GREAT LAKES
NET BASIN SUPPLY SIMULATION
BY A STOCHASTIC APPROACH

Final report, research project INRS-Eau, Hydro-Québec*:

GREAT LAKES

NET BASIN SUPPLY SIMULATION

BY A STOCHASTIC APPROACH

Rapport scientifique no 362

Louis Mathier
Laura Fagherazzi
Jean-Claude Rassam
Bernard Bobée

November 1992

INRS-Eau
Université du Québec
C.P. 7500
Sainte-Foy (Québec)
G1V 4C7

* Research contract DAC-91-ERE-342

CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

The present study was carried out by the Institut National de la Recherche Scientifique - Eau (INRS-Eau), and by Hydro-Québec Direction Production et Échange d'Énergie with guidance by, and interaction with an expert committee. We would like to thank the members of the committee Dr. Bill Hogg of Environment Canada, Mr. Dave Rockwood of SSARR consultants, Mr. Jean-Louis Bisson of Division Hydrométéorologie Hydro-Québec, Mr. Narut Kang of Groupe Équipement Hydro-Québec, Mr. Douglas Sparks of Groupe Équipement Hydro-Québec, Professor Michel Slivitzky of INRS-Eau, Professor Jose Salas of Colorado State University, and Dr. Frank Quinn of the Great Lakes Environmental Research Laboratory.

Other participants were Dr. Jan Grygier, consultant (California), Professor V.T.V. Nguyen of McGill University Montreal, Mr. David Fay of the Great Lakes Saint-Lawrence Office, Mrs Deborah M. Lee of the Great Lakes Environmental Research Laboratory, and Mr. Jacques Bourret Service Production of Hydro-Québec. We would like to stress the important input to this study by expert committee members Dr. Quinn, Professor Slivitzky, and Professor Salas who provided their assistance up to the very end.

We would like to thank the people who initiated and made possible this study, in particular Mr. Roger Larivière, previous head of Service Production, Hydro-Québec today retired and Dr. Jean-Pierre Lardeau, head of Service Hydraulique, Hydro-Quebec as well as Mr. Gilles Brosseau head of Division Hydrologie, Hydro-Québec.

INTRODUCTION

This final report presents the results of the stochastic approach used to simulate synthetic series of the Great Lakes net basin supply (NBS) for the Beauharnois-Les Cèdres extreme flood study (Rassam *et al.* 1992). A six months duration contract (DAC-91-ERE-342) was awarded to INRS-Eau to analyze the historical records (90 years long) of net basin supply (NBS) of the Great Lakes Basin, review the literature on large-scale multivariate generation techniques and simulate 555 series of 90 years long of NBS on a monthly base. The study has been performed from november 1991 to april 1992 at INRS-Eau in collaboration whith Hydro-Québec.

The Net Basin Supply (NBS) series for Lakes Superior (SU), Michigan-Huron (MH), Erie (ER) and Ontario (ON) were used as the data base for modelling and data generation. Figure 1.1 shows a map of the Great Lakes - Saint Lawrence River Basin. In this study the NBS have been computed as a residual of the water balance equation written as

$$NBS = Q_O \pm D + CU \pm \Delta S - Q_I \qquad (1a)$$

where:

Q_I is the connecting channel inflow
Q_O is the connecting channel outflow
D is the net diversion in or out of the lake
CU is the consumptive use
ΔS is the change in lake storage.

This definition of NBS is generally used for hydrological study of the Great Lakes Basin (IGLLB 1973). The NBS data have been validated and coordinated by the US Corps of Engineers and are provided by the Great Lakes Evironmental Research Laboratory (GLERL), Ann Arbor, Michigan. A description of the Great Lakes system and NBS definition is provided by Yevjevich (1975).

The alternative method of computing NBS is the component method expressed as

$$\mathrm{NBS} = \mathrm{P} + \mathrm{R} - \mathrm{E} \qquad (1b)$$

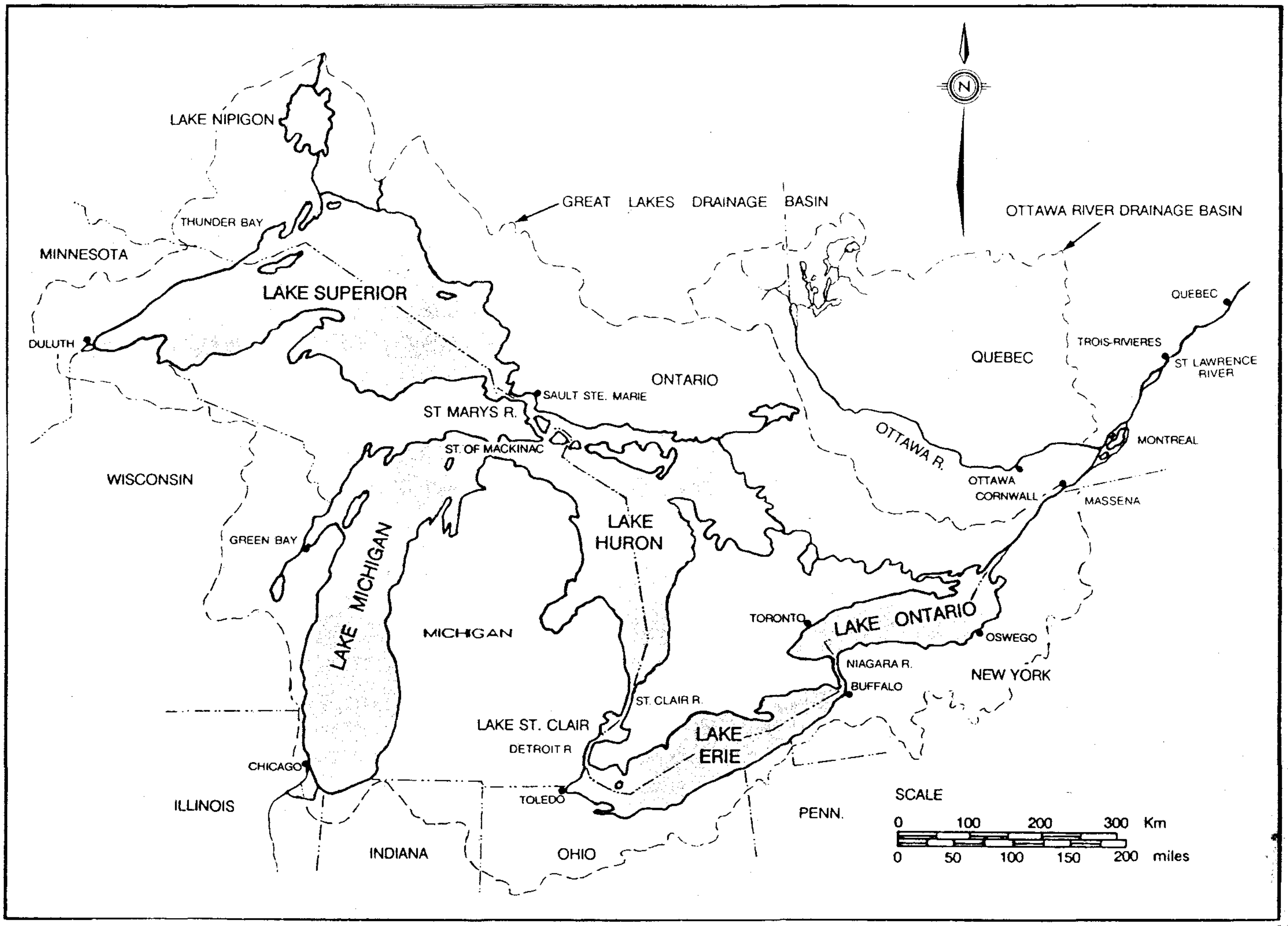


Figure 1.1: Great Lakes - Saint Lawrence River Basin (from Yee *et al.* 1990).

where:

P is the overlake precipitation
R is the runoff from the basin into the lake
E is the evaporation from the lake surface.

The differences in the procedures are discussed by Quinn and Guerra (1986). The significance of the differences is that, when evaluating management alternatives, it is necessary to evaluate past water supplies under current channel conditions. The use of equation (1a) could therefore bias the computations by incorporating errors in connecting channel flow measurements due to measurement techniques or the computation of regime changes. As noted by Quinn (1982), this could result in a considerable error in computing NBS prior to the current channel regimes. Quinn (1982) shows that a 5 % error in either the Detroit or Niagara River flows would result in a 34 % error in the Lake Erie NBS computed by equation (1a). A corresponding 5 % error in the precipitation, runoff, or evaporation terms in equation (1b) would result in a 4-5 % error in the Lake Erie NBS. Thus, while there is uncertainty in the evaluation of the components of equation (1b), the relative impacts are considerably less than in the case of the connecting channel flows. Therefore, it should be recognized that there are potentially large errors which could be introduced in the NBS computed for different channel regimes used in the study. One has to keep in mind that as far as water supply for the entire Great Lakes Basin are concerned the NBS data base computed by equation (1a) is the best existing information that allows us to treat the problem.

In the first section of this report properties of the historical records of NBS are presented for annual and monthly values. Description of the historical NBS is given in time, frequency and spectral domains. This analysis will be used for the selection of the properties to be explicitly preserved for the data generation. The second section presents a literature review of multivariate stochastic models. Section 3 describes the validation procedures and the final model selection. Three multivariate stochastic models were selected to generate annual and monthly NBS samples for the four lakes. For each model, the monthly NBS samples were used to simulate quarter monthly water level data at Lake Ontario. The three samples (annual NBS, monthly NBS, quarter monthly level) were used to validate the NBS and level data against historical records using several validation criteria. This phase is called the exploratory validation. Section 4 presents the final data simulation and validation of NBS for the Great Lakes. This phase is called the confirmatory validation. Finally, the conclusions of the study are presented in the last section.

1. Properties of the net basin supplies (NBS)

Properties of the NBS in the time, frequency and spectral domains are analyzed. The analysis of the basic data (historical NBS) considers both annual (long term) and monthly statistics. Annual statistics include: mean, standard deviation, skewness coefficient, serial correlogram, cross correlation, frequency analysis (using Weibull plotting position), run properties and test of normality. Likewise, monthly statistics include mean, standard deviation, skewness coefficient, month-to-month correlations and monthly cross-correlations.

Besides the analysis of basic statistics, annual time series for all sites are tested statistically to detect possible shift and trend. Analysis also includes basin precipitation time series to see whether apparent concurrent shifts or trends are observed in both NBS and basin precipitation data. Apparent long-term persistence characteristics of the series are also discussed.

1.1 Annual series

Time series of annual NBS for the period 1900-1989 were obtained and analyzed statistically. The NBS series are analyzed by considering the calender year (January to December) while Hydro-Québec typically uses the year from October to September. This difference in the year definition does not make much difference for the purposes of the study. If not specified otherwise, the calender year definition is used in this report. Table 1.1 shows the basic statistics of annual NBS in thousand cubic feet per second (tcfs) for the four lakes.

Table 1.1: Basic statistics of annual NBS (tcfs).

		MH	ER	ON
mean	870,1	1344,3	236,4	430,1
standard-deviation	204,0	312,1	110,3	98,1
coefficient of skewness	0,026	-0,052	0,093	0,471

1.1.1 Serial correlation

Figures 1.2a,b,c,d respectively show the lag-1 to lag-24 correlograms for the four Lakes. For Lake Superior and Lake Michigan-Huron the lag-1 correlation coefficients are low (respectively 0,16 and 0,19). For Lake Erie the lag-2 ($r = 0,22$) and lag-5 ($r = 0,28$) coefficients are low but significant at a 5% level of significance. For Lake Ontario the lag-1 ($r = 0,29$), lag-2 ($r = 0,25$), lag-3 ($r = 0,22$) and lag-four ($r = 0,21$) coefficients are also low but significant at the same level of significance.

1.1.2 Cross-correlation

Table 1.2 gives the matrix of correlation coefficients between annual NBS of the four lakes.

Table 1.2: Correlation coefficients between annual NBS of the four lakes.

	SU	MH	ER	ON
SU	1,00			
MH	0,54	1,00		
ER	0,30	0,50	1,00	
ON	0,27	0,62	0,66	1,00

In general the correlation structure shows a good spatial pattern (Fig. 1.1). Downstream from Lake Superior the correlation coefficients with other lakes are decreasing. Except for Michigan-Huron and Erie the correlations are always higher for two adjacent lakes. The correlations between neighboring lakes are higher going downstream or for smaller lakes.

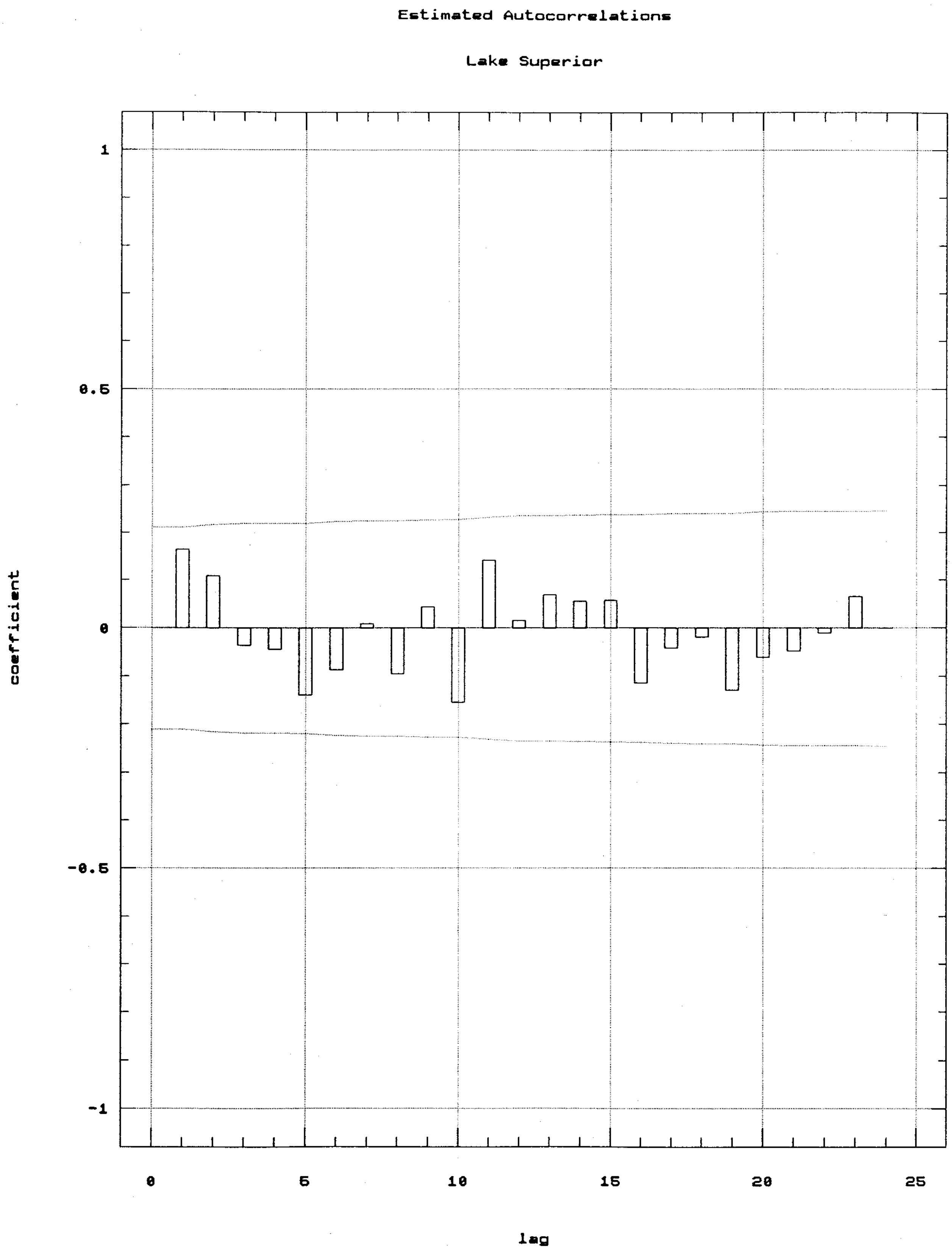


Figure 1.2 a: Lag-one to lag-24 correlogram for Lake Superior.

