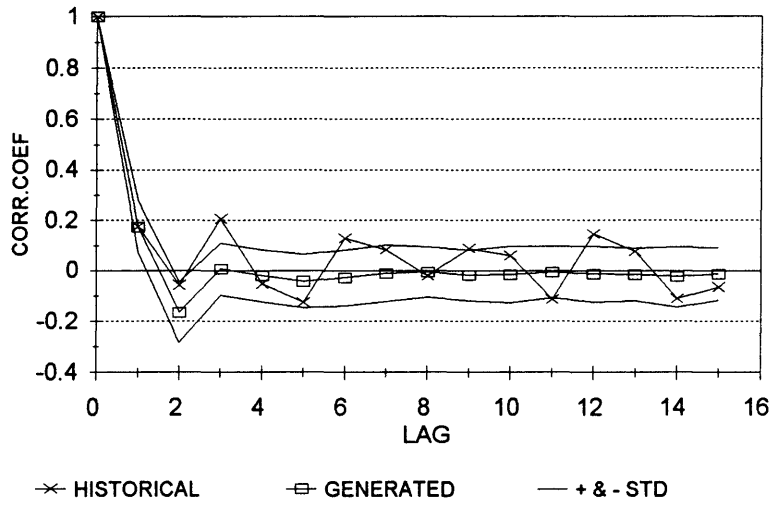   CARMA(2,1) Model.

Fig 5-11   Historic and generated autocorrelogram for Roseires based on
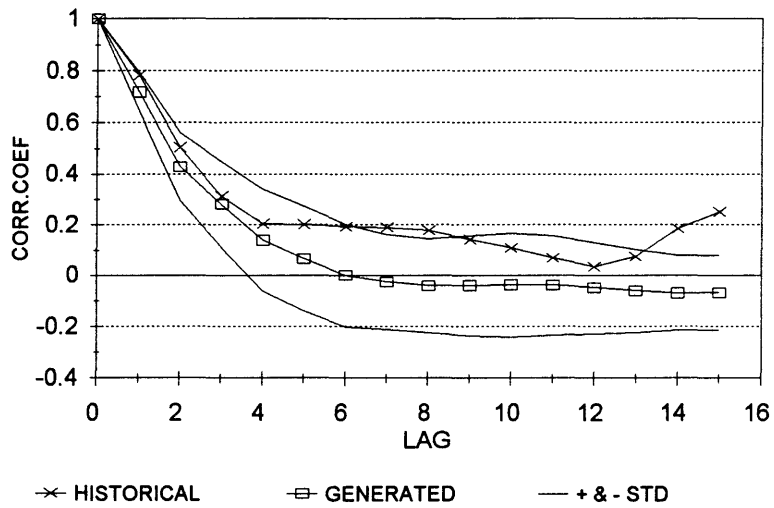CARMA(2,1) Model.

Fig 5-12   Historic and generated autocorrelogram for Malakal based on
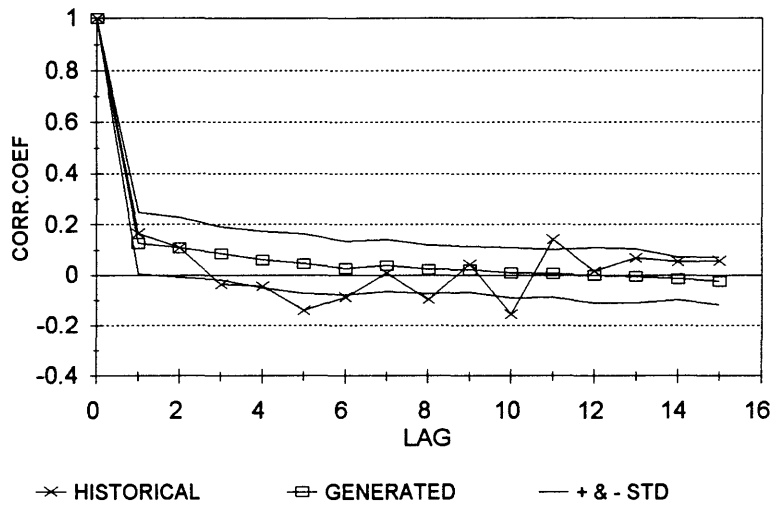CARMA(2,1) Model.

245

Fig 5-13   Historic and generated autocorrelogram for lake Superior based on CGNN model.
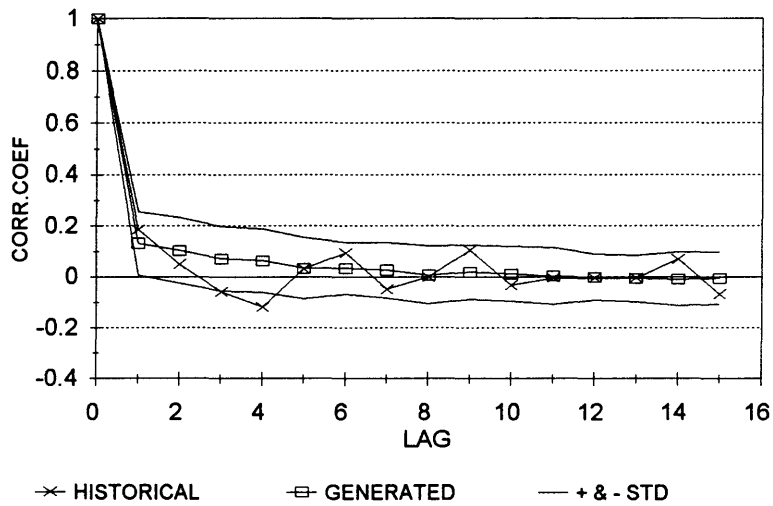

Fig 5-14   Historic and generated autocorrelogram for lake Michigan-Huron based on CGNN model.
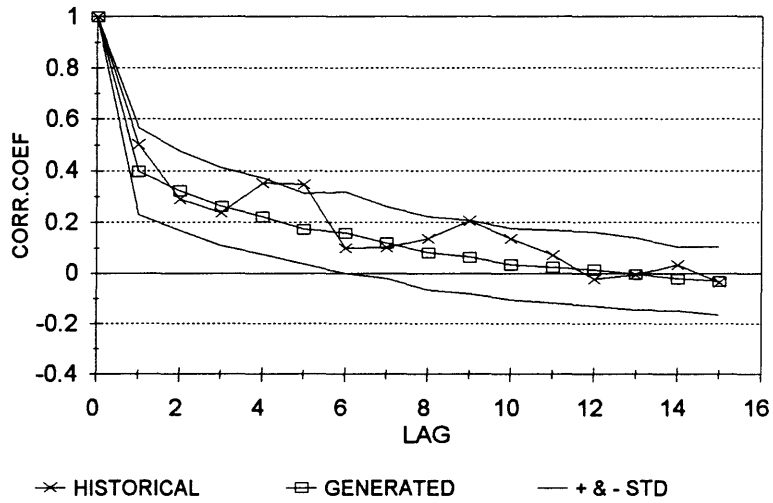
Fig 5-15   Historic and generated autocorrelogram for lake St.Clair based
on CGNN model.
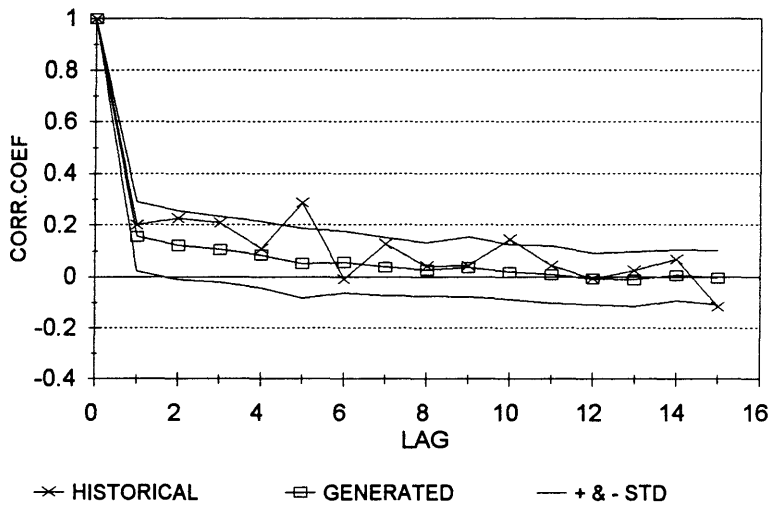

Fig 5-16   Historic and generated autocorrelogram for lake Erie based
on CGNN model.

Fig 5-17   Historic and generated autocorrelogram for lake Ontario based
            on CGNN model.

Fig 5-18   Historic and generated autocorrelogram for lake Superior based
on CARMA(1,1) model.

Fig 5-19   Historic and generated autocorrelogram for lake Michigan-Huron
based on CARMA(1,1) model.

249

Fig 5-20   Historic and generated autocorrelogram for lake St.Clair based
on CARMA(1,1) model.

Fig 5-21   Historic and generated autocorrelogram for lake Erie based
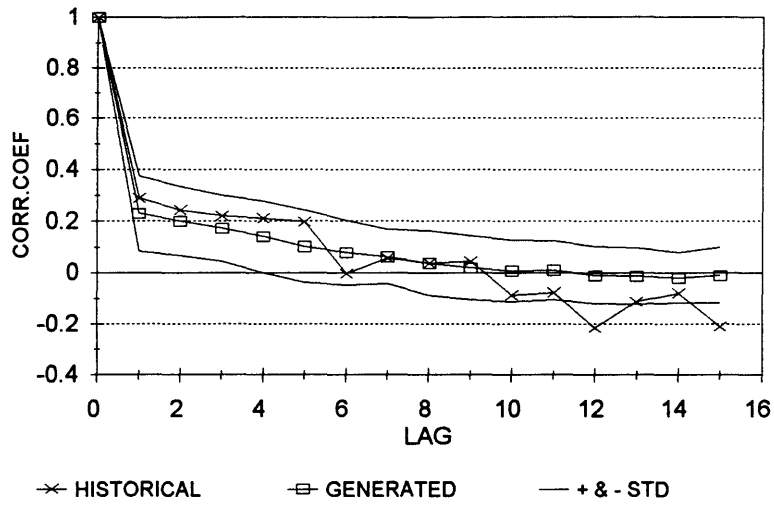on CARMA(1,1) model.

Fig 5-22   Historic and generated autocorrelogram for lake Ontario based
on CARMA(1,1) model.

Fig 5-23   Historic and generated autocorrelogram for lake Superior based
            on CARMA(2,1) model.

Fig 5-24   Historic and generated autocorrelogram for lake Michigan-Huron
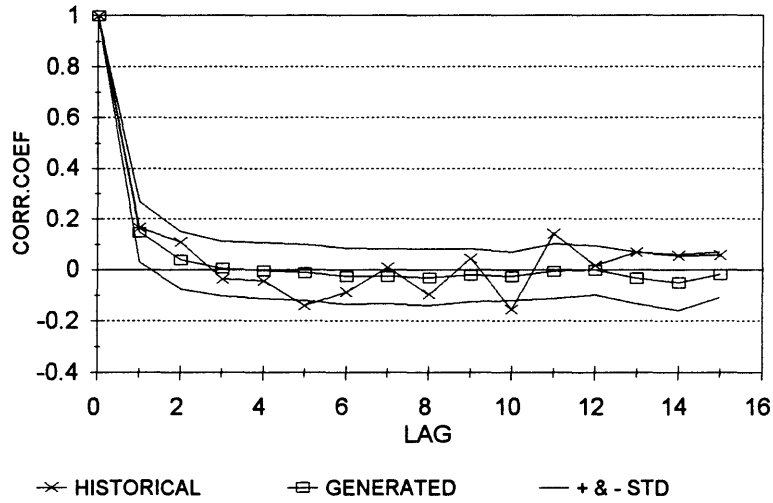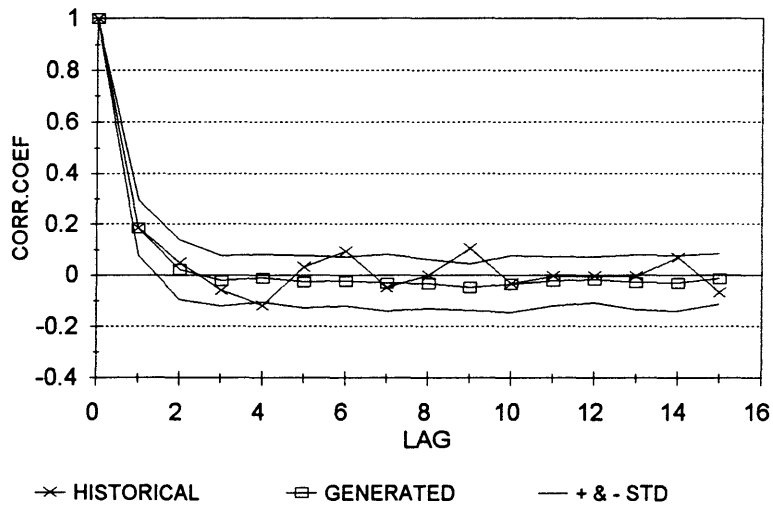   based on CARMA(2,1) model.

252

Fig 5-25   Historic and generated autocorrelogram for lake St.Clair based
on CARMA(2,1) model.

Fig 5-26   Historic and generated autocorrelogram for lake Erie based
on CARMA(2,1) model.

253

Fig 5-27   Historic and generated autocorrelogram for lake Ontario based
on CARMA(2,1) model.

Fig 5-28   Historic and generated autocorrelogram for Aswan based on
GNN-CARMA(1,1) Model.

Fig 5-29   Historic and generated autocorrelogram for Khartoum based on
GNN-CARMA(1,1) Model.