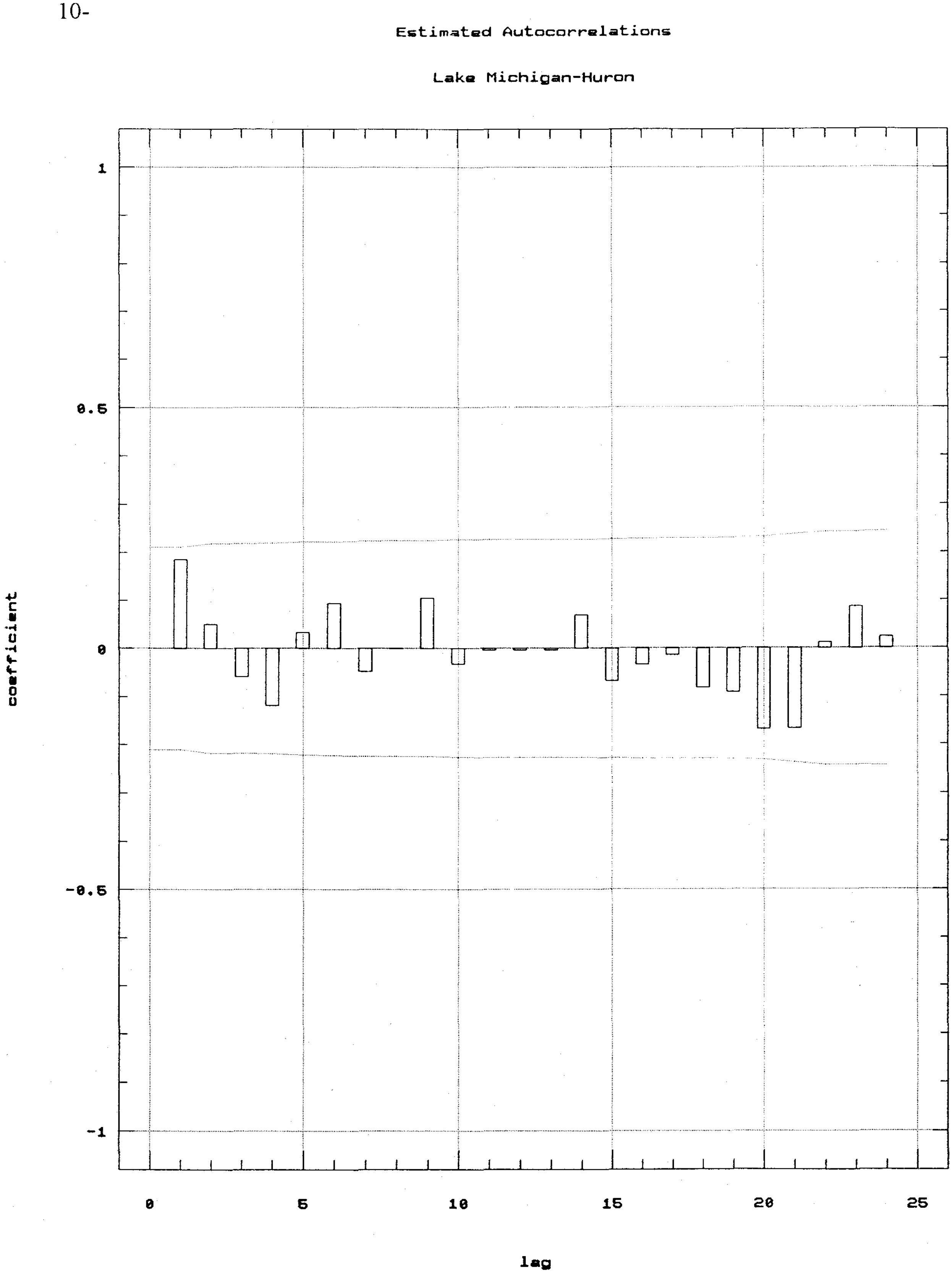


Figure 1.2 b: Lag-one to lag-24 correlogram for Lake Michigan-Huron.

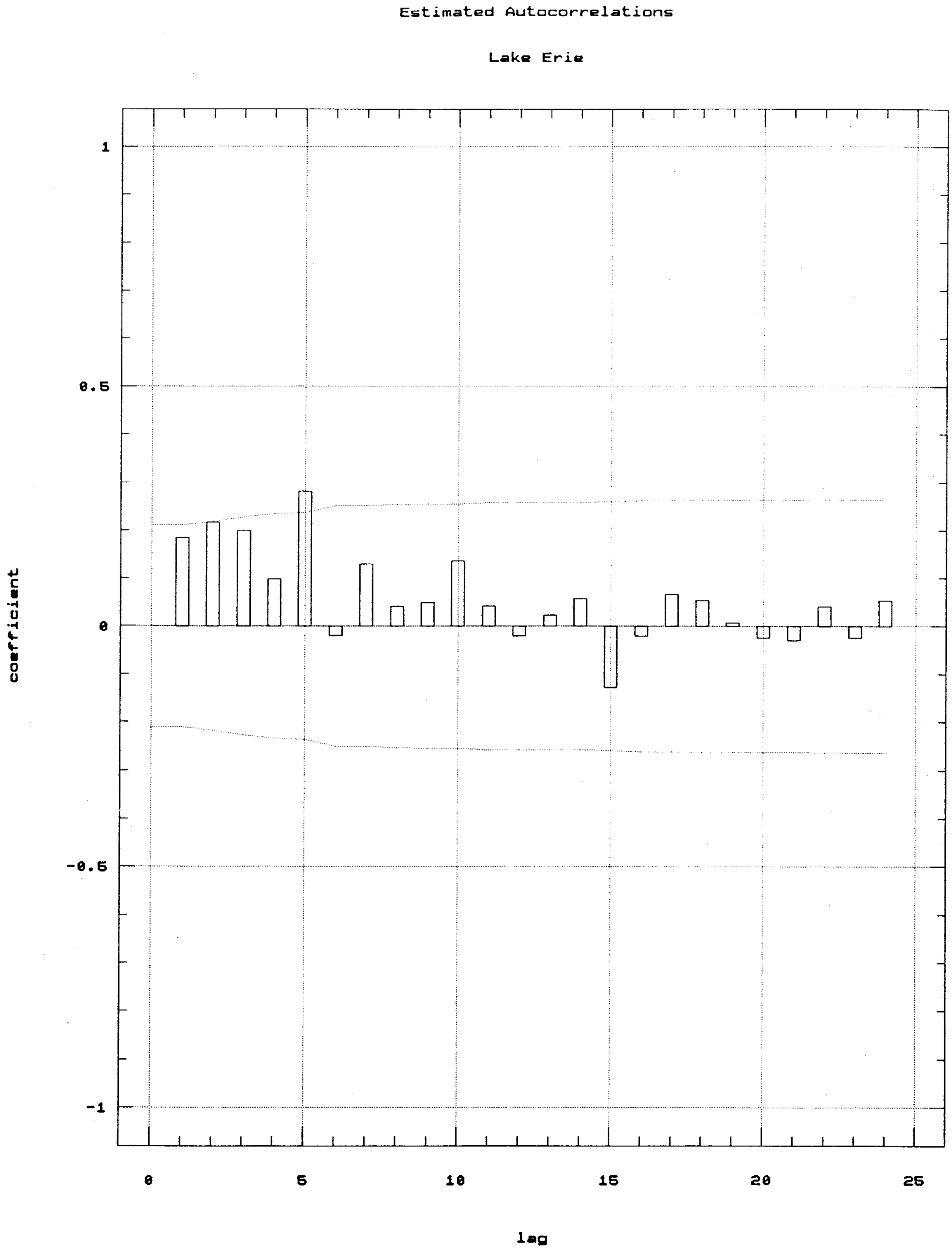


Figure 1.2 c: Lag-one to lag-24 correlogram for Lake Erie.

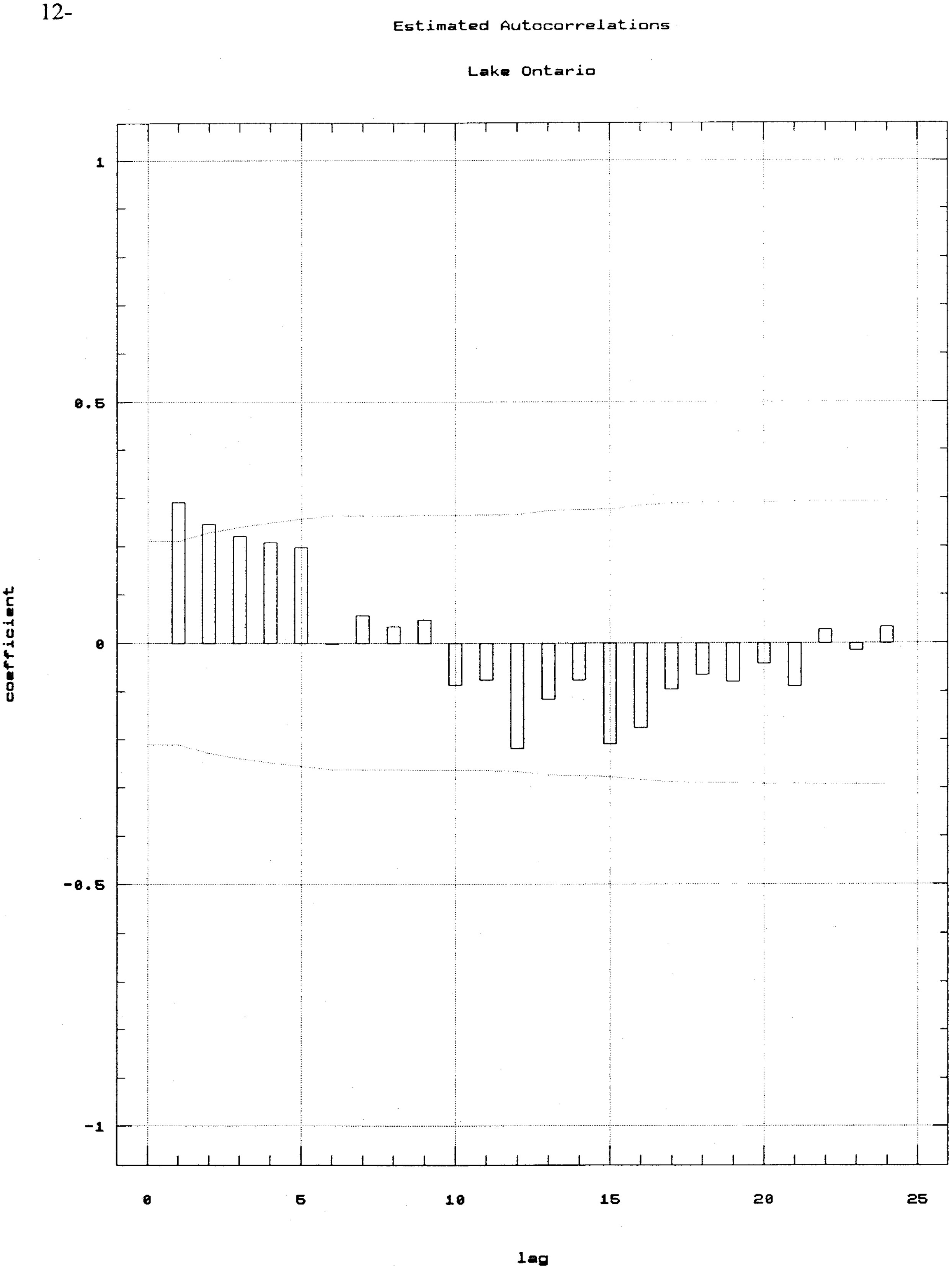


Figure 1.2 d: Lag-one to lag-24 correlogram for Lake Ontario.

Table 1.3 lists only the significant lag-1 to lag-4 cross-correlations at a 5% level of significance between annual NBS for the four Lakes.

Table 1.3: Significant lag-k cross-correlations between annual NBS for the four Lakes.

	SU (T)	MH (T)	ER (T)	ON (T)
lag-1				
SU (T-1)	-	0,32	-	0,28
ER (T-1)	-	0,21	-	0,22
ON (T-1)	-	-	0,24	-
lag-2				
ER (T-2)	-	-	-	0,22
lag-3				
ER (T-3)	-	-	-	0,23
ON (T-3)	-	-	0,26	-
lag-4				
SU (T-4)	-	-	0,20	-

Up to lag-4 some of the correlations are significant at a 5% level of significance, but in general the coefficients are low.

1.1.3 Frequency analysis and normality test

Figures 1.3a,b,c,d respectivly show the frequency plots of annual NBS using the Weibull plotting position formula on normal probability paper for Lakes Superior, Michigan-Huron, Erie and Ontario.

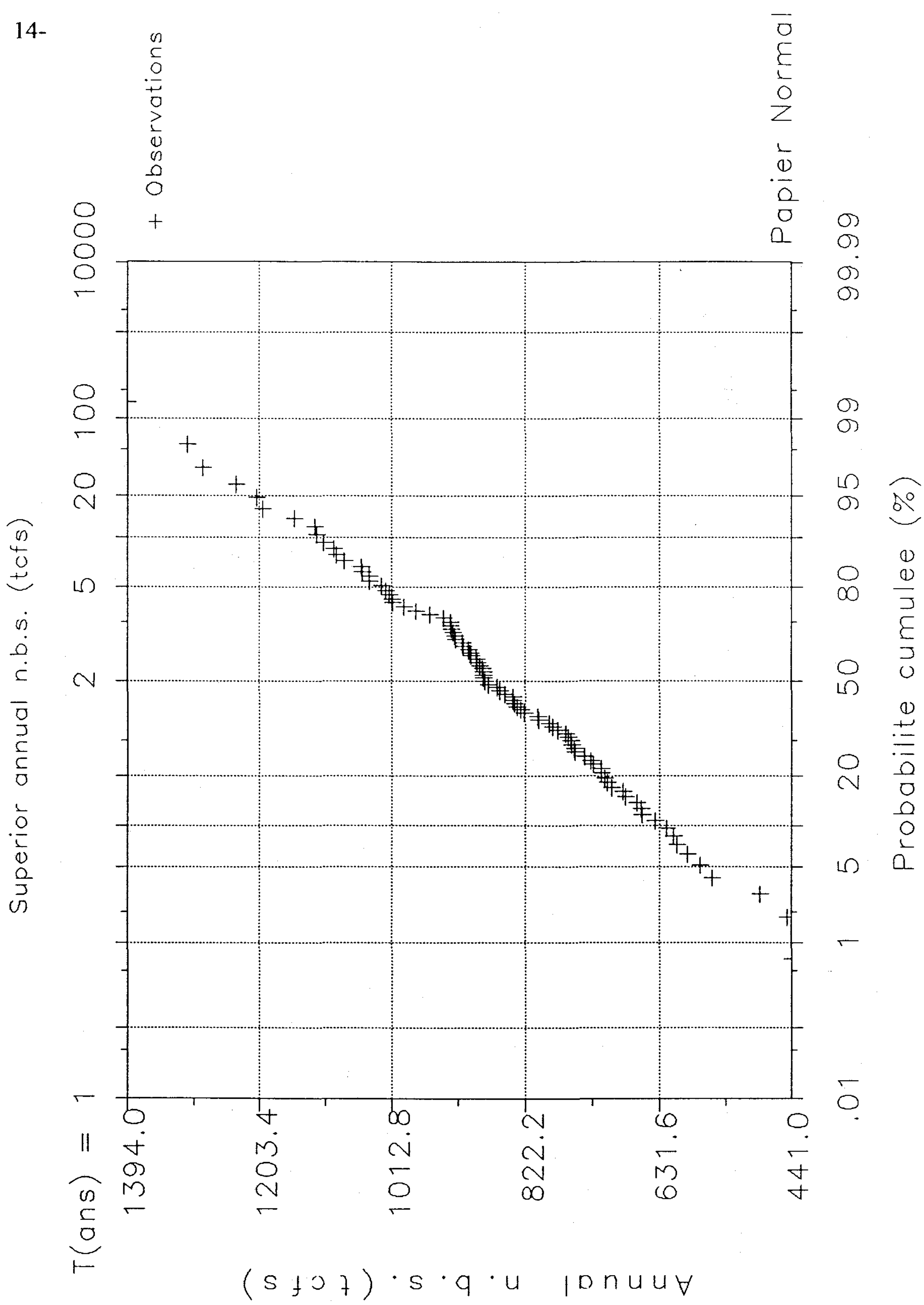


Figure 1.3 a: Frequency plot of annual NBS using the Weibull plotting position formula on normal probability paper for Lake Superior.