255

Fig 5-30   Historic and generated autocorrelogram for Roseires based on
GNN-CARMA(1,1) Model.
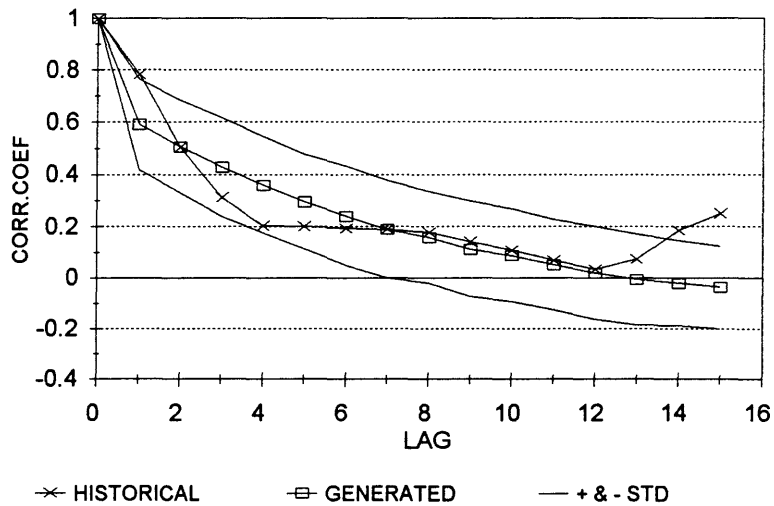
Fig 5-31   Historic and generated autocorrelogram for Malakal based on
GNN-CARMA(1,1) Model.

256
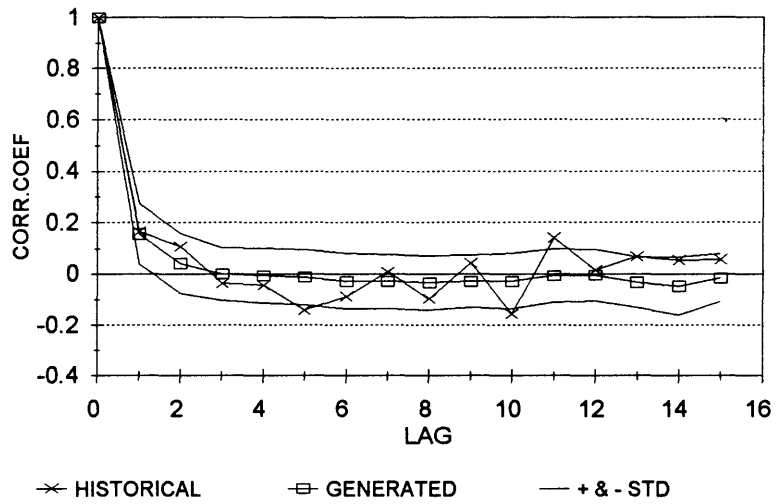
Fig 5-32   Historic and generated autocorrelogram for lake Superior based
            on GNN-CARMA(1,1) model.
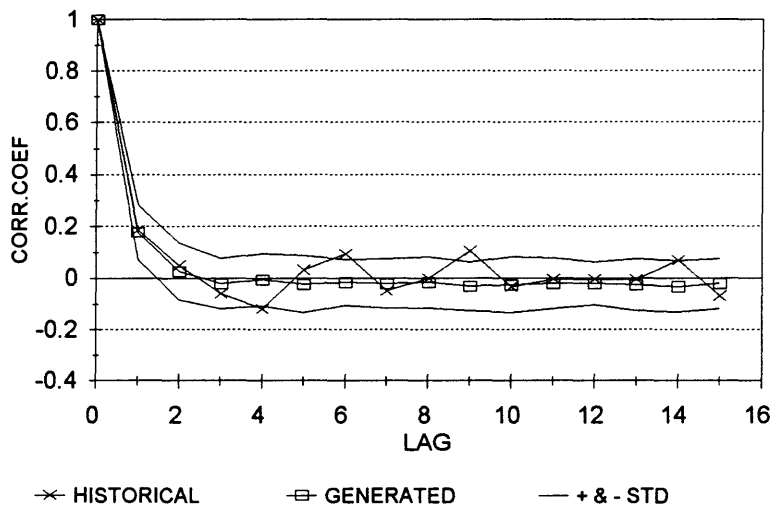
Fig 5-33   Historic and generated autocorrelogram for lake Michigan-Huron
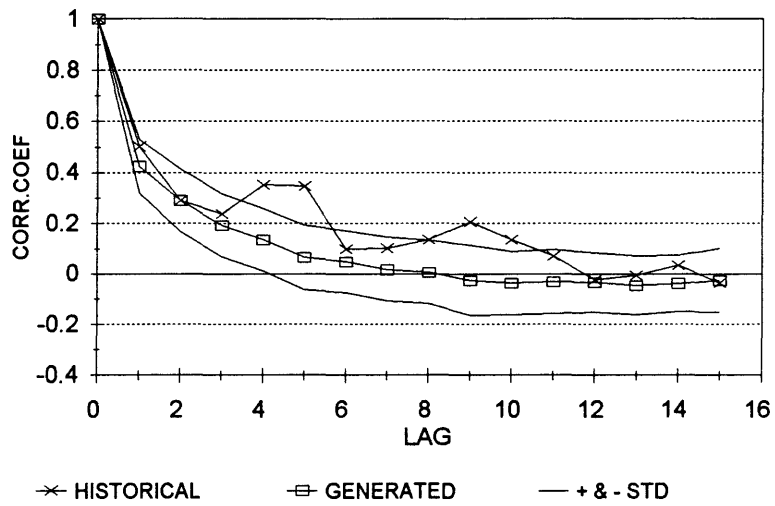   based on GNN-CARMA(1,1) model.
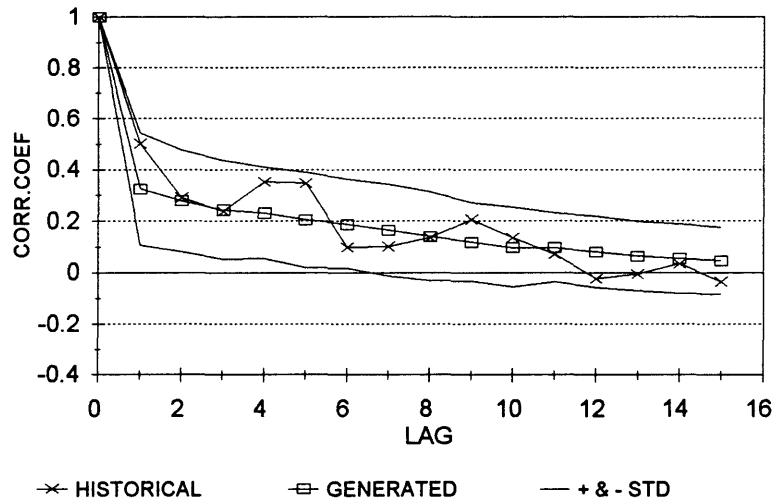
Fig 5-34   Historic and generated autocorrelogram for lake St.Clair based
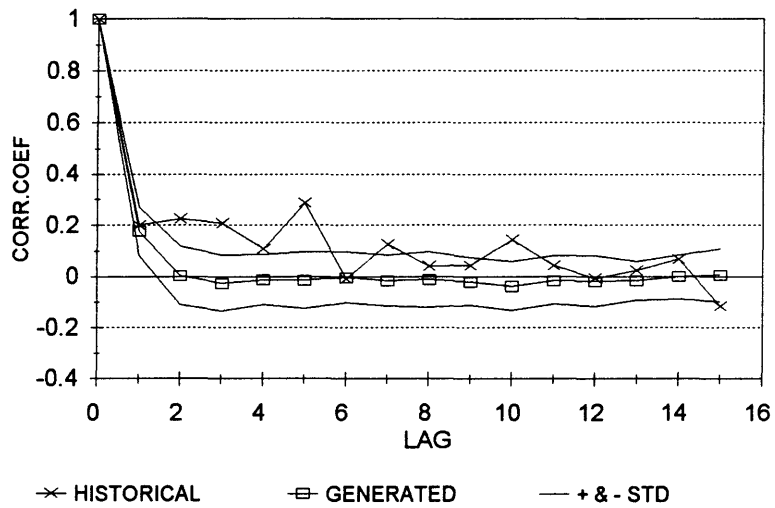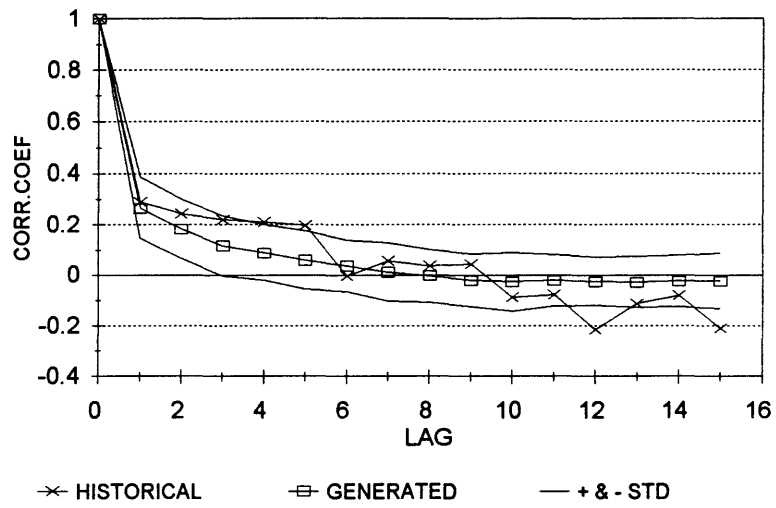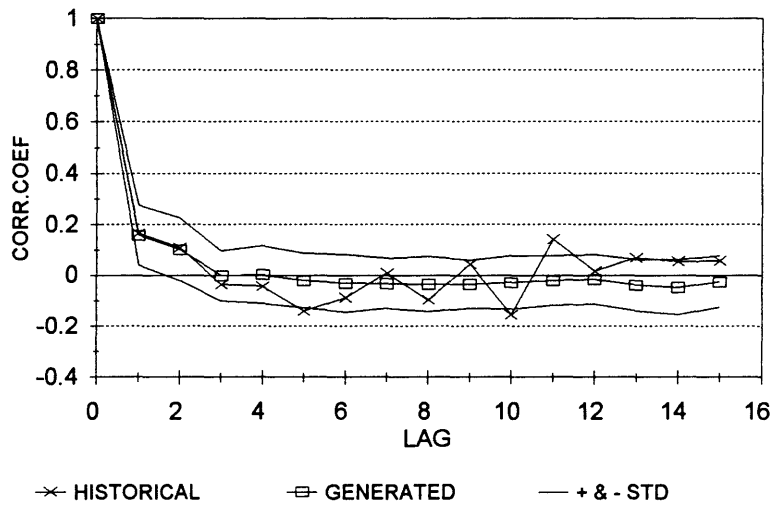