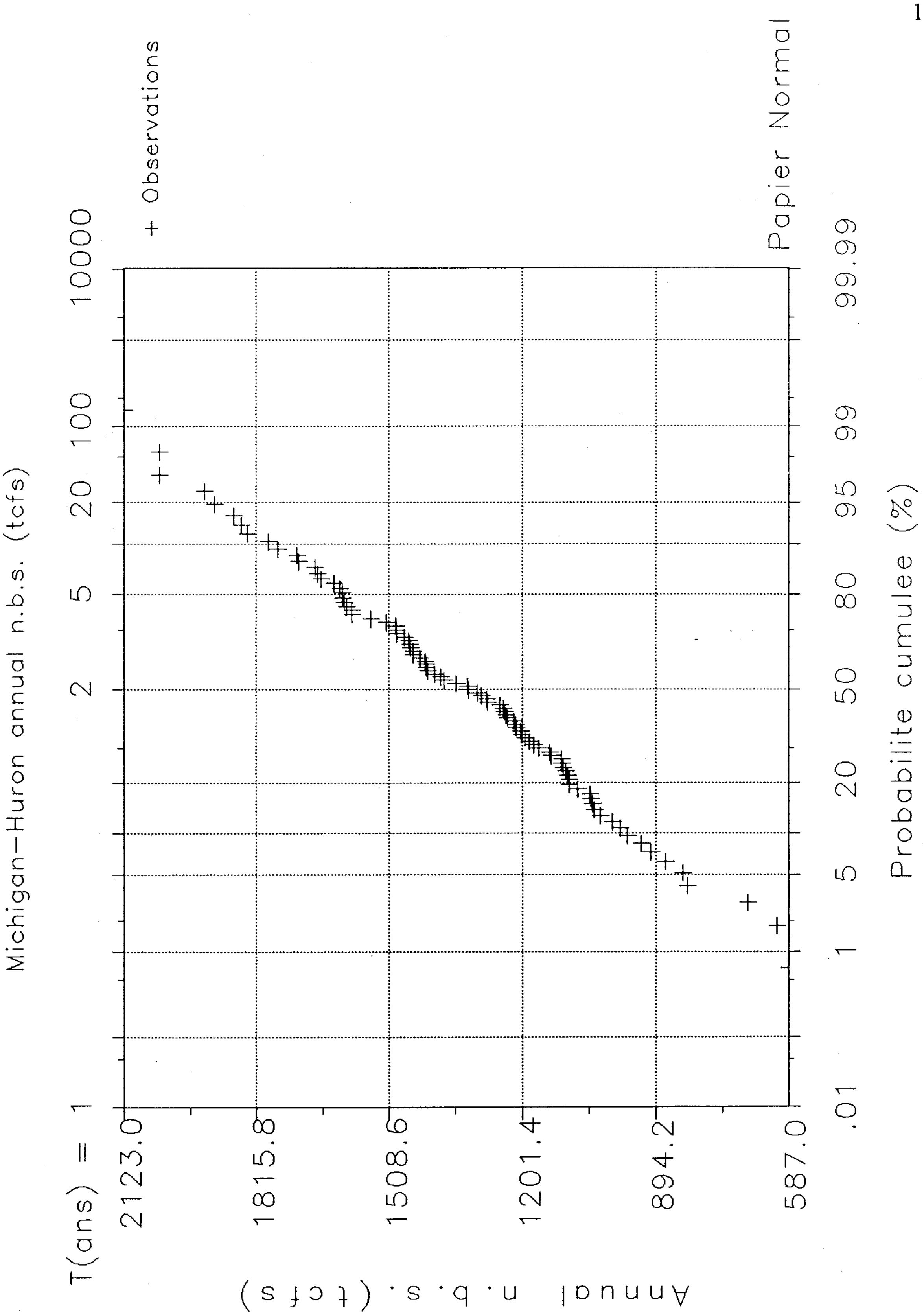


Figure 1.3 b: Frequency plot of annual NBS using the Weibull plotting position formula on normal probability paper for Lake Michigan-Huron.

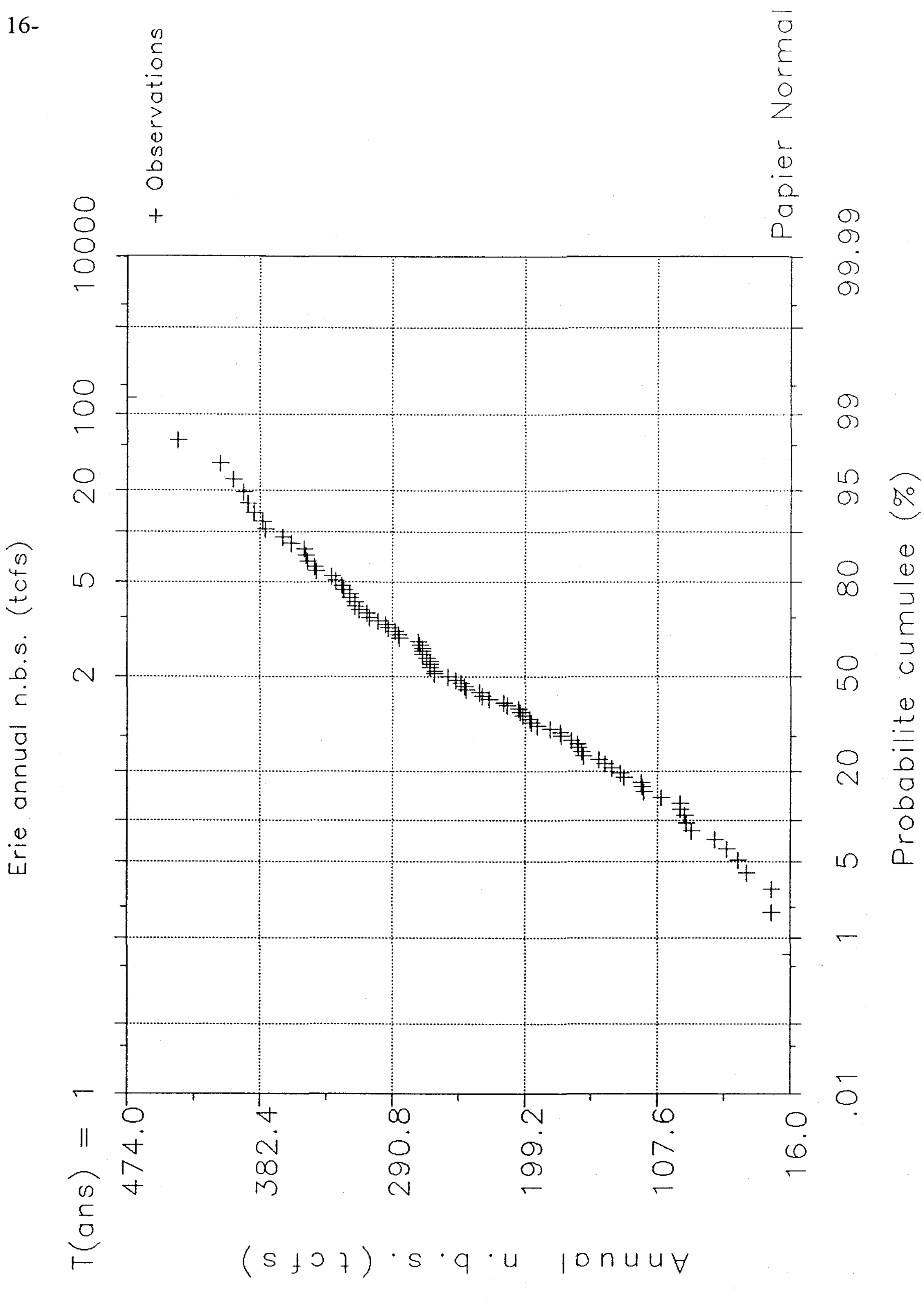


Figure 1.3 c: **Frequency plot of annual NBS using the Weibull plotting position formula on normal probability paper for Lake Erie.**

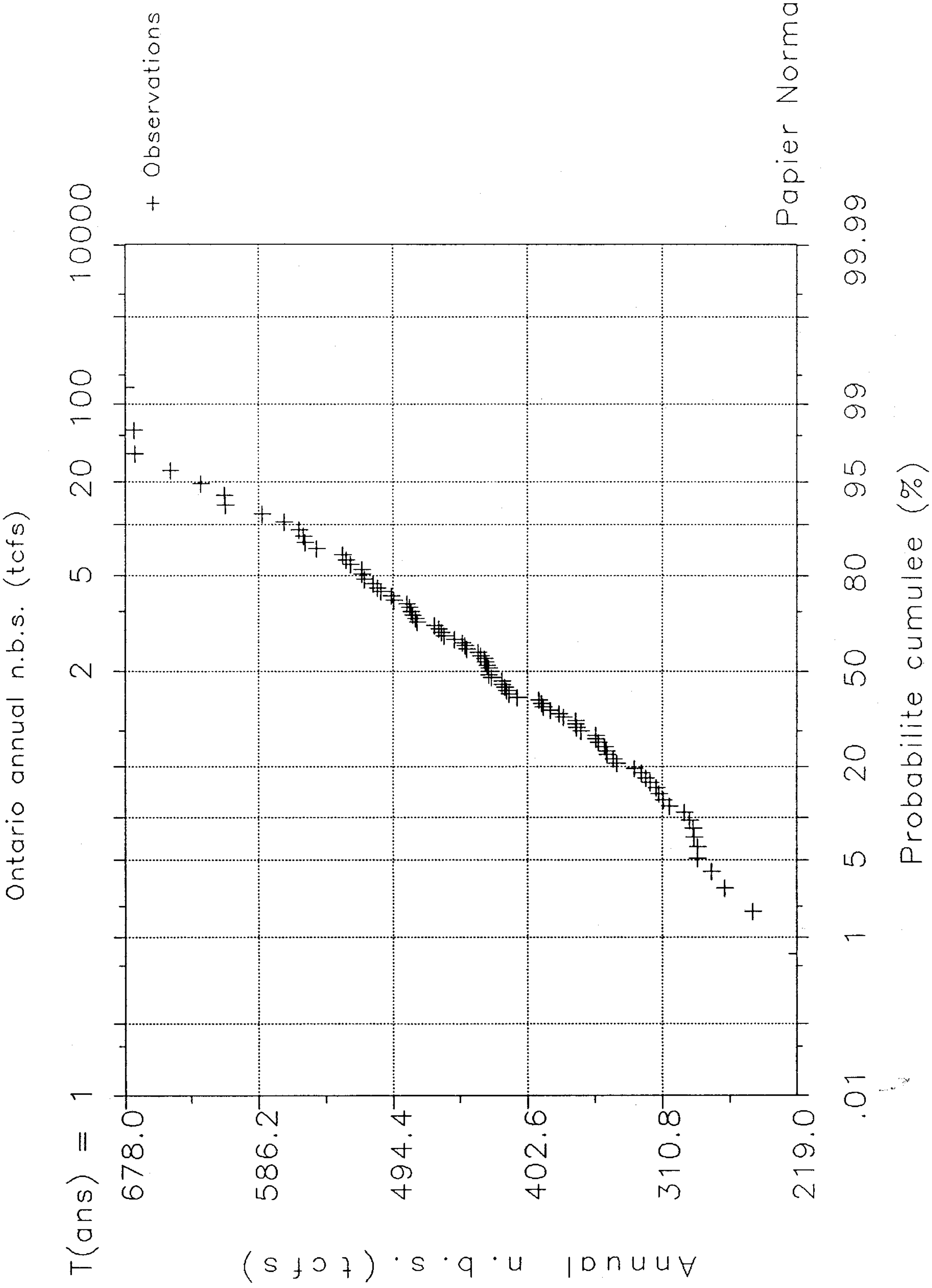


Figure 1.3 d: Frequency plot of annual NBS using the Weibull plotting position formula on normal probability paper for Lake Ontario.

Table 1.4 gives the result of the Kolmogorov-Smirnov goodness-of-fit test applied to the annual NBS series, where DN is the maximum absolute difference between the empirical distribution and the theoretical normal distribution with the parameters estimated from the observations. The P value is the probability of exceedance corresponding to the DN statistic obtained from the sample.

Table 1.4: Kolmogorov-Smirnov goodness-of-fit test.

	SU	MH	ER	ON
DN statistic	0,085	0,062	0,087	0,056
P value	0,539	0,885	0,515	0,940

Based on a Kolmogorov-Smirnov goodness-of-fit test (Table 1.4) and on the normal probability plots (Fig. 1.3) it can be assumed that annual NBS for the four Lakes are normally-distributed.

1.1.4 Run properties

The run properties used in this study are defined according to Salas and Boes (1980). Let $X_1,..., X_i,..., X_n$, be a sample of size n. A positive value of X_i - c, where c is a constant threshold, is called a surplus. A consecutive sequence of exactly L surplus is called a surplus run of length L, and the sum of the surplus X_i -c over such a run is called the surplus run sum. The maximum surplus run length (RL) is the longest of all the surplus run lengths in the sample. The largest surplus run sum in the sample is the maximum surplus run volume (RS). Table 1.5 gives the maximum surplus run length (RL) and the maximum surplus run volume (RS) of historical data for a truncation level (or constant threshold c) equal to the historical annual NBS mean. Both the traditional definition (one year) and the 4-year running average run criteria were used. In Table 1.5 RS is given in Hm^3 x 10^6 units.

Table 1.5: Maximum surplus run length (RL) and run volume (RS) of historical data.

	SU	MH	ER	ON
RL (year) 1-year	7	9	15	10
RL (year) 4-year	18	22	22	18
RS ($Hm^3 \times 10^6$) 1-year	0,767	1,708	1,661	0,994
RS ($Hm^3 \times 10^6$) 4-year	1,414	3,531	1,724	1,525

Table 1.5 shows that the values of RL and RS are high, especially for Lakes Michigan-Huron and Erie.

1.1.5 Hurst coefficient

The Hurst coefficient (H) (Hurst 1951, Boes and Salas 1978) is used as an indicator of the persistence of a time series process. Persistence in streamflow is defined as a tendency for low flows to follow low flows and high flows to follow high flows. The Hurst phenomenon arises due to non-normality of the data, serial correlation and nonstationarity in the underlying mean of the process. Asymptotically, for large normal independent series H = 0,5. Hurst has shown that the average value of H, observed for a large number of hydrological time series, is approximately 0,73.

The observed Hurst coefficients for the annual NBS of the four Lakes are:

- 0,647 for Lake Superior;
- 0,731 for Lake Michigan-Huron;
- 0,757 for Lake Erie, and;
- 0,752 for Lake Ontario.

The four sites have a high value of the Hurst coefficient, which indicates long-term persistence in annual NBS.

1.1.6 Spectral analysis

In the spectral analysis, the observed time series is considered as a random sample of a process in time that is made up of oscillations of all possible frequencies. The variance spectrum partitions the variance into a number of intervals or bands of frequency (f). The spectral density is the amount of variance per interval of frequency. For a completely random series of uncorrelated numbers, the spectral density function (G(f)) is constant and is termed white noise. This indicates that no frequency interval contains any more variance than any other frequency interval.

Figure 1.4 shows the variance spectra of Great Lakes standardized series of annual NBS.

For the four Lakes, some frequencies contain more variance than the others (G(f) is not constant). For Lakes Superior, Michigan-Huron, Erie and Ontario there is a small peak in the amplitude of the frequencies occuring respectively at f = 0,11, 0,16, 0,25 and 0,20. These peaks show that there is more variance in the standardized NBS for cycle periods of approximately 9 and 6 years for Lake Superior and Lake Michigan-Huron, and of 5 and 4 years for Lake Erie and Lake Ontario. For Lakes Erie and Ontario the cycle periods of the maximum variance spectra correspond to the observed lag-5 and lag-4 autocorrelation coefficients. Thus, maximum variance spectra of the Great Lakes may indicate some long-term persistence in the annual NBS.

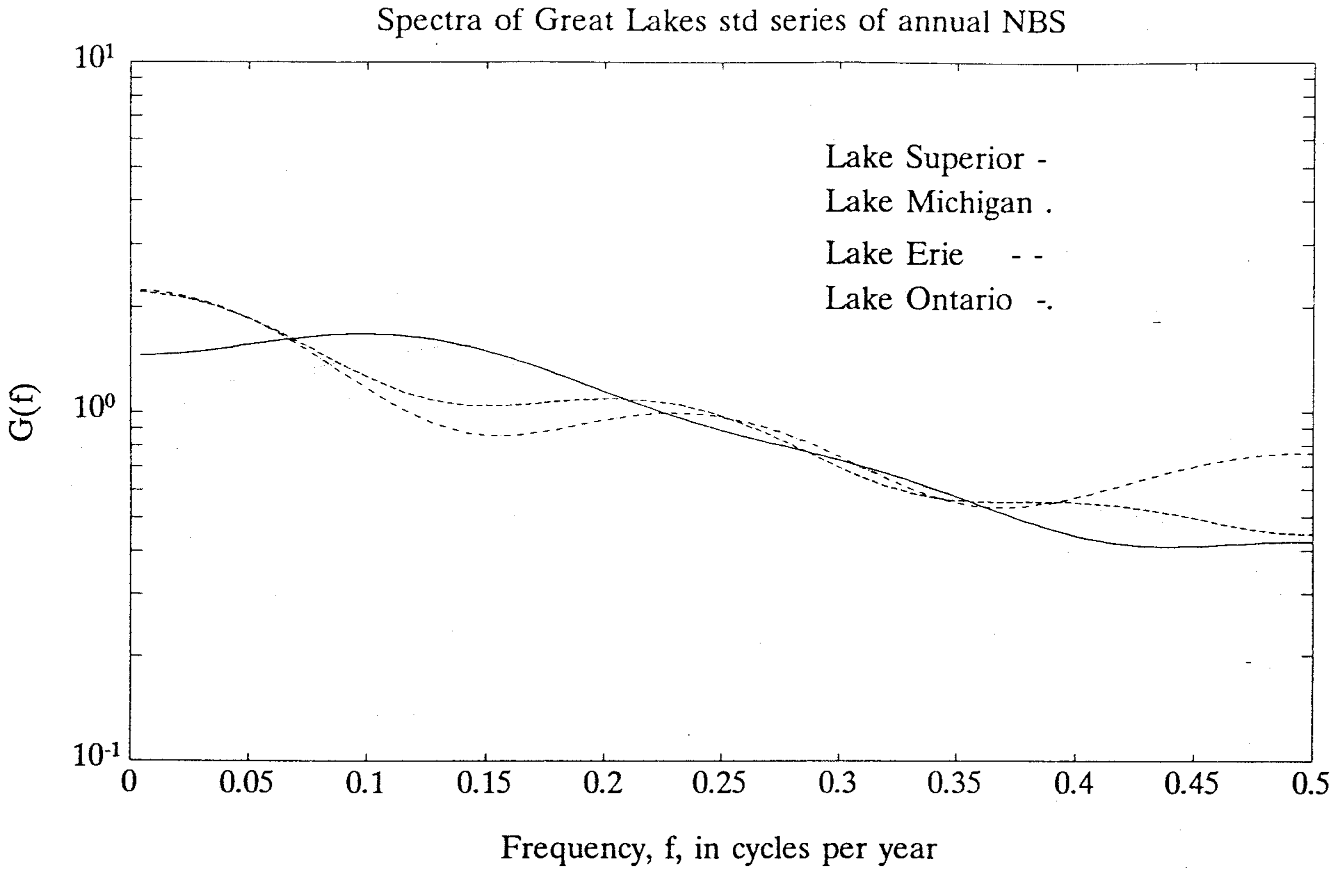


Figure 1.4: Variance spectra of Great Lakes standardized series of annual NBS.

1.1.7 Homogeneity

Figures 1.5a and 1.5b give the series of the annual NBS respectively for Lakes Superior and Michigan-Huron. The examination of figure 1.5 shows that the NBS for Lakes Superior and Michigan-Huron appear to be stationary. On the other hand, NBS series of Lake Erie (Fig. 1.6a) and Lake Ontario (Fig. 1.7a) show a positive jump or shift in the late sixties early seventies. Statistical tests have been performed to see whether these shifts are natural or man-made or a combination of both. Previous analysis of the data by the GLERL appear to indicate that such positive shifts may be due to similar shifts in the precipitation regime. It was decided to include in the analyses the series of annual over-basin precipitation. Figures 1.6b and 1.7b show these series for Lakes Erie and Ontario. Visual examination of figures 1.6a, b and 1.7a, b shows that there is a positive shift in both NBS and precipitation series. A bayesian approach was used to detect the most probable year of the occurence of the shift (Lee and Heghinian 1977, Bruneau and Rassam 1983) in the precipitation and NBS series.

The most significant change in the mean of precipitation and NBS series for Lakes Erie and Ontario occures simultaneously in or close to 1970. Since there is a concordance between the shifts in NBS and precipitation it can be assumed that the observed shifts in NBS are natural and that these shifts occured close to 1970. Based on this information a Mann-Whitney test of homogeneity (Mann and Whitney 1947) has been carried out on the four NBS series to see whether differences between the means of two subsamples (1900 to 1969 and 1970 to 1989) are significant. The results of the test of homogeneity for Lakes Superior and Michigan-Huron indicate that in both cases the two subsamples are homogenious at the 1% level of significance. For Lakes Erie and Ontario there is a significant difference between the means at a 1% level of significance. Thus, results of the bayesian approach and homogeneity test indicate that there is a shift in the annual NBS for Lake Erie and Lake Ontario. Based on this type of analysis it can be assumed that the shifts are an inherent property of the series which has to be taken into account in the stochastic model selection for the simulation.

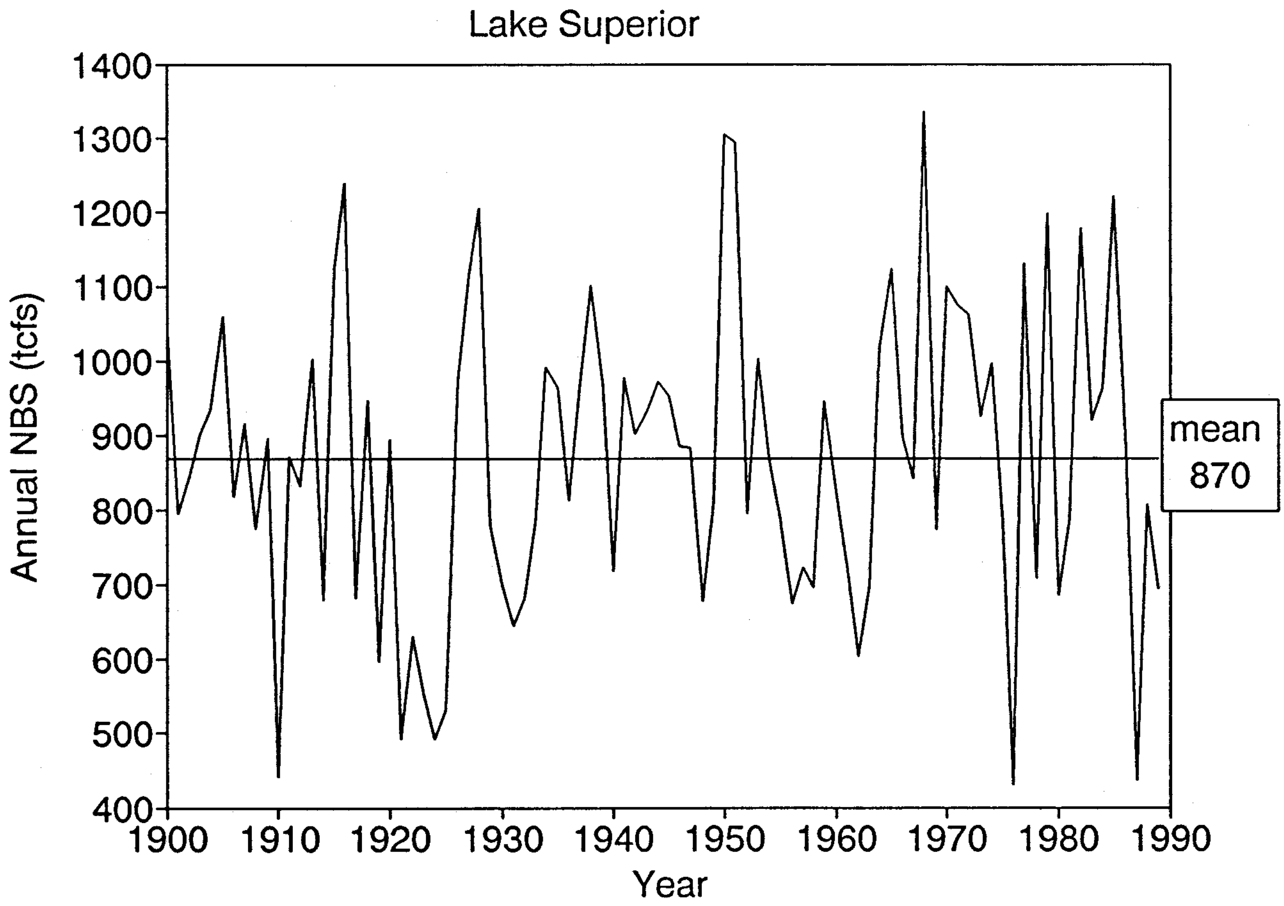


Figure 1.5 a: Annual net basin supply for Lake Superior.

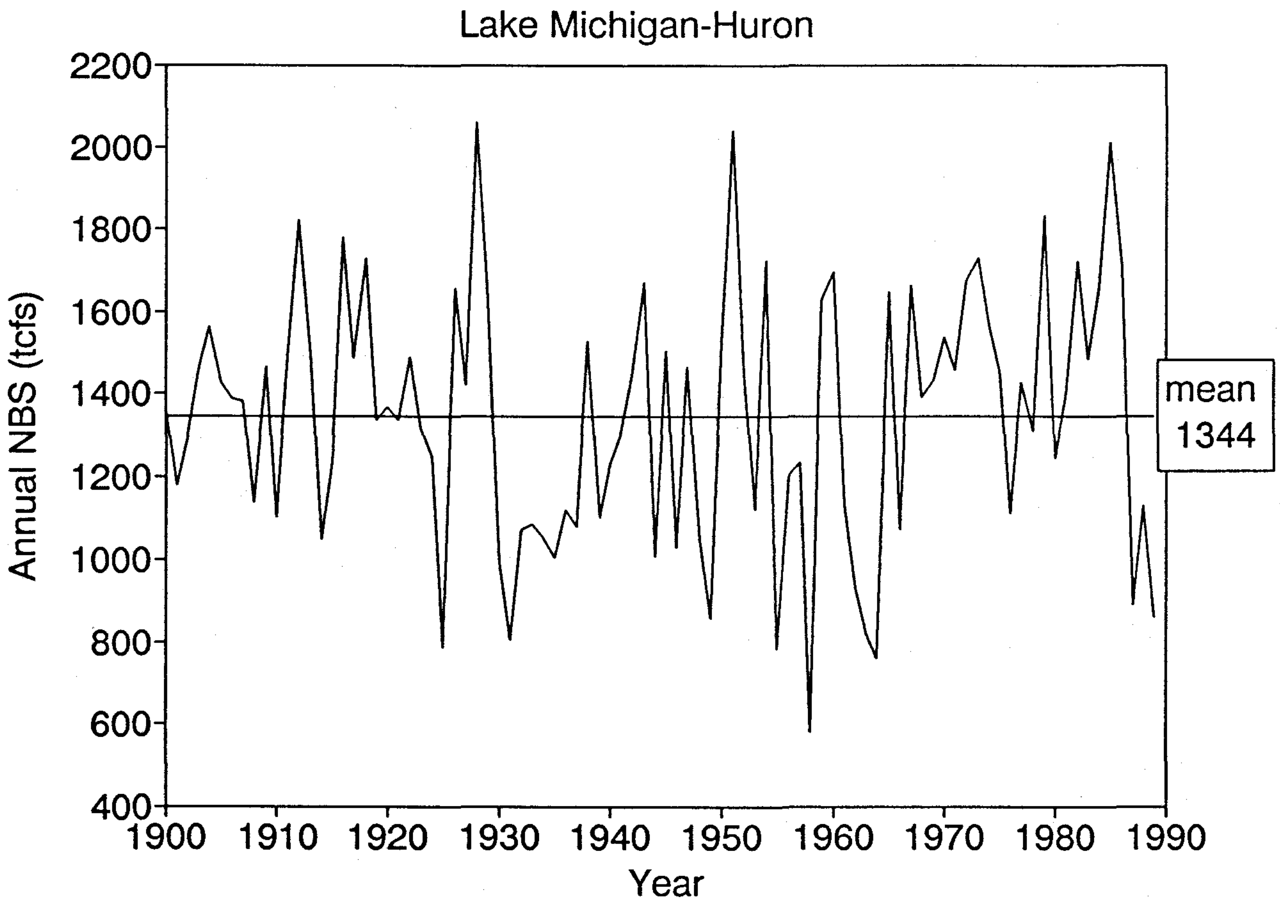


Figure 1.5 b: Annual net basin supply for Lake Michigan-Huron.

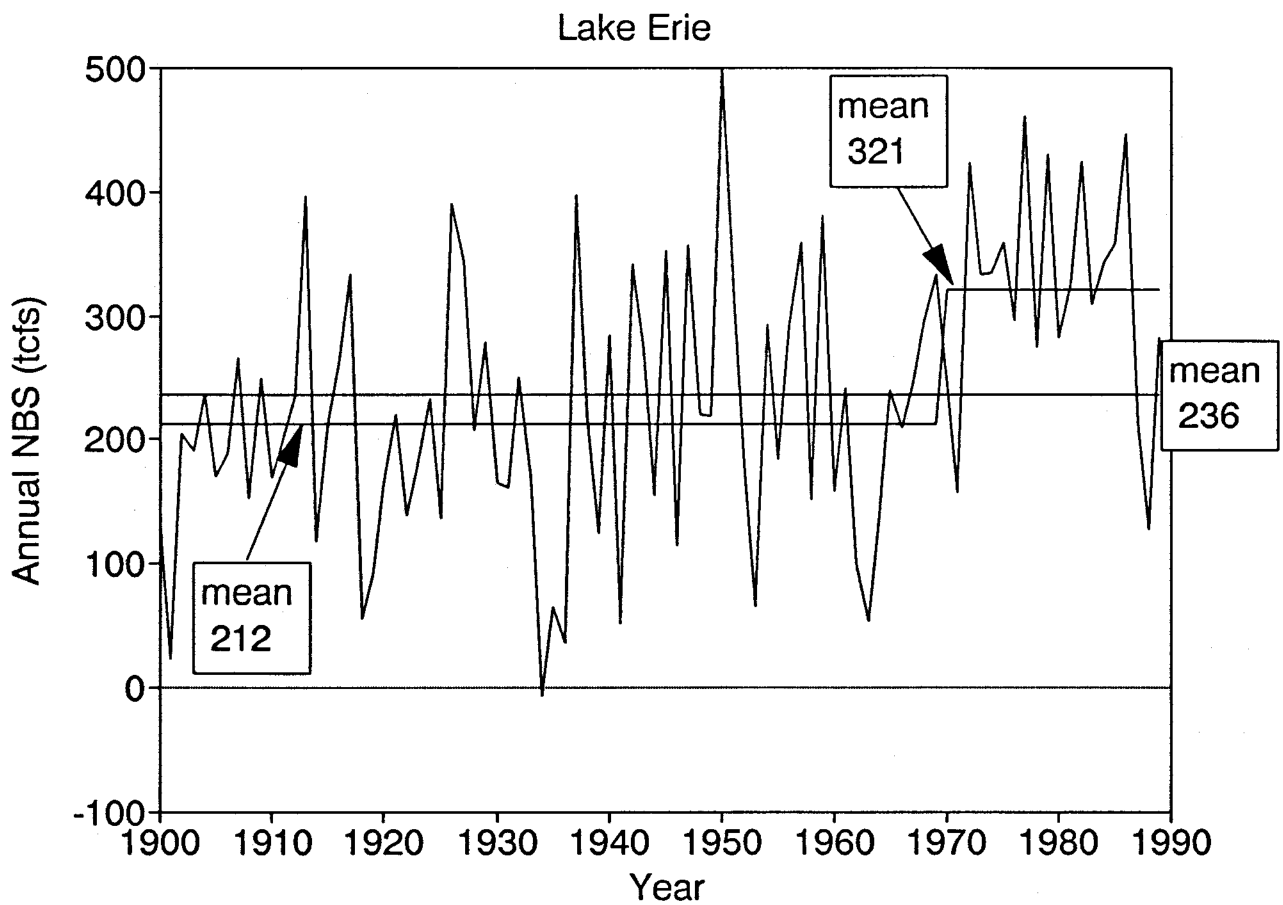


Figure 1.6 a: Annual net basin supply for Lake Erie.

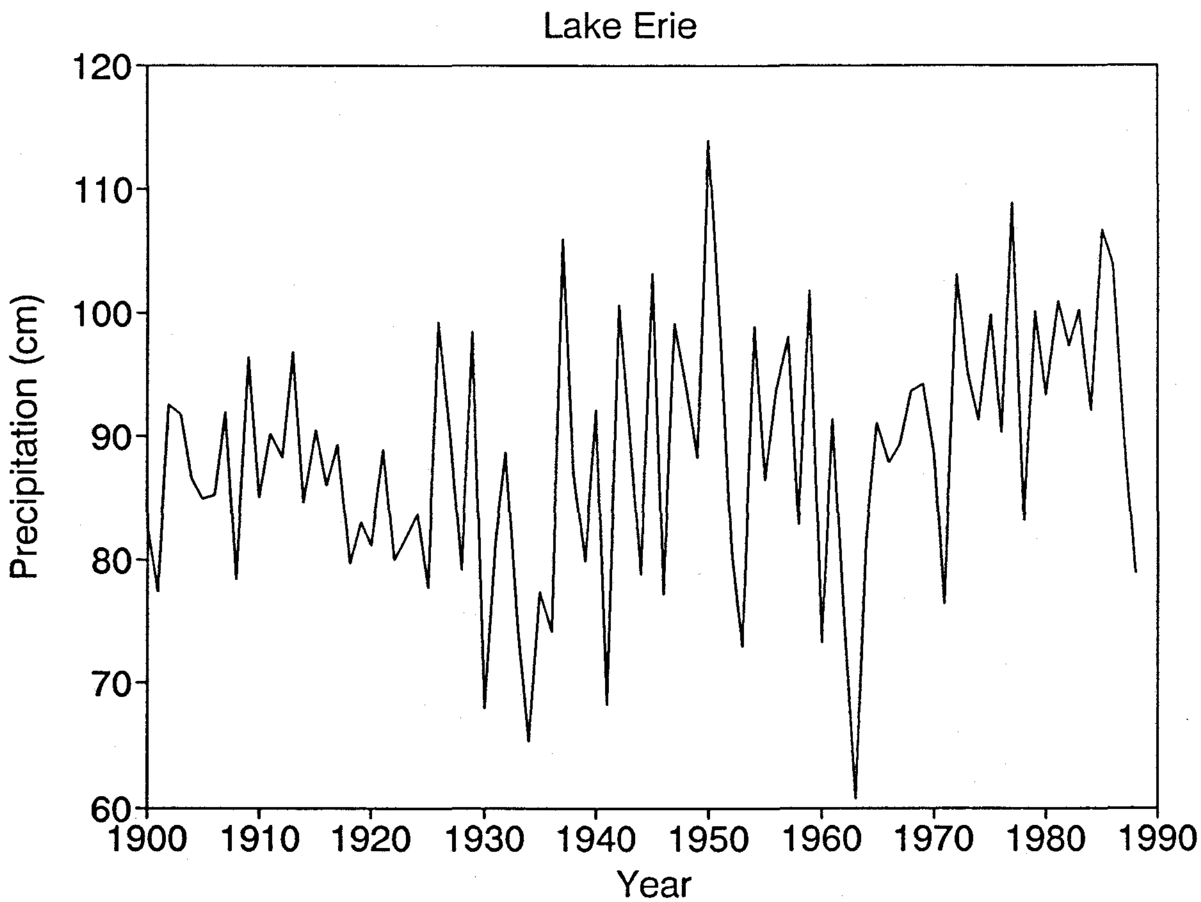


Figure 1.6 b: Annual over basin precipitation for Lake Erie.