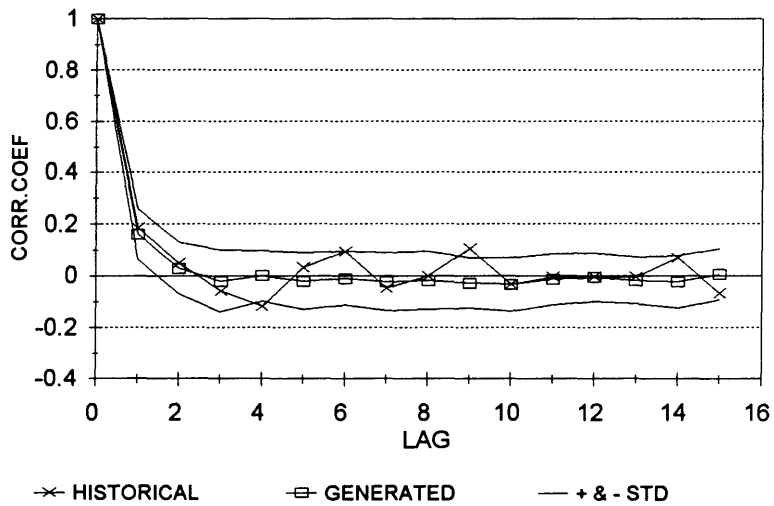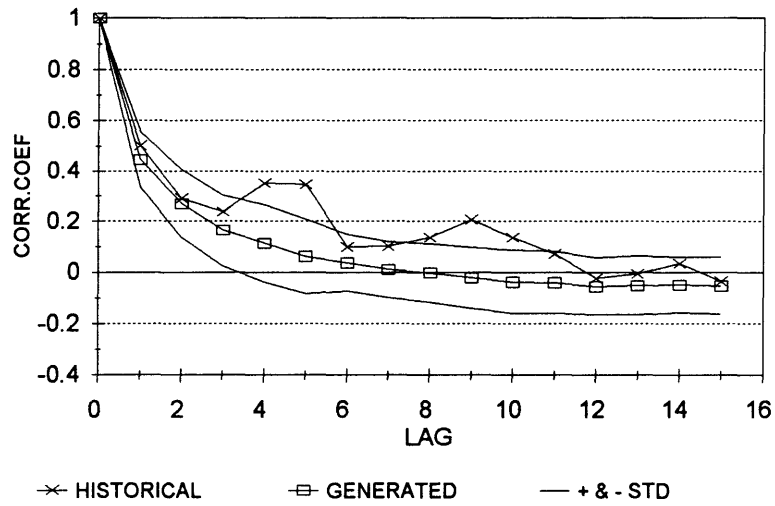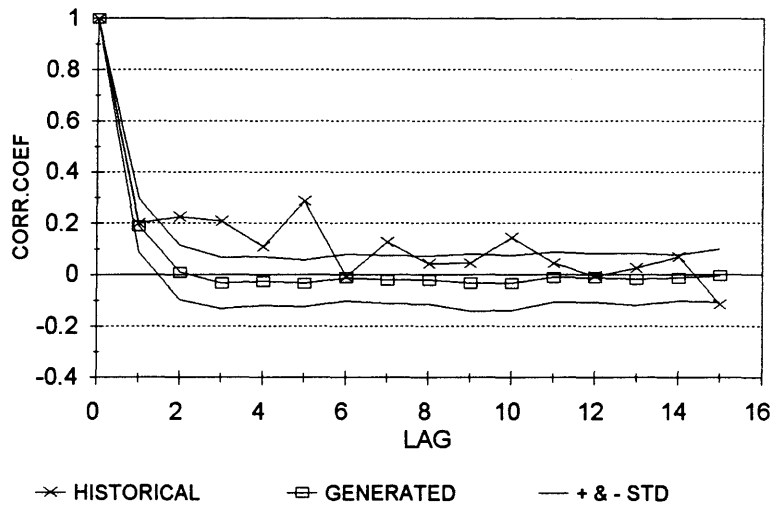on GNN-CARMA(1,1) model.
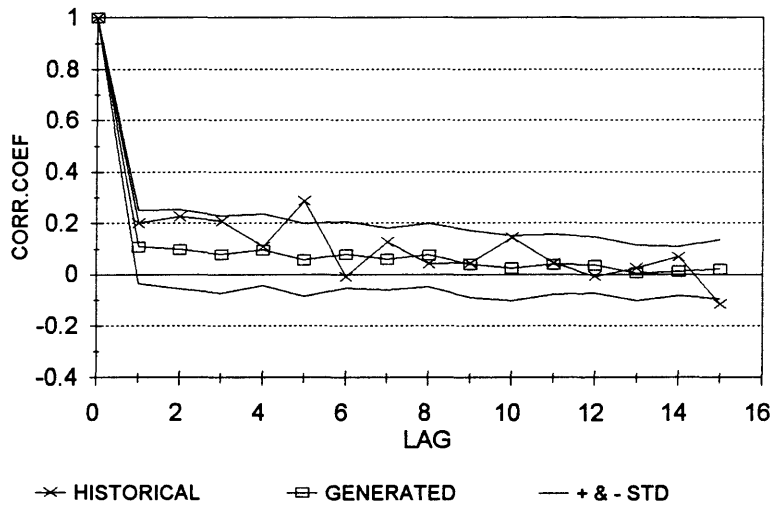
Fig 5-35   Historic and generated autocorrelogram for lake Erie based
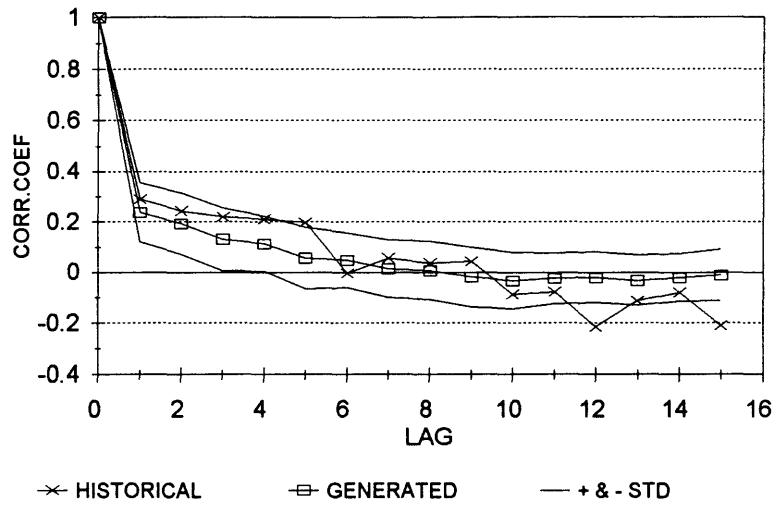on GNN-CARMA(1,1) model.
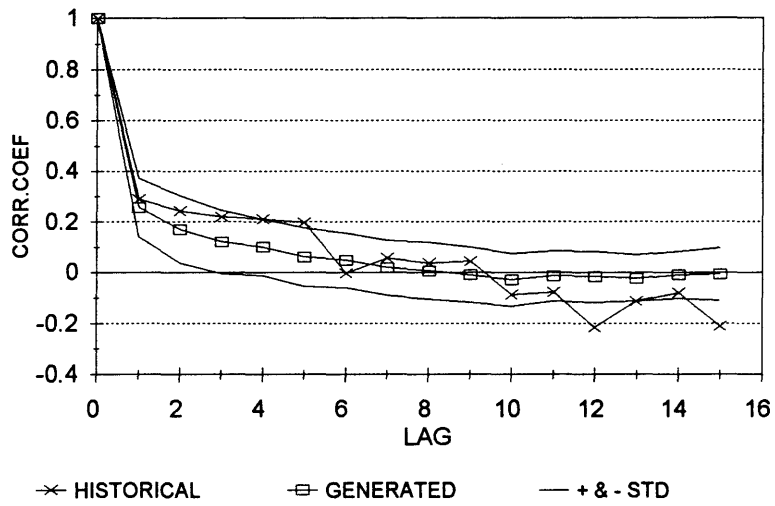
Fig 5-36   Historic and generated autocorrelogram for lake Ontario based
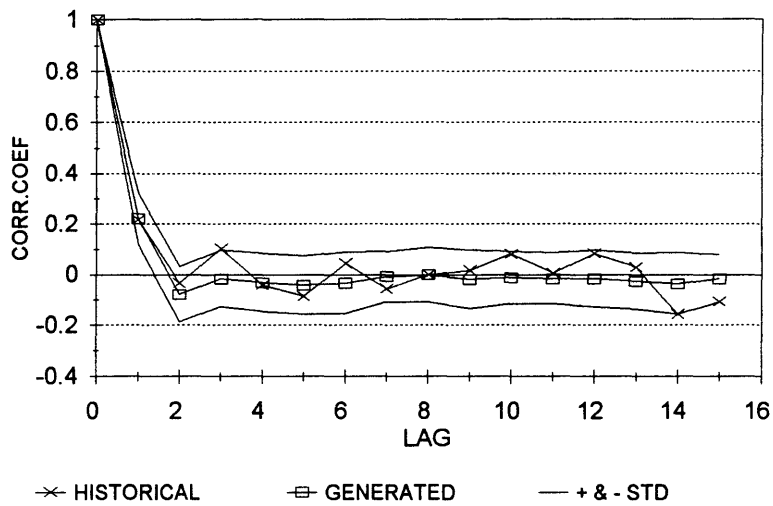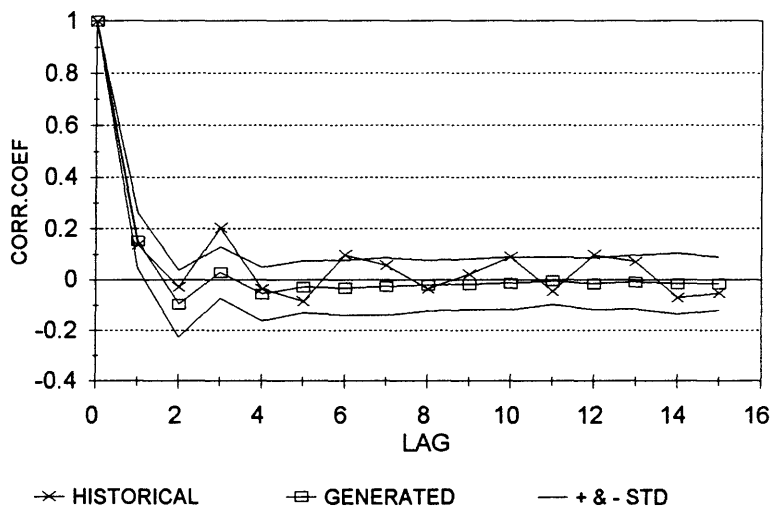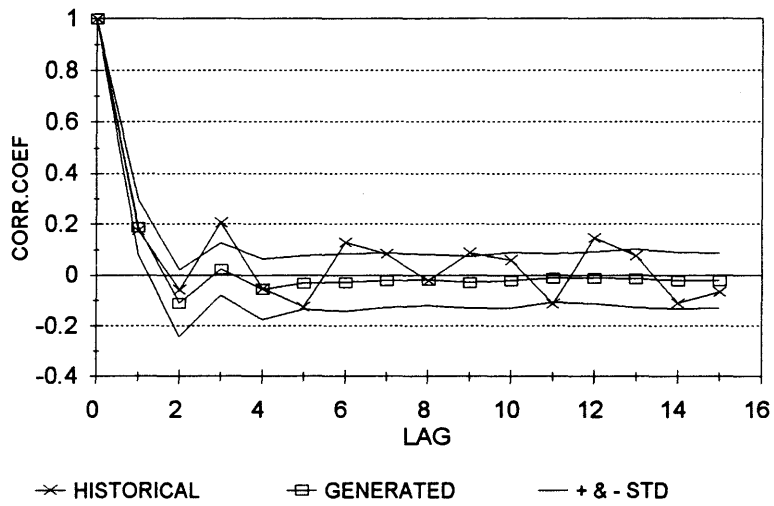   on GNN-CARMA(1,1) model.