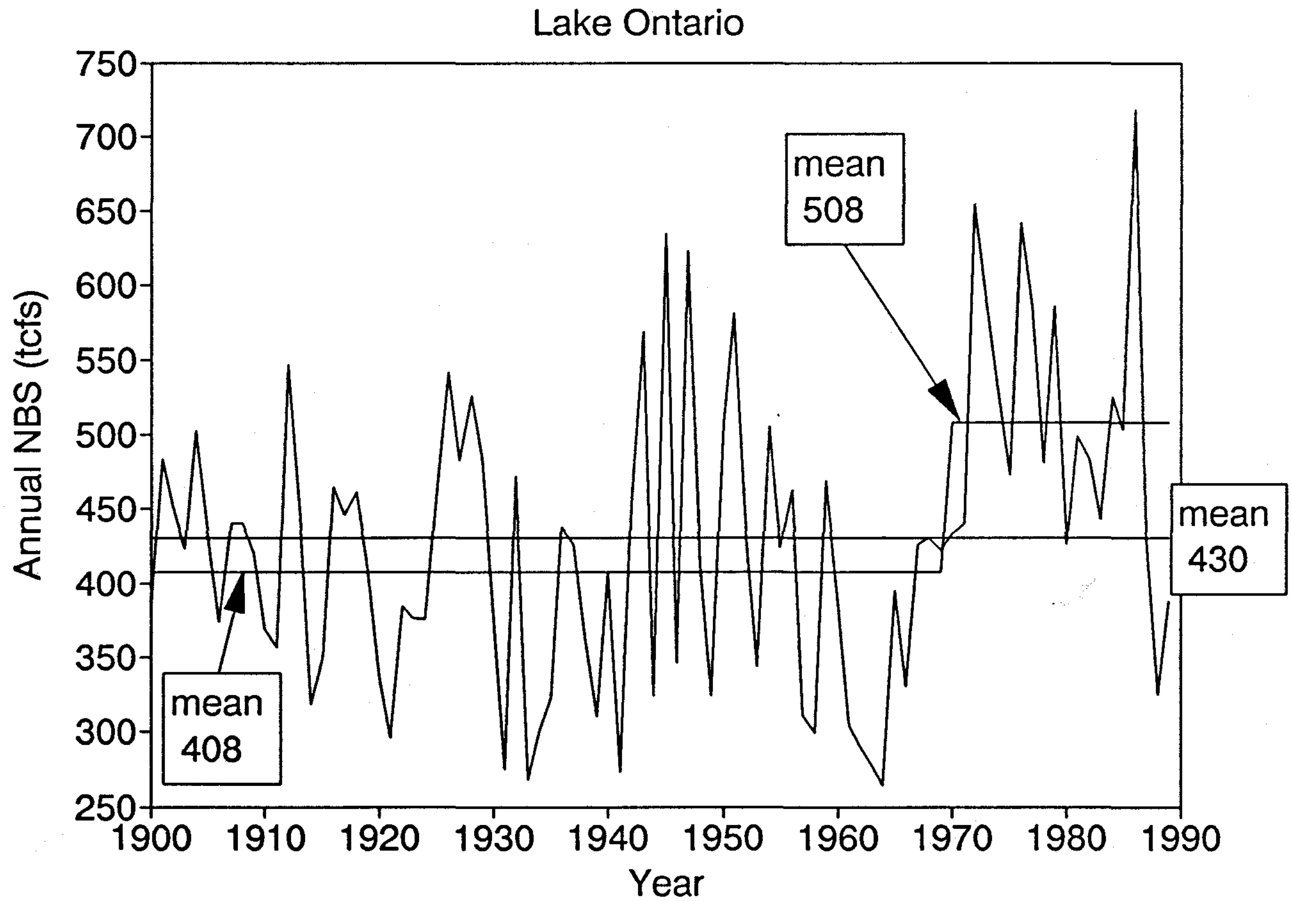


Figure 1.7 a: Annual net basin supply for Lake Ontario.

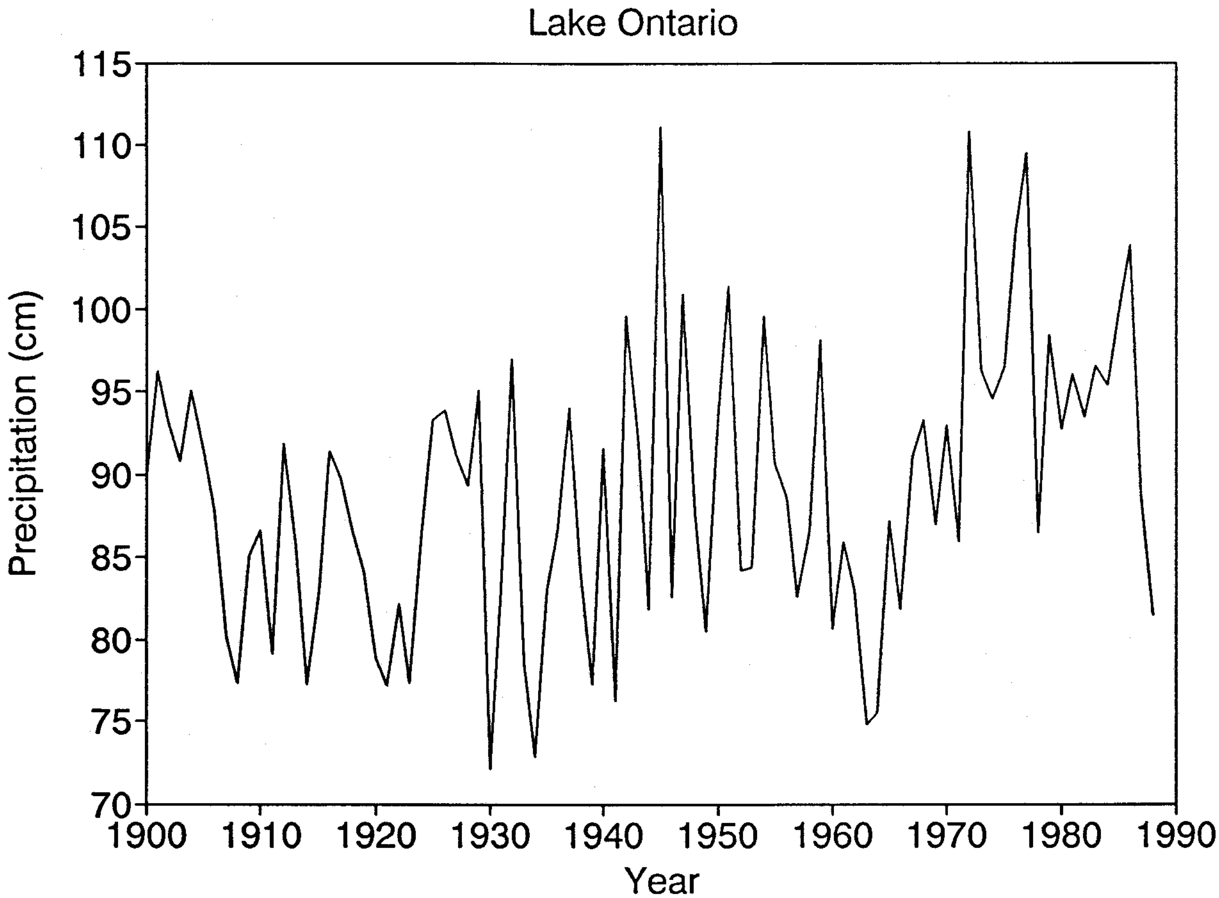


Figure 1.7 b: Annual over basin precipitation for Lake Ontario.

Results presented in the previous sections also indicate that long-term persistence is an important property of the series. This is shown by the high values of the run properties (Section 1.1.4), high values of the Hurst coefficient (Section 1.1.5) and peaks in the variance spectra (Section 1.1.6). It has also been noticed that some of these peaks coincide with high order autocorrelation coefficients (Section 1.1.1).

Since unrealistic autocorrelations can result from the presence of a shift in the series (Salas and Boes 1980, Salas *et al.* 1981), the validity of these statistics showing long-term persistence for Lakes Erie and Ontario may be questionable. Due to the uncertainties induced in the series by the presence of a shift, the results can not be directly explained in term of persistence. Nevertheless, these characteristics (autocorrelation, RL, RS, Hurst coefficient, variance spectra) can be used to describe the data. Results of the normality test for Lakes Erie and Ontario (Section 1.1.3) may also be influenced by the non-homogeneity of the data induced by the shift (see Section 2.2.4).

On the other hand, for Lakes Superior and Michigan-Huron the results presented in the previous sections indicate that long-term persistence is an important characteristic of the series which has to be taken into account in the validation of the simulated data.

1.2 Monthly series

Time series of monthly NBS for the period 1900-1989 were obtained and analyzed statistically. Figures 1.8a,b,c,d and Table 1.6 respectivly show the mean, standard deviation and skewness coefficient of monthly NBS for the four lakes.

Mean peaks of monthly NBS occure in March on Lake Erie, in April on Lakes Ontario and Michigan-Huron, and in May on Lake Superior.

Most of the monthly series have a large coefficient of skewness, which indicate that the monthly observations are not normally-distributed. Thus, data transformation would probably be needed (see section 2.2.4).

1.2.1 Month-to-month correlation

Figures 1.9a,b,c,d show the lag-1 to lag-3 month-to-month correlation coefficients for the four lakes.

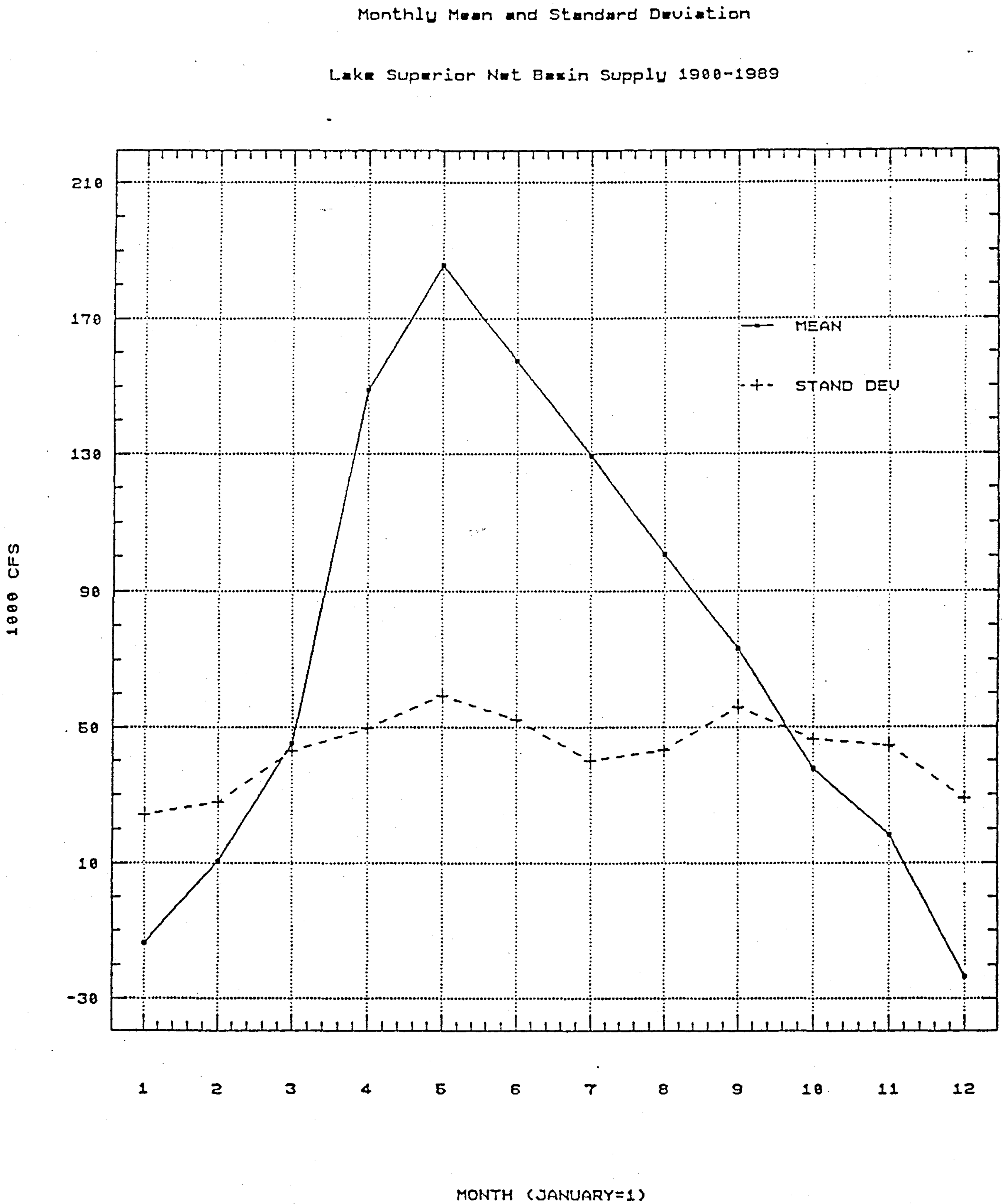


Figure 1.8 a: Mean and standard deviation of monthly NBS (tcfs) for Lake Superior.

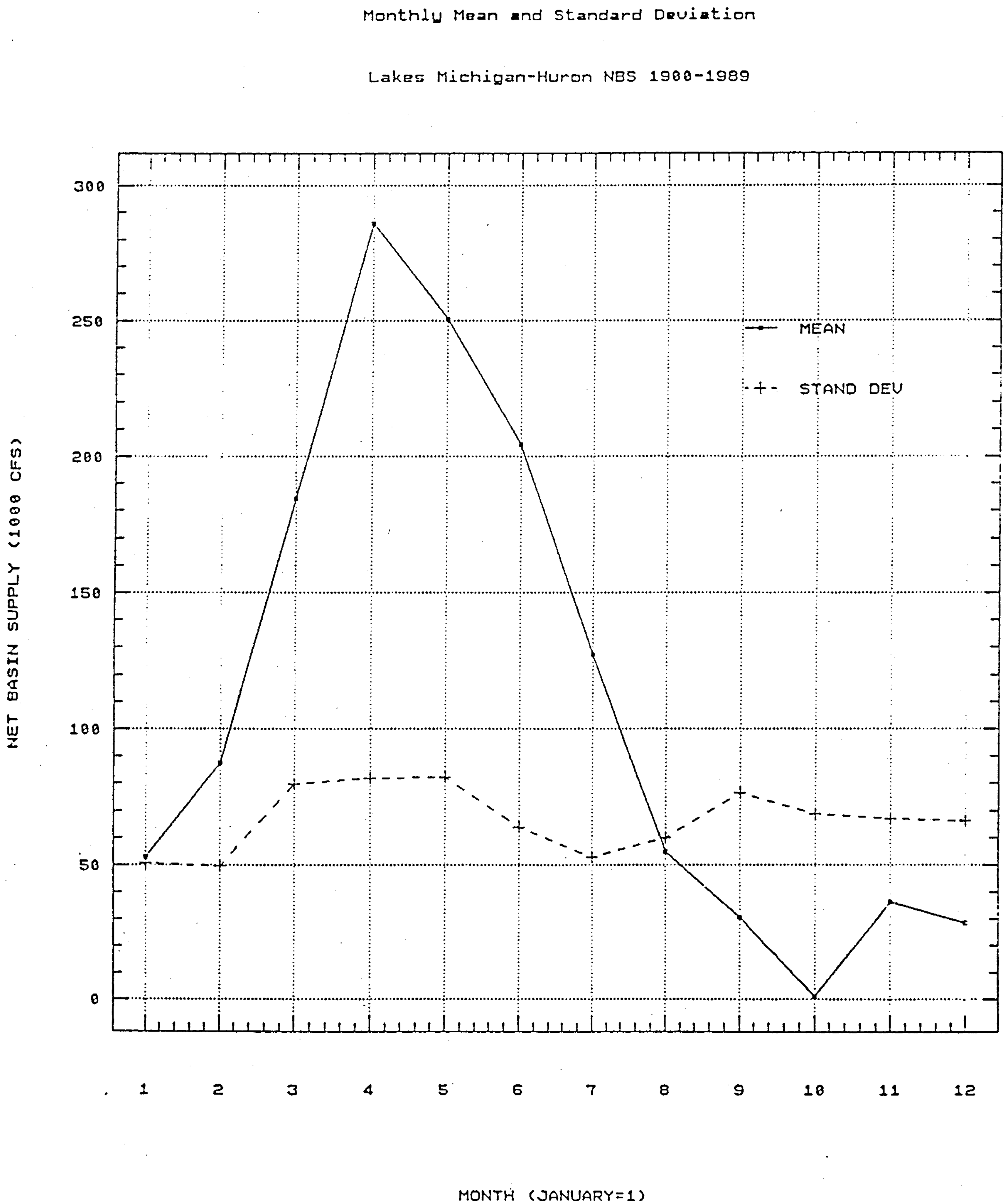


Figure 1.8 b: Mean and standard deviation of monthly NBS (tcfs) for Lake Michigan-Huron.