CHAPTER VI

SUMMARY AND CONCLUSIONS

The general objective of this study was to develop certain types of shifting level models that can be used in modelling and simulation of processes that exhibit changes. Specific objectives were: to compare alternative methods for estimating the parameters of such models, to develop multivariate models which can be used to model hydrologic processes at more than one site. Chapter IV of this dissertation addressed the parameter estimation of the GNN model. The MOM was analyzed as a potential method to be used in parameter estimation. In this study the general moment equations of the GNN model were derived for any lag-k autocorrelation. As such these equations can be used to estimate the model parameters without the restrictions of using only the lag-1 and lag-2 autocorrelation coefficients. Other methods to estimate the parameters were also suggested. These methods include regression analysis, fitting the autocorrelation function (ACF), using the range properties, and using the run properties. The performance of these methods to preserve the statistical characteristics of the shifting mean model was tested by using simulation experiments. Data was generated from GNN with different population parameters. Then these estimations methods were used to estimate the GNN model parameters. The bias and RMSE were calculated to test the performance of the methods. The regression analysis and range methods, in general, performs better than the other methods.

The GNN model was then applied to Malakal annual flows and lake St.Clair annual NBS. Simulation experiments were conducted to test the ability of the GNN model to preserve important observed statistics such as the mean, standard deviation, skewness,

rescaled range, Hurst coefficient, longest drought, maximum deficit, and surplus. Results show that the GNN model, in general, performs well in preserving these statistics. The performance of the GNN model was also compared with different ARMA models that were also used to model the observed data for Malakal and St.Clair. Results indicate that the GNN has some advantages over ARMA models especially in preserving the rescaled range and Hurst coefficient.

An extended version of the GNN model called GNN-1 was also derived and analyzed in this study. Different methods of estimation were suggested to estimate the model parameters. Results indicate that this model is an alternative for modeling observed hydrologic time series which exhibit changes. However, in comparing this model with the simpler GNN model, it does not give a better performance.

In chapter V a contemporaneous multivariate version of the GNN model called CGNN was formulated. The covariance structure of the model was derived and different methods to estimate the model parameters were suggested. The CGNN model can be used in situations were the sites to be analyzed have apparent shifts in the mean for all sites. The model was used to simulate annual streamflow records at several sites in the Nile River basin. The model was also used to model the NBS records of the Great Lakes System. In both cases, the CGNN model performed well in preserving both basic and long term statistics for these sites.

Another contemporaneous multivariate model called as GNN-CARMA(p,q) was developed. This model is a combination of the CGNN and CARMA models that can be used to model multisite systems. This model can be applied to systems in which some stations

exhibit shifts and other stations in the same system do not appear to exhibit shifts. A good example is the Great Lakes system were the observed annual NBS for lakes St.Clair and Erie strongly suggest the use of a shifting level model. On the other hand the observed record for lakes Superior, Michigan, and Ontario do not indicate a need to use a shifting level model and a classical CARMA model is suitable for them. As such the GNN-CARMA model may be useful in this case. Results indicate that this model performed quite well. The model was able to preserve the short and long term basic statistics such as the mean, standard deviation, rescaled range, and Hurst coefficient.

In summary, the study of GNN models in general reveals that such models are powerful models that can be used to model hydrologic processes that exhibit changes. One advantage of using such models is the fact that such models are easy to use and operationally simple. Another advantage lies in the fact that this model can be argued to be physically sound to be used to model hydrologic processes. In other words, the GNN model is not only an operational model but also have some physical justification to be used in modeling hydrologic processes which exhibit changes.

Finally, further research of the GNN class models is needed. One of the obvious research areas is to develop a periodic or seasonal version of the model that can be used to model seasonal observed data.

REFERENCES

1. Ballerini, R., and Boes, D. "Hurst behavior of shifting level processes". *Water Resources Research*, 21(11),pp (1642-1648),1985.

2. Bates, G.T, Giorgi,F and Hostetler, S.W., " Toward the simulation of the effects of the great lakes on regional climate ", *Monthly Weather Review*. Vol(121),pp.1373-87, 1994.

3. Bishop, C.T." Historical variation of water levels in lakes Erie and Michigan-Huron", *Journal of Great Lakes Research*. vol.16, No.3, pp.406-425, 1990.

4. Boes, D and Salas,J. "Nonstationarity of the mean and the Hurst Phenomena". *Water Resources Research.*,14(1),pp (135-143), 1978.

5. Bras, R.L and Rodriguez-Iturbe,I. *Random Functions and Hydrology*, Addison-Wesley publishing company,1985.

6. Camacho, F., Mcleod,I and Hipel,K.W. "Contemporaneous Auto regressive-Moving Average (CARMA) modeling in Water Resources". *Water Resources Bulletin*, 21(4): pp(709-720),1985.

7. Eltahir, E.A.B. "El Nino and the natural variability in the flow of the Nile River",*Water Resources Research.*,32(1),pp(131-137),1996.

8. Flaschka, I., Stokton,W and Boggess,W.R."Climatic variation and surface water resources in the great basin region".*Water Resources Bulletin*, 23(1): pp(47-57),1987.

9. Gutowski, W.J, McMahon, G.F, Schluchter, S.S and Kirshen, P.H " Effects of global warming on hurricane-induced flooding", *ASCE, Journal of Water Res Planning and Management*. 120(2),pp(176-185),1994.

10. Hipel, K.W., and A.J. Mcleod. "Preservation of the rescaled adjusted range, 1, a reassessment of the Hurst phenomenon". *Water Resources Research*, 14(3),pp (491-516),1978.

11. Hirsch, R.M, Walker, J.F, Day, J.C and Kallio, R." The influence of man on hydrologic systems",In: The Geology of North america, Volume 0-1, Surface Water Hydrology, Edited by M.G Wolman and H.C riggs, 1990.

12. Hubert, P., Carbonnel,J.P and Chaouche,A "Application ADES SERIES DE PRECIPITATIONS ET DE DEBITS DE L'AFRIQUE DE L'OUEST " *Journal*

of Hydrology., 110 (349-367), 1989.

13. Hurst, H.E.," Long-term storage capacities of reservoirs". *trans.Am.soc.civ Engrs.*,vol 116,pp (770-808),1951.

14. Irvine, K.N. and Eberhardt, A.J., " Multiplicative, seasonal ARIMA models for lake Erie and lake Ontario ", *Water Resources Bulletin.* 28(2),pp.1271-1275, 1992.

15. Jackson, B.B. " The use of streamflow models in planning".*Water Resources Research.*, 11(1), pp (54-63), 1975.

16. Jackson, B.B. " Birth-death models for differential persistence".*Water Resources Research.*, 11(1), pp (75-95),1975.

17. Kiladis, G.N, and Diaz, H.F."Global climatic anomalies associated with extremes in the southern oscillation", *Journal Of Climate.*,Vol 2 ,pp(1069-1090),1989.

18. Kite, G.W." Use of time series analysis to detect climate change", *Journal of Hydrology.* vol(111),pp(259-279),1989.

19. Klemes, V.," The Hurst phenomenon: A puzzle ?,*Water Resources research.*,10(4),pp (675-688),1974.

20. Lau, K.M. and Sheu, P.J."Annual cycle, Quasi-Biennial oscillation, and southern oscillation in global precipitation",*Journal Of geophysical resarch.*,93(D9) ,pp(10975-10988), September 1988.

21. Lettenmaier, D.P, and Sheer, D.P." Climatic sensitivity of California water resources", *ASCE, Journal of Water Res Planning and Management.* 117(1),pp(108-125),1991.

22. Lettenmaier, D.P, and Gan, T.Y." Hydrologic sensitivities of the Sacramento-San Joaquin river basin, California, to global warming", *Water Res research.* 26(1), pp(69-86),1990.

23. Lin, Y and Lye, L.M."Modeling long-term dependance based on cumulative departures of annual flow series",*Journal of Hydrology.*,V. 160, pp(105-121),1994.

24. Lins, H.F, Hare, F.K, and singh, K.P." Influence of atmosphere",In: The Geology of North america, Volume 0-1, Surface Water Hydrology, Edited by M.G Wolman and H.C riggs, 1990.

25. Mandelbrot, B.B., and J. VanNess. " Fractional Brownian Motions, Fractional Noises and Applications", *Soc. Ind. Appl. Math. Rev.*,10(4), pp(422-437), 1968.

26. Mandelbrot, B.B., and J.R. Wallis. " Noah, Joseph, and Operational hydrology", *Water Res research.* 4(5), pp(909-918),1968.

27. Matalas, N.C., and C.S. Huzzen."A Property of the Range of Partial Sums". proceedings of the International Hydrology Symposium, vol 1, pp(252-257), Colorado State University, Fort Collins, Colorado, (1967).

28. Matalas, N.C " Mathematical assessment of synthetic hydrology". *Water Resources Research.*,3(4),pp(937-945),1967.

29. Matalas, N.C and Wallis,J.R. " Statistical properties of multivariate fractional noise process". *Water Resources Research.*,7(6),pp(1460-1468),1971.

30. Mechoso, C.R and Iribarren, G.P."Streamflow in Southeastern South America and Southern Oscillation",*Journal Of Climate.*,Vol 5 ,pp(1535-1539),1992.

31. Mesa, O. J. and Poveda, G."The Hurst effect: The scale of fluctuation approach",*Water Resources Research.*,29(12),pp(3995-4002),1993.

32. Mitchell, J.F, " The "green house" effect and climate change ", *Review of Geophysics.* 27(1),pp.115-140, 1989.

33. Obeysekera, J.B. Run and range properties of shifting level models. Ph.d Dissertation, Colorado State University, Fort Collins, Colorado, (1981).

34. O'Connell, P.E. Presistance in streamflow sequences. Ph.D Dissertation, Univ of London (1974).

35. Potter, K.W."Comment on the 'hurst phenomenon: a puzzle?' by V.Klemes". *Water Resources Research.*,11(2),pp(373-374),1975.

36. Potter, K.W."Evidence for nonstationarity as a physical explanation of the Hurst phenomenon",*Water Resources Research.*,12(5),pp(1047-1052),1976.

37. Potter, K.W."Annual Precipitation in the Norteast United States: Long Memory, Short Memory, or No Memeory?",*Water Resources Research.*,15(2),pp(340-346),1979.

38. Privalsky, V.," Statistical analysis and predictability of lake Erie water level

variations", *Journal of Great Lakes Research*. vol.18, No.1, pp.236-243, 1992.

39.   Quinn, F.H. and Sellinger, C.E., " Lake Michigan record levels of 1838, A present perspective", *Journal of Great Lakes Research*. vol.16, No.1, pp.133-138, 1990.

40.   Quinn, F.H., " Lake Superior regulation effects", *Water Resources Bulletin*. 12(6) ,pp.1271-1275, 1976.

41.   Riggs, H.C. and Harvey, K.D." Temporal and spatial variability of streamflow",In The Geology of North america, Volume 0-1, Surface Water Hydrology, Edited by M.G Wolman and H.C riggs, 1990.

42.   Ropelewski, C.F."Quantifying southern oscillation-precipitation relationships",*Journal of Climate.*, Vol 9, pp(1043-1059),May 1996.

43.   Ropelewski, C.F."North American precipitation and temperature patterns associated with El Nino / Southern oscillation (ENSO)", *Monthly Weather Review.*, Vol 114(2), pp(2352-2362),December 1986.

44.   Salas, J. and Boes,D.," Shifting level modeling of hydrologic series", *Adv. Water Resources.*, vol.3,pp(59-63),1980.

45.   Salas, J., Boes, D.,Yevjevich, V and Pegram,G.G.S." Hurst phenomenon as a pre-asymptotic behavior" *Journal of Hydrology.*, 44(1-15), 1979.

46.   Salas, J., Tabios III, G.Q. and Bartolini, P." Approaches to multivariate modeling of water resources time series" *Water resources Bulletin.*, 21(4),pp(683-707),1985.

47.   Salas, J., Delleur, J.W., Yevjevich, V and Lane, W.L. Applied modeling of hydrologic time series. Water Resources Publications, Littleton, Colorado, (1980).

48.   Scott, D.F." The hydrological effects of fire in South african mountain catchments", *Journal of Hydrology*. vol(115),pp(409-432),1993.

49.   Sen, Z." Note on the cyclic features in cumulative departrures of annual flow series", *Journal of Hydrology*. vol(125),pp(47-54),1991.

50.   Sene, K.J. and Plinston, D.T., " A review and update of the hydrology of lake victoria in East Africa", *Hydrological Sciences Journal*. 39(1),pp.47-63, 1994.

51.   Shalash, S." effect of Owen Falls dam on the hydrological regime of lake Victoria". In: The influence of man on the hydrological regime with special

reference to representative and experimental basins(proc. Helsinki symposium,June 1980), 239-243. IAHS pub.no 130.

52.   Siddiqui, M.M., " The Asymptotic Distribution of the Range and other Functions of partial sums of Stationary Processes", *Water Resources Research.* 12(6),pp.1271-1275, 1976.

53.   Slack, J.R. " Hydro-Climatic Data Network (HCDN), A USGS streamflow data set for the United states for the study of climate variations, 1874-1988, USGS open-file report 92-129, Reston, Virginia 1992.

54.   Stedinger, J. and Taylor, M." Synthetic streamflow generation,1. Model verification and validation", *Water Resources Research.*,18(4),pp(908-918),1982.