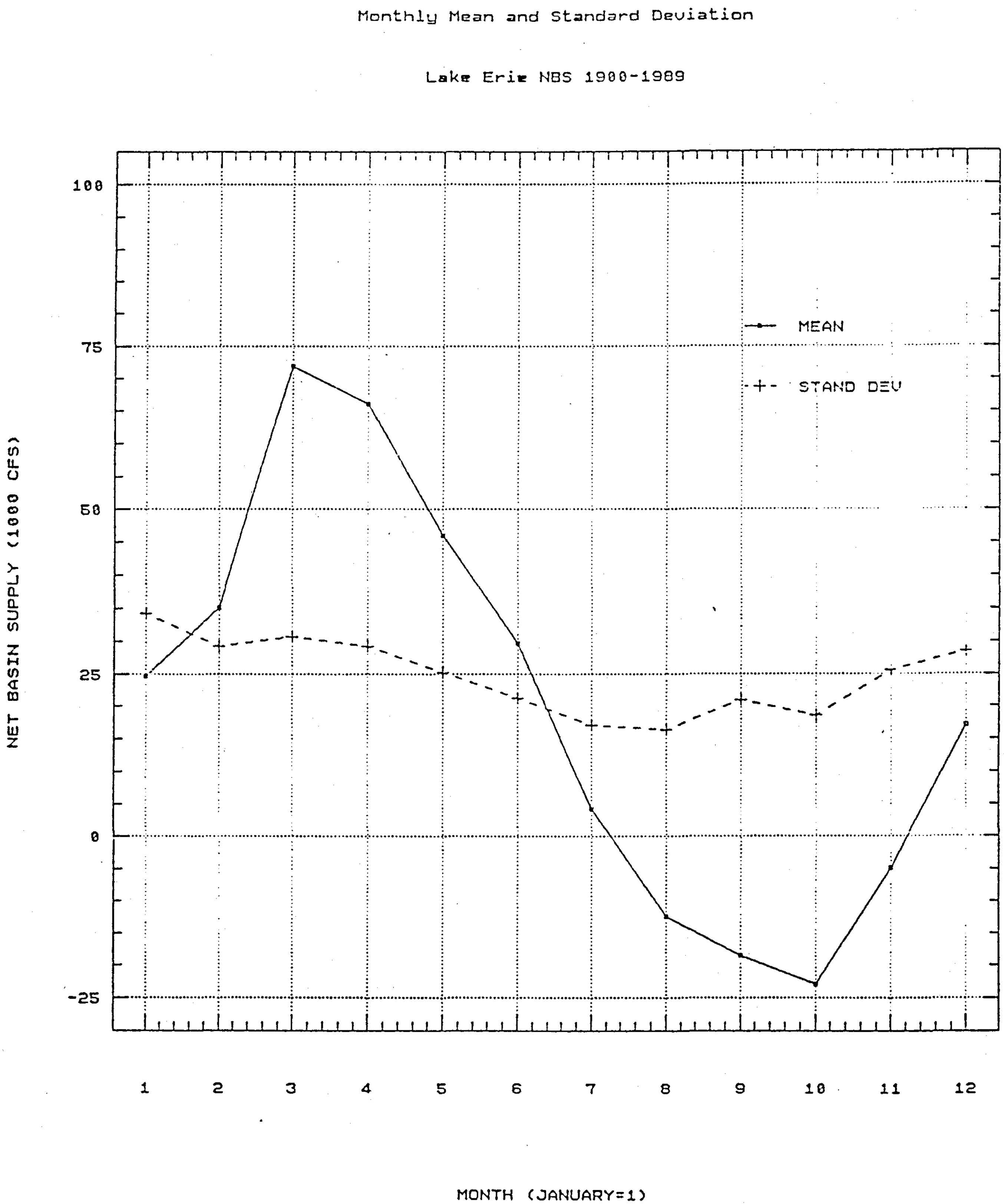


Figure 1.8 c: Mean and standard deviation of monthly NBS (tcfs) for Lake Erie.

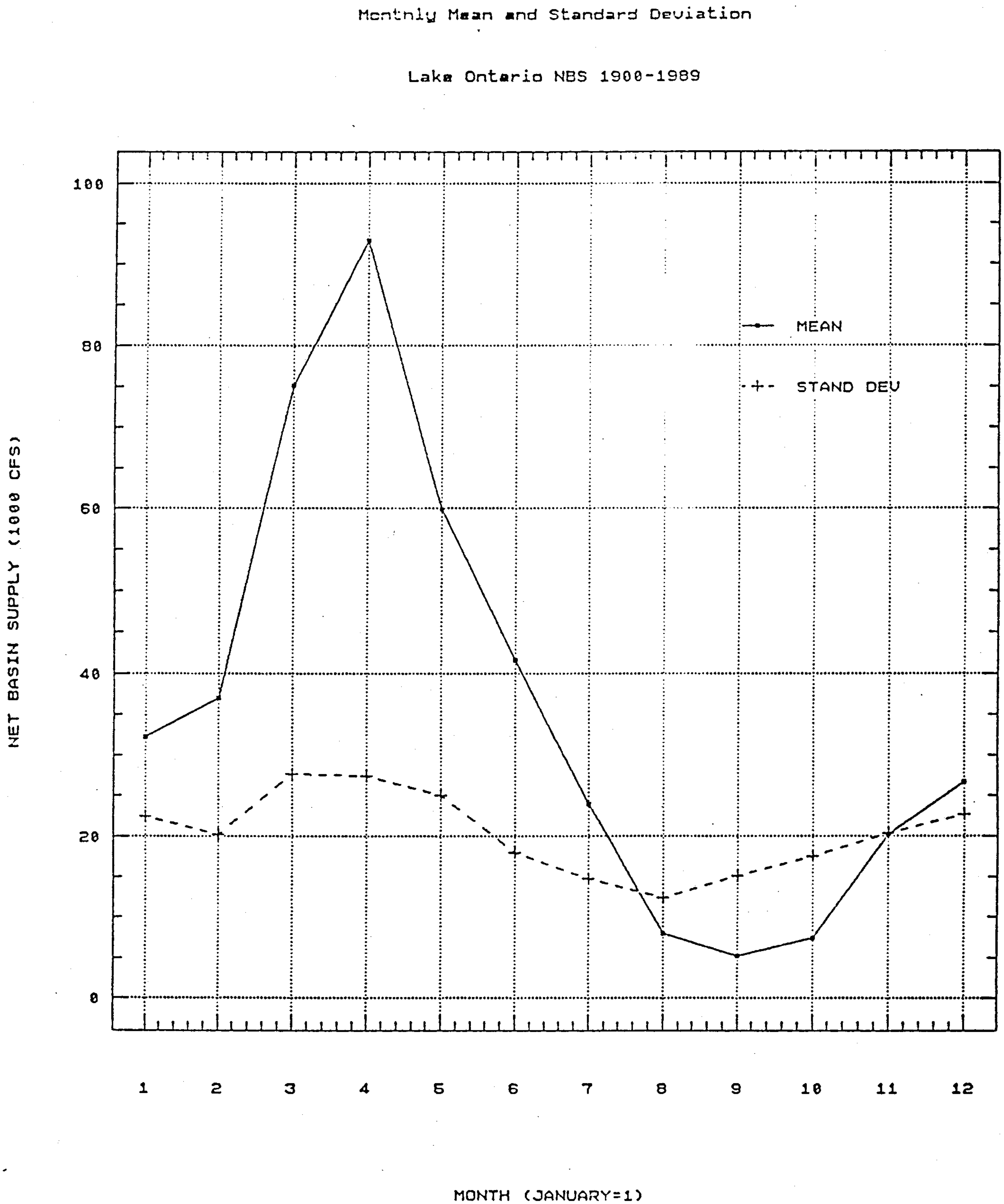


Figure 1.8 d: Mean and standard deviation of monthly NBS (tcfs) for Lake Ontario.

Table 1.6: Skewness coefficients of monthly NBS for the four lakes.

	SU	MH	ER	ON
Jan	0,217	0,563	1,164	0,960
Feb	0,479	0,349	0,275	0,559
Mar	0,588	0,398	0,601	0,611
Apr	0,378	0,406	0,043	0,066
May	0,229	0,478	0,852	1,079
Jun	0,407	0,469	0,677	1,262
Jul	0,582	0,601	1,093	0,970
Aug	0,456	0,278	1,139	0,705
Sep	0,806	1,057	1,515	1,307
Oct	0,181	0,739	0,969	0,935
Nov	0,571	0,328	1,171	0,726
Dec	-0,127	0,314	0,239	0,435

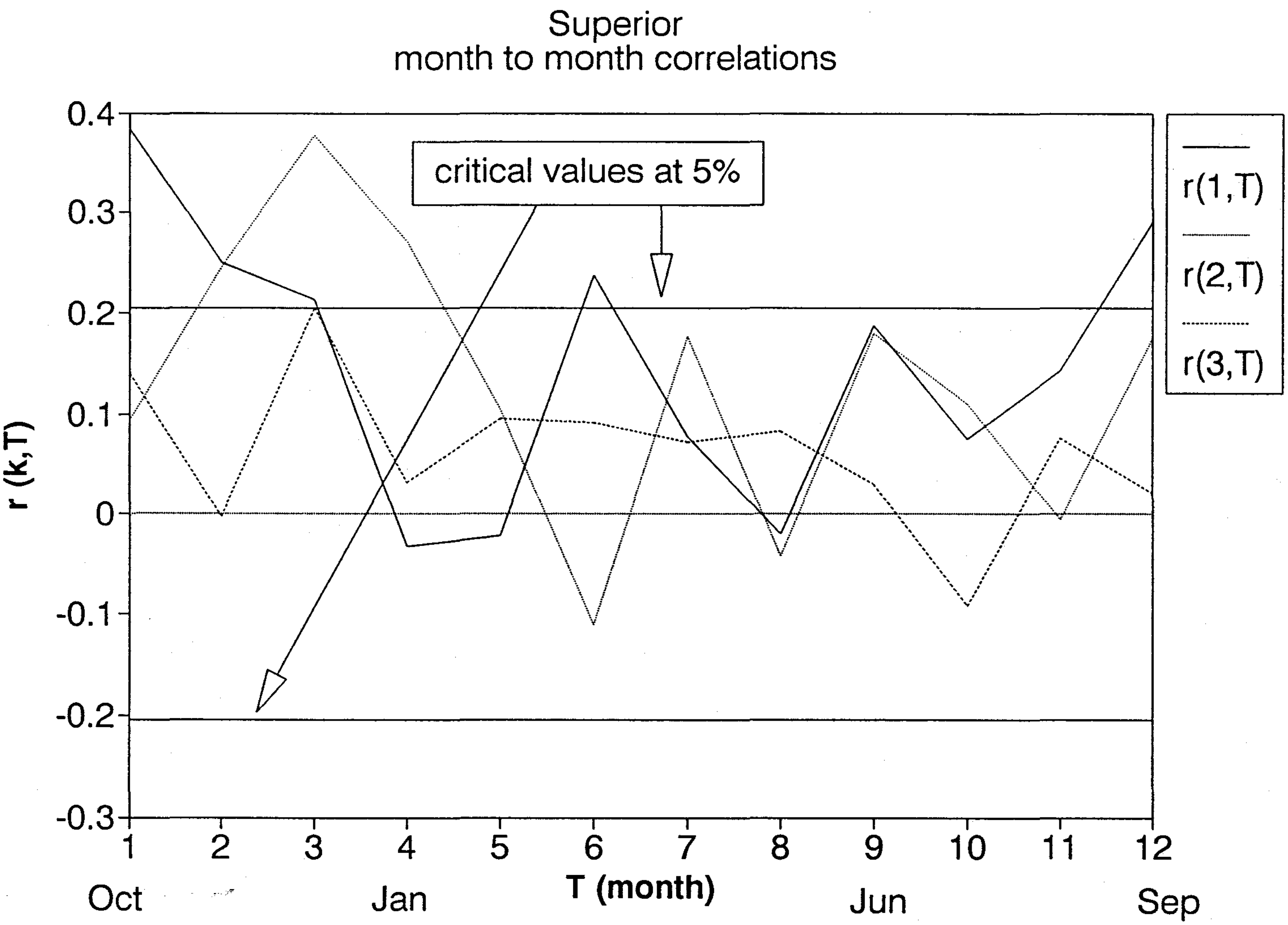


Figure 1.9 a: Lag-1 to lag-3 month-to-month correlation coefficients for Lake Superior.

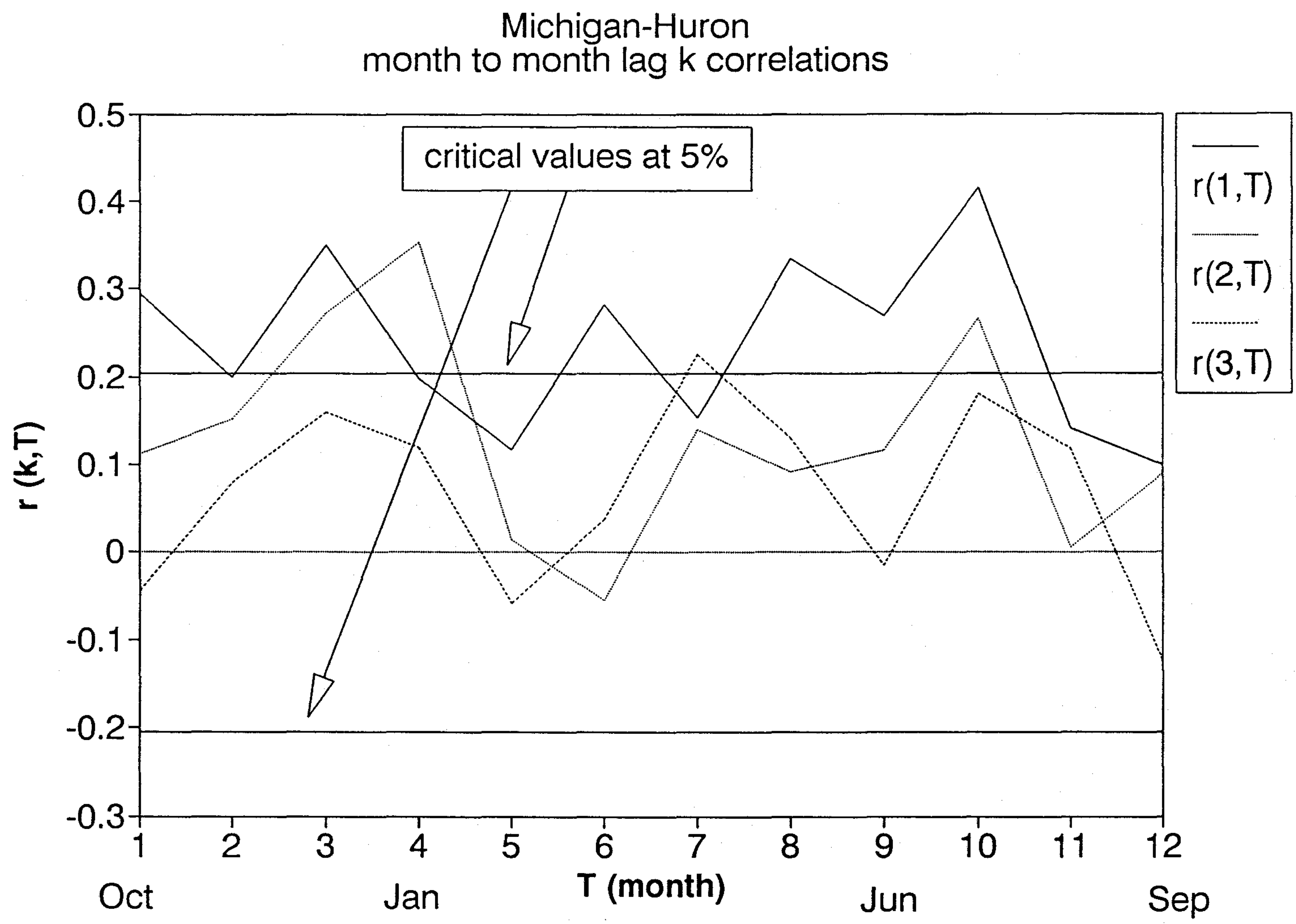


Figure 1.9 b: Lag-1 to lag-3 month-to-month correlation coefficients for Lake Michigan-Huron.

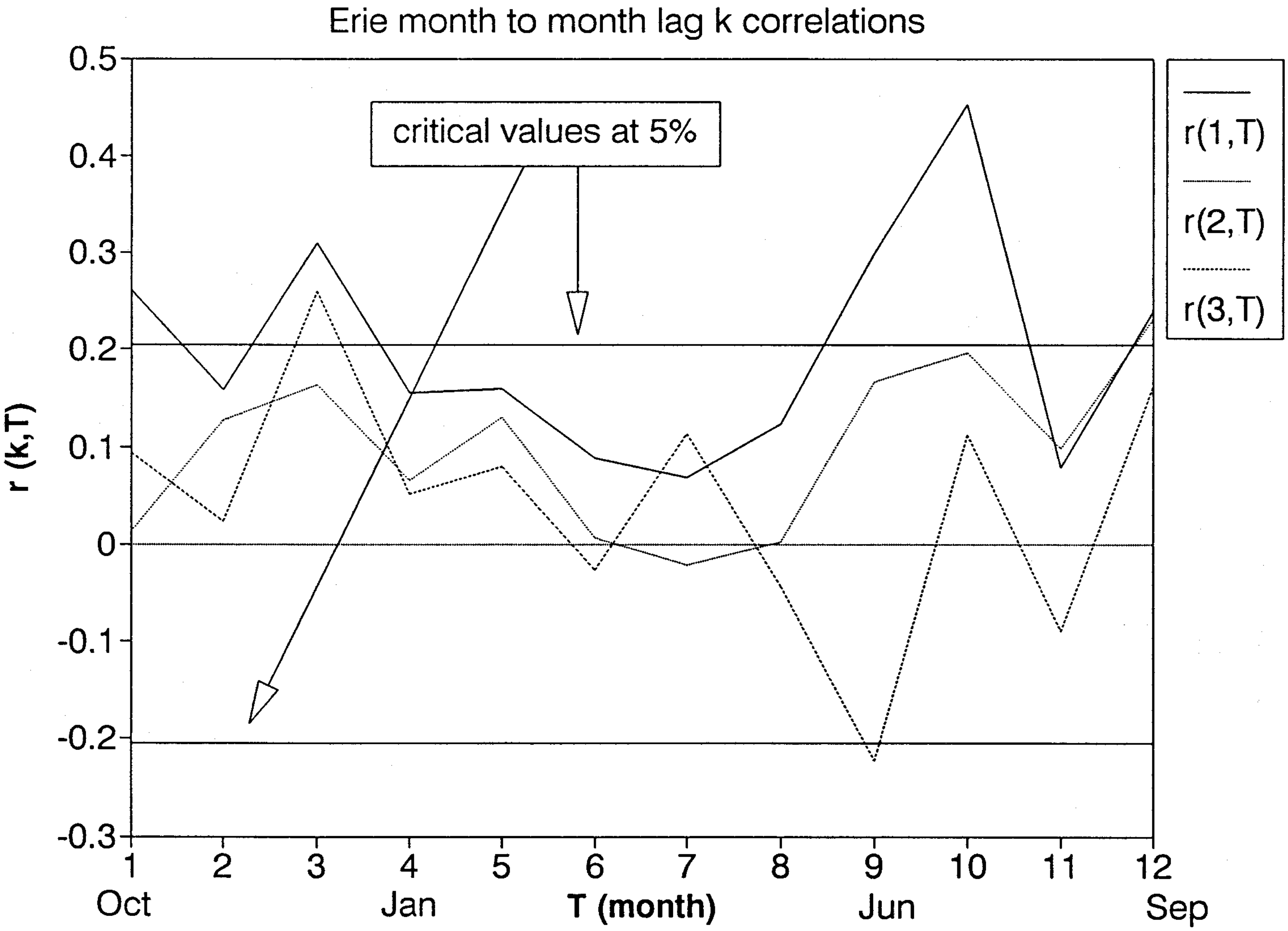


Figure 1.9 c: Lag-1 to lag-3 month-to-month correlation coefficients for Lake Erie.

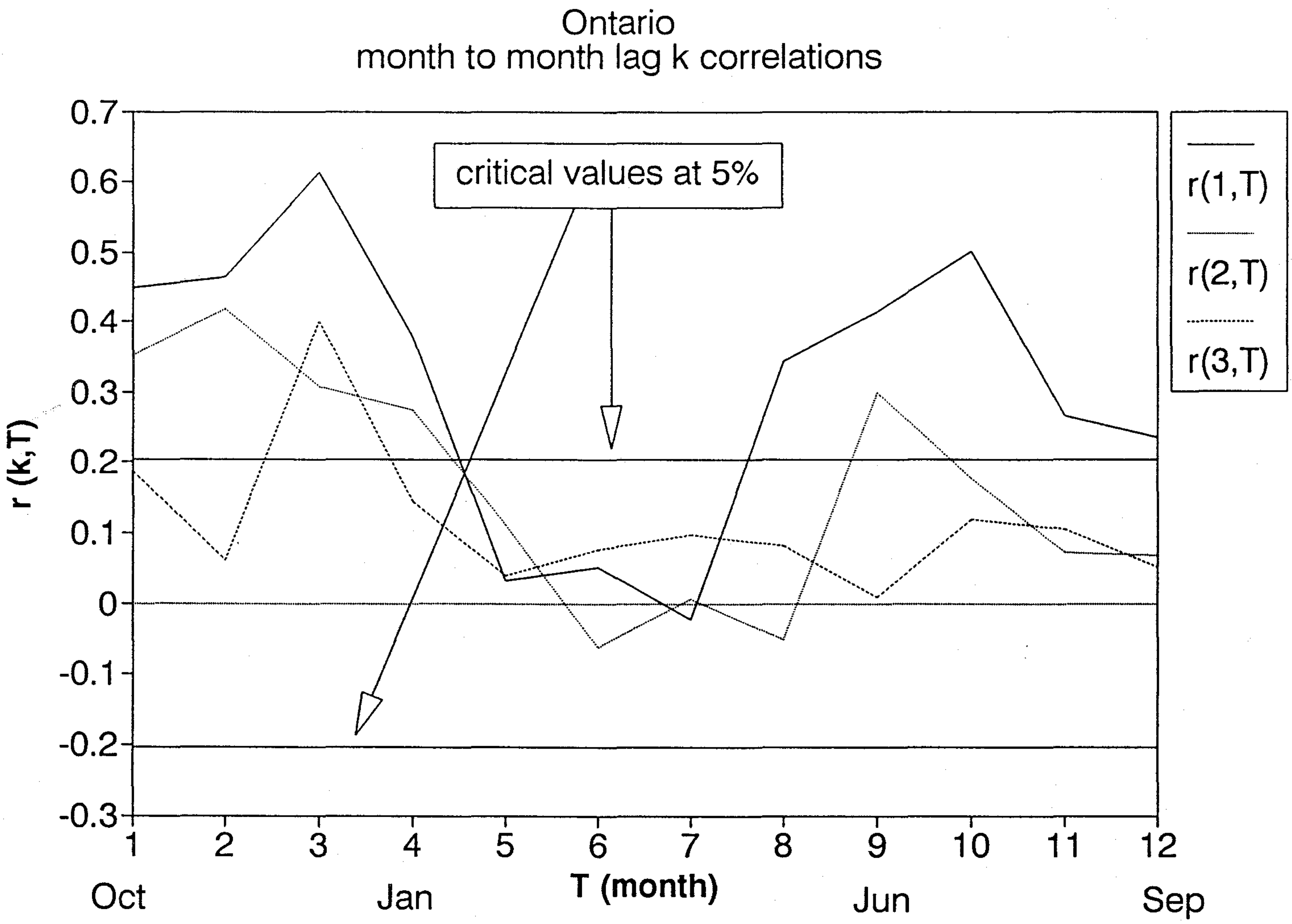


Figure 1.9 d: Lag-1 to lag-3 month-to-month correlation coefficients for Lake Ontario.

In general, the month-to-month coefficients of correlation are low. For lag-1 and 2, the highest correlations are between autumn months, followed by spring and summer months. The lowest correlations are between winter months. Almost none of the lag-3 correlations are significant.

1.2.2 Monthly cross-correlations

Table 1.7 gives the significant lag-0 cross-correlations at a 5% level of significance between monthly NBS for the four Lakes.

Table 1.7: Significant lag-0 cross-correlations between monthly NBS for the four Lakes.

	SU-MH	SU-ER	SU-ON	MH-ER	MH-ON	ER-ON
Jan	0,29	-	-	0,54	0,60	0,77
Feb	0,45	0,30	0,34	0,60	0,72	0,746
Mar	0,69	0,40	0,59	0,57	0,71	0,66
Apr	0,46	-	-	0,40	0,48	0,49
May	0,41	-	-	0,51	0,64	0,73
Jun	0,56	0,21	0,33	0,52	0,54	0,54
Jul	0,34	-	-	0,43	0,60	0,49
Aug	0,51	-	0,25	0,48	0,43	0,36
Sep	0,52	0,27	-	0,48	0,52	0,78
Oct	0,46	-	0,21	0,59	0,41	0,67
Nov	0,57	0,39	0,25	0,52	0,52	0,75
Dec	0,59	0,42	0,40	0,58	0,61	0,74

In general the correlation structure of monthly NBS shows a similar spatial pattern as the one observed for annual NBS (Table 1.2). As we go downstream from Lake Superior (Fig. 1.1) the correlation coefficients with other lakes decreases for each month. The highest correlations occure between Lakes Erie and Ontario. The lowest correlations between sites are in summer months.

1.3 Properties to be explicitly preserved by simulation

The examination of the statistical characteristics of historical annual and monthly NBS are used to select the properties to be explicitly preserved by the simulation.

In brief, it has been shown (section 1.1.7) that the annual series of Lakes Superior and Michigan-Huron are stationary. Likewise, annual NBS series for Lakes Erie and Ontario show a significant shift in 1970. Serial correlations are low but significant, showing a complex dependance structure for Lakes Erie and Ontario (section 1.1.1) that may be induced by the observed shift in the series. The cross-correlations, ranging from 0,27 to 0,66, indicate that the spatial structure is also important (section 1.1.2). The analysis of surplus run properties (section 1.1.4), Hurst coefficient (section 1.1.5) and spectral analysis (section 1.1.6) show some persistence in the series.

On the other hand, series of monthly NBS (section 1.2) show large skewness and important seasonal variations of the mean, standard deviation and coefficients of correlation, which are in general low but significant at a 5% level of significance.

Given these characteristics of NBS series and, given that this study relates to extreme events, more attention should be payed to reproduce properties at the annual time scale than at the monthly time scale. The analysis of the correlograms, run properties, Hurst coefficients and variance spectra clearly show the long-term persistence of the annual series. For Lakes Erie and Ontario, such complex dependence structure may be the result of the apparent shifts in the annual NBS series. Thus, emphasis will be placed on annual properties, although monthly properties will be considered as well.

The properties to be preserved or explicitly modeled are:

- cross-correlation of order zero (spatial relationship);
- serial correlation as shown in historical records (temporal relationship and persistence);
- upward shifts for Lakes Erie and Ontario;
- mean and standard deviation;
- basic properties of monthly series (monthly mean and standard deviation).

Some other implicit properties will be checked through validation criteria. Since previous studies on the Great Lakes Basin (see for example Loucks 1989) have suggested that certain models may not be able to reproduce observed historical Lake levels, even if the NBS are adequately generated, the analysis of the data would also includes properties of historical and simulated Lake levels provided by the GLERL (see section 3).

2. Review of stochastic models and model selection

The purpose of this section is to review and evaluate the models available in computer program form for stochastic generation of monthly NBS at several sites.

The literature on stochastic hydrology includes several univariate models. The autoregressive (AR) type of models are by far the most widely used in hydrology. Descriptions of univariate AR models applied in hydrology are given by Salas *et al.* (1980) and Fiering and Jackson (1971). Univariate models can be classified in three classes:

- Autoregressive model, AR(p), where p denotes the order of the AR term;
- Autoregressive Moving Average model, ARMA(p,q), where p and q denote the orders of the AR and MA terms;
- Autoregressive Integrated Moving Average model, ARIMA(p,d,q), where d denotes the differencing component of the model and p and q are as defined before.

These models in their univariate form do not take into account spatial correlations between sites. They are not directly useful for multivariate modelling although, in certain modelling strategies they can be quite useful.

2.1 Multivariate models

The principal aims of multivariate models is to take into account the cross-correlation between sites. Multivariate AR and ARMA models can be constructed by fitting univariate models to each of the stations under study independently and then modelling the spatial correlation through the residuals. Generally, only the lag-0 cross-correlation is considered by restricting the parameters matrix of the model to be diagonal (Salas *et al.* 1980). These models are referred to as contemporaneous such as the CARMA model. Review of multivariate models can be found in Salas *et al.* (1985), Stedinger *et al.* (1985a), Grygier and Stedinger (1990) and C.E.A. (1990). Some theoretical aspects of multivariate models are discussed by Fiering (1964), Matalas (1967) and Bernier (1971).

Multivariate models can be classified in direct modelling approach and indirect modelling approach (Grygier and Stedinger 1988, C.E.A. 1990). Direct modelling is used to build the model directly based on the monthly time series. In the indirect approach the annual series are first generated and then the annual values are disaggregated into monthly or smaller time units.

A number of data generation studies has been made related to streamflow simulation (Young and Pisano 1968, Srikanthan *et al.* 1983, 1984, Nathan *et al.* 1989, Salas and Abdelmohsen 1991). Studies performed by Megerian and Pentland (1968), IGLLB (1973), Yevjevich (1975) and Loucks (1989) include simulation of the Great Lakes NBS. The latter studies have been performed using the direct approach. Generally, monthly NBS statistical properties are well reproduced by the direct approach. Unfortunately, the direct approach usually fail to adequately reproduce statistical properties and persistence of annual NBS (Grygier and Stedinger 1988). To preserve the long-term persistence characteristics observed in the annual NBS, a special attention will be payed to the indirect modelling approach. However, the direct approach would also be considered as a back-up model.

A condensed disaggregation procedure proposed by Grygier and Stedinger (1988) has been choosen for the indirect approach. Multivariate step disaggregation (Santos and Salas 1991) is an alternative but the computer program was not available at the time of this study. Likewise the LAST model (Lane 1979, Lane and Fervert 1988) is an other alternative but the Grygier and Stedinger model is a more recent one allowing automatic selection of data transformation. Three reasons justify the choice of the Grigier and Stedinger (1988) model:

1) the complete procedure (including data transformation) is available in computer form in the SPIGOT Synthetic Streamflow Generation Software Package (Grygier and Stedinger 1990, 1991);

2) the Grygier and Stedinger (1990, 1991) condensed procedure reduces the number of parameters to be estimated (Grygier and Stedinger 1988) in comparison with other models like the Valencia-Schaake model (Valencia and Schaake 1973), LAST model (Lane 1979, Lane and Fervert 1988), and the Generalized SPC model (Stedinger *et al.* 1985b);

3) the procedure can explicitly reproduce the correlations between monthly and annual flows, the correlations between consecutive monthly flows, and the cross-correlations at different sites (Grygier and Stedinger 1990), although it cannot reproduce the correlation of the last season of the previous year with the first season of the current year.

The direct approach will be considered using a monthly-annual Singular Value Decomposition model (SVD) (Cavadias 1985), which is based on the principal component analysis (Fiering 1964, IGLLB 1973). The SVD model software is not available but it is relatively easy to code using available statistical packages. The SVD model has the main advantage of explicitly taking into account all the correlation structures of the data set in time and space.

2.2 Description of the pre-selected models

The model used to generate NBS should capture the important statistics of the record data. Since the multivariate AR(1) model available in SPIGOT may not be able to reproduce persistence, run characteristics and shifts in historical records, the use of alternative models are considered. Based on theoretical considerations, three models are selected for samples simulation. These models are described in the following sections. The choice of one model for the final simulation will be based on the results of an exploratory validation of generated samples (see section 3).

For the indirect approach, two different procedures are used to reproduce the shifts at the annual time scale. The first procedure is built to directly introduce the shifts in the generated sequences of annual NBS using a multivariate AR(1) model with shifting levels. This procedure is based on the idea of shifting level modelling of hydrologic series proposed by Boes and Salas (1978), Salas and Boes (1980) and Salas *et al.* (1981). The second procedure will try to fit the general shape of the observed correlograms (Fig. 1.2) using a multivariate ARMA(1,1) model. In both cases, the generated annual NBS will be disaggregated into monthly NBS.

For the direct approach, it will be assumed that the correlation structures of the monthly-annual data vectors, which are explicitly preserved using the SVD model, will be able to implicitly reproduce the shifts and other characteristics observed in the historical records.