55.   Stockton, C.W." Interpretation of past climatic variability from paleoenvironmental indicators;In Climate, Climate Change and Water Supply, Geophys, Res Board, Nat. Res. Council., 1977, Washington, DC 1977.

56.   Turner, K.M."Evaluating reliability of storage schemes with dendohydrology and the Hurst phenomena", *Water Resources Bulletin.*,32(1),pp(47-55),1996.

57.   Wallis, J.R.," Climate, climatic change and water supply",*E.O.S., American Geophysical Union.*pp(1012-1023), Nov 1977.

58.   Williams, G.P and Wolman, M.G." Effects of dams and reservoirs on surface-water hydrology; Changes in rivers downstream from dams; USGS Professional paper 1286.

59.   Young, G.K, and Pisanu, W.C." Operational hydrology using residuals", *ASCE, Journal of Hydraulic Division.* Hy (4),pp(909-923), July 1968.

APPENDIX A

Proof of Equation 4-45:

From Eq. (4-22) (see Obeysekera ,1980):

$$\beta^2 = 1 + 2 \sum_{k=1}^{\infty} \rho_x(k) \tag{A-1}$$

$$\beta^2 = 1 + 2 \rho_x(1) + 2 \sum_{k=2}^{\infty} \rho_x(k)$$

and we have shown that (see section 4.1.2):

$$\rho_x(k) = \frac{\sigma_m^2}{\sigma_x^2}(1-p)^k \qquad for \ k > 1 \tag{A-2}$$

subtituting Eq. (A-2) into Eq. (A-1) gives

$$\beta^2 = 1 + 2 \rho_x(1) + 2 \frac{\sigma_m^2}{\sigma_x^2}[(1-p)^2 + (1-p)^3 + (1-p)^4 + \ldots\ldots] \tag{A-3}$$

Since

$$\frac{1-p}{p} = [(1-p) + (1-p)^2 + (1-p)^3 + (1-p)^4 + \ldots\ldots] \tag{A-4}$$

Eq (A-3) becomes:

$$\beta^2 = 1 + 2 \rho_x(1) + 2 \frac{\sigma_m^2}{\sigma_x^2}\left[\frac{(1-p)^2}{p}\right] \tag{A-5}$$

268

Also, for k=2 Eq.(A-2) gives

$$\frac{\sigma_m^2}{\sigma_x^2} = \frac{\rho_x (2)}{(1-p)^2}$$
(A-6)

then, Eq.(A-5) can be written

$$\beta^2 = 1 + 2 \rho_x (1) + \frac{2 \rho_x (2)}{p}$$
(A-7)

or

$$p = \frac{2 \rho_x (2)}{\beta^2 - 1 - 2 \rho_x (1)}$$
(A-8)

APPENDIX B

Proof that the CGNN model mentioned in section 5.2.2 is not stationary. The proof will be done for the bivariate case.

$$
\begin{aligned}
E(M_1^{*^{(1)}} M_1^{*^{(2)}}) &= E\{ \ [M_1^{(1)} I_{\{1,2,3,\ldots\}}(N_1^{(1)})] \ [M_1^{(2)} I_{\{1,2,3,\ldots\}}(N_1^{(2)})] \ \} \\
&= E[M_1^{(1)} M_1^{(2)}] \ E[I_{\{1,2,3,\ldots\}}(N_1^{(1)})] \ E[I_{\{1,2,3,\ldots\}}(N_1^{(2)})] \\
&= E[M_1^{(1)} M_1^{(2)}] \quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\text{(B-1)}
\end{aligned}
$$

where I is the indicator function

$$
\begin{aligned}
E(M_2^{*^{(1)}} M_2^{*^{(2)}}) = E\{ \ &[M_1^{(1)} I_{\{2,3,\ldots\}}(N_1^{(1)}) + M_2^{(1)} I_{\{1\}}(N_1^{(1)})] \\
&[M_1^{(2)} I_{\{2,3,\ldots\}}(N_1^{(2)}) + M_2^{(2)} I_{\{1\}}(N_1^{(2)})] \ \}
\end{aligned}
\quad\text{(B-2)}
$$

$$
\begin{aligned}
= \ &E\{ \ [M_1^{(1)} I_{\{2,3,\ldots\}}(N_1^{(1)})][M_1^{(2)} I_{\{2,3,\ldots\}}(N_1^{(2)})]\} \\
+ \ &E\{ \ [M_1^{(1)} I_{\{2,3,\ldots\}}(N_1^{(1)})][M_2^{(2)} I_{\{1\}}(N_1^{(2)})]\} \\
+ \ &E\{ \ [M_2^{(1)} I_{\{1\}}(N_1^{(1)})][M_1^{(2)} I_{\{2,3,\ldots\}}(N_1^{(2)})]\} \\
+ \ &E\{ \ [M_2^{(1)} I_{\{1\}}(N_1^{(1)})][M_2^{(2)} I_{\{1\}}(N_1^{(2)})]\}
\end{aligned}
\quad\text{(B-3)}
$$

and since the M process is spatially correlated only at lag zero, the second and third terms will drop from Eq. (B-3). Therefore, Eq. (B-3) will simplify to :

$$
\begin{aligned}
= \ &E\{ \ [M_1^{(1)} M_1^{(2)}] \ E[I_{\{2,3,\ldots\}}(N_1^{(1)}) \ I_{\{2,3,\ldots\}}(N_1^{(2)})] \ \} \\
+ \ &E\{ \ [M_2^{(1)} M_2^{(2)}] \ E[I_{\{1\}}(N_1^{(1)}) \ I_{\{1\}}(N_1^{(2)})] \ \}
\end{aligned}
\quad\text{(B-4)}
$$

The four terms in the RHS of Eq. (B-4) can be written as:

$$
E[\ I_{\{2,3,\ldots\}}(N_1^{(1)}) \ ] = (1 - p^{(1)})
$$

$$
E[\ I_{\{2,3,\ldots\}}(N_1^{(2)}) \ ] = (1 - p^{(2)})
$$

270

$$E[\ I_{\{1\}}(N_1^{(1)})\ ] = p^{(1)}$$

$$E[\ I_{\{1\}}(N_1^{(2)})\ ] = p^{(2)}$$

Then equation (B-4) becomes:

$$E(M_2^{*(1)}\ M_2^{*(2)}) = E\{\ [M_1^{(1)}\ M_1^{(2)}]\ \}\ (1 - p^{(1)})\ (1 - P^{(2)})$$
$$+ E\{\ [M_2^{(1)}\ M_2^{(2)}]\ \}\ \ p^{(1)}\ p^{(2)} \tag{B-5}$$

and since $E\,[M_1^{(1)}\ M_1^{(2)}] = E\,[M_2^{(1)}\ M_2^{(2)}]$, Eq. (B-5) becomes:

$$E(M_2^{*(1)}\ M_2^{*(2)}) = E\,[M_1^{(1)}\ M_1^{(2)}]\ \{\ (1 - p^{(1)})\ (1 - p^{(2)}) + p^{(1)}\ p^{(2)}\} \tag{B-6}$$

Then from Eqs. (B-1) and (B-4), one can find that

$$E(M_2^{*(1)}\ M_2^{*(2)}) \neq E(M_1^{*(1)}\ M_1^{*(2)})$$

which means that the CGNN model is not stationary in this case.