2.2.1 Multivariate AR(1) model with shifting levels

In this model two sets of parameters of a multivariate AR(1) model are estimated from the two parts of the historical records (first 70 years, last 20 years). To generate the annual NBS, SPIGOT will use one model for a certain time, then switch to the other and so on. The time spent at either level is taken from a geometric distribution. Parameter of the geometric distribution is set by the user to give average length of time to stay at each level over the whole generated sequences in the same proportion as the one observed in the historical records (in our case 70/20). The geometric distribution is used here as a mixing process only, and no attempt is made to say that the duration of the observed shifts in annual NBS are geometrically-distributed. The second step in the data generation uses a multivariate annual to monthly disaggregation model to generate monthly NBS at the four sites.

A complete description of the stage disaggregation sheme used in this project is given in the SPIGOT Technical Description, Version 2.6 (Grygier and Stedinger 1990) and will not be repeted here. In the SPIGOT Technical Description report:

- the methodology of the Scheme III - Multivariate Annual Flow Generation and Disaggregation is presented in page 13;
- Section 3.6 (page 23) gives the general description of the Multivariate Annual-Monthly Model;
- Section 3.3 (page 19) gives the Multivariate Autoregressive Model used to generate normally-distributed annual flow vectors (equa. 2*);
- the disaggregation model is described in section 3.5, equations (22) and (23).

A procedure added to SPIGOT by Jan Grygier to generate random variates from the geometric distribution is described in the following section.

Generating random variates from the geometric distribution

We want to generate random variate N_i representing the length of a series at one level in the shifting levels model, with $\mu(N) = 1/p = 20$ or 70 years as an example.

The geometric distribution has:

$$P[N > n] = (1-p)^n \quad (2)$$

$$P[N = n] = p(1-p)^{n-1} \quad (3)$$

We will generate variates from the uniform distribution U(0,1) and then transform them.

The uniform distribution has:

$$P[X > x] = 1 - x \quad (4)$$

We want a transformation N = g(x) so that:

$$P[N > n] = P[X > x] \quad (5)$$

where n = g(x). Thus,

$$(1-p)^n = 1 - x$$

$$n \log(1-p) = \log(1-x) \quad (6)$$

$$n = \log(1-x) / \log(1-p) = g(x)$$

To get geometrically-distributed variates, we take samples from the U(0,1) distribution and apply the transformation g to them.

Because (1-x) is distributed exactly the same as x when x comes from U(0,1) the SPIGOT program just uses

$$g(x) = (\log x) / \log (1 - p) \tag{7}$$

In the current implementation SPIGOT asks the user to input the average length of time to stay at each level for the generation. In output, SPIGOT gives the time duration of each generated sequence and the average time spent at each level for the whole generation.

2.2.2 Multivariate ARMA(1,1) model

Autoregressive Moving Average time series models have proven to be a flexible tool for use in water resources planning. The ARMA(1,1) model, in particular, has a physically reasonable correlation structure which can reflect the long-term persistence observed in some geophysical time series. It has been showed that persistence may result either from long memory in hydraulic processes (Mandelbrot and Wallis 1968), or from shifts in the mean of these processes related to climatic changes (Boes and Salas 1978). The ARMA(1,1) structure may be considered compatible with either explanations that have been advanced.

The general multivariate ARMA(1,1) model may be written

$$X_t = \phi\, X_{t-1} + V_t - \theta\, V_{t-1} \tag{8}$$

where X_t is an m x 1 vector of normally-distributed flow residuals (zero mean) in period t with covariance matrix S_0,

V_t is an m x 1 vector of time-independent normally-distributed random fluctuations with covariance matrix G,

and ϕ and θ are m x m coefficient matrices.

Salas *et al.* (1980) and Loucks *et al.* (1981) have suggested that ϕ and θ be taken as diagonal matrices. Then the elements of each matrix are essentially the parameters of univariate ARMA(1,1) models fitted to the flows at each site (contemporaneous ARMA or CARMA model). This method has the advantage that each site is described with well known properties and that the multisite model is hydrologically reasonable. With the assumption of diagonality for ϕ and θ, the model fitting process is performed in two independent steps:

1) estimation of ϕ and θ;
2) estimation of the G matrix.

The MATLAB, IDENT procedure (MATLAB 1990) was used to estimate the parameters ϕ and θ for Lakes Superior, Michigan-Huron and Ontario. For Lakes Superior and Michigan-Huron, the parameter θ of the MA process is not significant, which simply gives an AR(1) model with parameter ϕ.

For Lake Erie, due to the shape of the autocorrelation function (Fig. 1.2) we could not obtain good estimates of the parameters ϕ and θ. To avoid this problem we simulate several autocorrelation functions for different values of ϕ and θ (Salas *et al.* 1980). We choose the set of parameters which gives the best fit of the observed correlogram as a whole. The validity of this procedure will be checked a posteriori through the validation criteria.

When the parameters ϕ and θ are estimated, V_t in Equation (8) can be algebraically derived. To arrive at white noise residuals, a model of the form

$$V_t = B\,\varepsilon_t \tag{9}$$

may be used, where B is an m x m matrix of coefficients and ε_t is an m x 1 vector of residuals with mean zero and variance one, which are uncorrelated in time and space.

In regard to the estimation of the covariance of the residuals (matrix G) for the CARMA(1,1) model, Stedinger *et al.* (1985a) have recommended the following estimator

$$\hat{g}_{ij} = \frac{\left(\hat{S}_0\right)_{ij}\left(1-\hat{\phi}_{ii}\,\hat{\phi}_{jj}\right)}{1-\hat{\phi}_{ii}\,\hat{\theta}_{jj}-\hat{\theta}_{ii}\,\hat{\phi}_{jj}+\hat{\theta}_{ii}\,\hat{\theta}_{jj}} \tag{10}$$

where $\left(\hat{S}_0\right)_{ij}$ is the ij th element of $\hat{S}_0$ the lag zero cross-covariance of X_t. Then, the matrix B of equation (9) is given by solving

$$BB' = \hat{G} \tag{11}$$

In the current implementation SPIGOT cannot estimate the parameters for a general ARMA model, but it can generate ARMA(1,1) annual flows if the user puts the right parameters in the parameters file using estimated ϕ and θ and equations (9), (10) and (11). For this model we are using the same SPIGOT disaggregation procedure as the one described in section 2.2.1.

2.2.3 SVD model

The multivariate simulation method proposed in the present study is based on the singular value decomposition (SVD) theorem (Cavadias 1985). Consider the N x p data matrix X, and let r be the rank of X. The singular decomposition of a matrix X can be written as follows:

$$X = VD^{1/2}U' \tag{12}$$

in which V is an N x r matrix of eigenvectors of XX'; D is a r x r diagonal matrix of the positive eigenvalues of XX'; and U is a p x r matrix of the eigenvectors of X' X.

Let X_S be the standardized matrix of X, that is, each column of X_S has zero mean and unit variance. On the basis of equation (12), the matrix V_S of the standardized principal components of X_S is given by

$$V_s = X_s U D^{-1/2} \tag{13}$$

Note that the colums of V_S are orthonormal, that is $V_s' V_s = I$, and hence they can be generated independently following the distribution function of each column of V_S.

Equations (12) and (13) provide the basis for the generation of monthly and annual NBS for the four Lakes considered. More specifically, the steps in the calculation of the simulated data matrix $\hat{X}$ are as follows:

1. Transform the observed data matrix X, where N represents the number of years of observation; and $p = 52$ indicates that for each of the four sites there are 13 data columns, into a normal data matrix X* using a three-parameter lognormal transformation.

2. Apply the first-order autoregressive models to the data matrix X*, and let E be the matrix of the residuals.

3. Compute the standardized matrix E_S.

4. Compute the correlation matrix R_{ES} from E_S, and its associated eigenvalue and eigenvector matrices D and U respectively.

5. Compute the principal component matrix V_S using equation (13).

6. Generate the matrix $\hat{V}_s$ of random numbers following the probability distributions of the columns of V_S.

7. Compute the standardized matrix V_{S1} from V_S, its corresponding correlation matrix R_{VS1} and its associated eigenvalue and eigenvector matrices, U_{VS1} and D_{VS1}, respectively.

8. Compute the matrix $\hat{V}_{s2}$ using equation (13). This step is necessary to insure the orthonormality of the generated random matrix V_{S2} .

9. Compute the matrix $\hat{E}_s$ of standardized synthetic residuals using equation (12).

10. Compute the matrix of non-standardized residuals $\hat{E}$ using the means and standard deviations of the residual matrix E.

11. Compute the matrix $\hat{X}*$ of normalized synthetic flows using the fitted autoregressive models and $\hat{E}$.

12. Compute the matrix $\hat{X}$ of synthetic flows from $\hat{X}*$ using the inversed log-transformation.

2.2.4 Data transformation

The models used in SPIGOT assume that annual and monthly NBS can be transformed into normally-distributed variables. If necessary, the transform variables are obtained from the real NBS by taking one of the three transformations given in SPIGOT (Grygier and Stedinger 1990):

- 2-parameter lognormal distribution;
- 3-parameter lognormal distribution (using a quantile lower bound estimator);
- 3-parameter gamma distribution (Pearson Type III distribution).

In the SPIGOT parameter estimation module, univariate probability models are developed for the data at all sites in each month and for annual series. The results of a Filliben's correlation test are used to determine the best transformation (Grygier and Stedinger 1991). Table 2.1 gives the transformation applied to the NBS data.

For the SVD model, the 3-parameter lognormal transformation has been applied to all data vectors.

Table 2.1: Data transformation based on the highest Philliben correlation for annual and monthly NBS at all sites, 1 = normal, 2 = 2-parameter lognormal distribution, 3 = 3-parameter lognormal distribution, 4 = 3-parameter gamma distribution.

	SU	MH	ER	ON
Annual	1	1	1	4
Jan	1	3	3	3
Feb	3	3	3	3
Mar	4	3	3	3
Apr	3	3	1	1
May	3	3	3	2
Jun	3	3	3	2
Jul	3	3	3	3
Aug	3	4	3	3
Sep	3	3	3	3
Oct	1	3	3	3
Nov	4	4	3	3
Dec	1	3	3	4

3. Samples simulation, validation procedures and final model selection

For the simulation of the Great Lakes monthly NBS, the three multisite stochastic models have been used:

- Multivariate AR(1) model with shifting levels using SPIGOT for disaggregation, which will be referred to as SL in the following sections;

- Multivariate contemporaneous ARMA(1,1) or CARMA model using SPIGOT for disaggregation, which will be referred to as ARMA in the following sections;

- Annual-Monthly Singular Value Decomposition model, which will be referred to as SVD in the following sections.

3.1 Samples simulation

The three models were used to generate samples of annual NBS data (100 series x 90 years) and monthly NBS data (11 series x 90 years x 12 months) at four sites: Lake Superior, Lake Michigan-Huron, Lake Erie, and Lake Ontario.

For each model, the monthly NBS was used by the GLERL to simulate quarter monthly water level data (11 series x 90 years x 48 quarter months) at Lake Ontario.

The three generated data sets (annual NBS, monthly NBS, and quarter monthly levels), were used to validate the NBS and level data against historical record for the NBS and levels of the Great Lakes resulting from the application of modified regulation plans 1977-A for Lake Superior and 1958-D for Lake Ontario (Rassam *et al.* 1992, Section 5). This so-called historical record is used as a basis of comparison (BOC) in the validation process. The first comparison of historical and generated data was based on several characteristics like the basic statistics, correlation coefficients, frequency plots and runs statistics. A total of 181 figures and tables were obtained and compared for the validation of the NBS (annual and monthly) and level data.

A selection of the most significant results of the annual validation at Lake Erie (most difficult site to simulate) and of the monthly NBS and quarter month level at Lake Ontario (outflow of the Great Lakes system) were the basis for the choice of the model for the final simulation.

During a meeting held in Hydro-Québec, a board of specialists reviewed the results of the validation for 16 prespecified comparison characteristics, which will be described in the following sections. The participants were asked to assign a unanimous appreciation code to each model:

- -1: bad performance,
- 0: good performance,
- 1: excellent performance.

The tabulated results of the appreciation code will be used in order to select the model having the overall best performance (see Section 3.3 and Table 3.4).

3.2 Validation procedures

The annual NBS, monthly NBS and quarter monthly level validation characteristics used for the selection of the model are described in the following sections. The validation characteristics are also listed in Table 3.4 along with the appreciation code assigned by the participants to each model. This process, labelled exploratory validation aims at identifying the best among the three generation models, according to the following characteristics, each of which being used as a basis for a selection criterion. The validation procedures are divided into three (3) categories, namely: validation of annual series of NBS, validation of monthly series of NBS and validation of quarter monthly levels of Lake Ontario.

3.2.1 Validation of annual series: comparison of historical and generated annual NBS

The generated annual NBS data were arranged in 100 series of 90 years from which generated annual validation characteristics were computed. The mean, standard deviation, maximum and minimum of each validation characteristic were obtained and compared with those obtained from the historical record. Most of this type of validation on annual series was carried on Lake Erie data because it was felt that they were the most critical to simulate.

Characteristic #1: Basic annual statistics

The first three rows of Table 3.1 show a comparison of the historical and generated mean, standard deviation and skewness. The mean, standard deviation, maximum and minimum values for each statistic are indicated. It could be noticed that the SL and ARMA models generally perform well whereas the SVD model fails to reproduce the variance of the basic annual statistics.

Characteristic #2: Hurst coefficient

The last row of Table 3.1 shows a comparison of the historical and generated Hurst coefficients. The mean, standard deviation, maximum and minimum values for this coefficient are indicated. It could be said that all three models reproduce fairly well the Hurst coefficient.

Characteristic #3: Annual cross-correlation

Table 3.2 shows a comparison of the historical and generated annual lag-zero cross-correlations among the sites. The mean, standard deviation, maximum and minimum values for the coefficients are indicated. The spatial properties of annual NBS are reproduced equally well by the three models although the SVD fails again to indicate any variance of the simulated cross-correlation coefficients.

Table 3.1: Basic annual statistics and Hurst coefficient for Lake Erie.

		Generated Annual Statistics			
	Historical		SL	ARMA	SVD
Mean (tcfs)	236,36	mean	237,62	234,68	241,78
		s.d.	29,40	21,37	2,02
		max	324,38	284,41	246,64
		min	187,72	186,42	237,29
Standard Deviation (tcfs)	110,26	mean	106,06	109,24	103,20
		s.d.	9,70	10,02	5,07
		max	138,10	144,07	120,54
		min	83,97	89,91	94,18
Skewness	0,093	mean	0,067	0,010	0,038
		s.d.	0,252	0,248	0,268
		max	0,616	0,546	1,018
		min	-0,594	-0,598	-0,582
Hurst	0,757	mean	0,713	0,723	0,723
		s.d.	0,085	0,066	0,063
		max	0,899	0,862	0,873
		min	0,508	0,537	0,546

Table 3.2: Annual cross-correlations.

		Generated			
	Historical		SL	ARMA	SVD
SU-MH	0,54	mean	0,49	0,53	0,51
		s.d.	0,12	0,08	0,03
		max	0,77	0,71	0,59
		min	0,14	0,31	0,43
SU-ER	0,30	mean	0,22	0,24	0,21
		s.d.	0,11	0,10	0,05
		max	0,55	0,50	0,34
		min	-0,07	-0,03	0,11
SU-ON	0,27	mean	0,20	0,21	0,23
		s.d.	0,10	0,10	0,05
		max	0,43	0,45	0,37
		min	-0,11	-0,06	0,12
MH-ER	0,50	mean	0,54	0,53	0,51
		s.d.	0,10	0,06	0,04
		max	0,83	0,67	0,60
		min	0,30	0,37	0,40
MH-ON	0,62	mean	0,60	0,61	0,59
		s.d.	0,07	0,06	0,04
		max	0,74	0,77	0,67
		min	0,44	0,44	0,50
ER-ON	0,66	mean	0,65	0,67	0,61
		s.d.	0,09	0,07	0,05
		max	0,83	0,80	0,73
		min	0,41	0,51	0,51

Characteristic #4: Autocorrelation

Figures 3.1a,b,c respectively show for the three models (SL, ARMA, SVD) a graphical comparison of the historical and generated (average of the 100 series) lag 1 to lag 14 autocorrelation coefficients. Although the SL and ARMA models provide the same general trend of the historical autocorrelogram, the SVD out performs them by following more closely the variations of the correlations coefficients for each time step lag.

Characteristic #5: NBS frequency curve

The frequency plots of the sorted historical values versus the sorted median and sorted 25% and 75% quantiles of the generated ones are shown in Figures 3.2a,b,c. It would be argued that the median of the values generated by the SL and ARMA models follow closely the historical frequency plot, while they depart from it for the SVD model. It would be seen also that the SVD model tends to generate annual NBS of which the upper tail of the distribution is lower than historical values and the lower tail is higher than historical values.

Characteristic #6: NBS return period

The frequency plots with a return period scale using the Weibull formula of the historical and generated annual NBS for each model are shown in Figures 3.3a,b,c. Although this characteristic is the same as the previous one, nevertheless it highlights a slight superiority of the SL over the ARMA model and a further superiority over the SVD model. From this type of plot, it is clear that the SVD fails to reproduce the distribution properties of the annual NBS.

a)

b)

c)

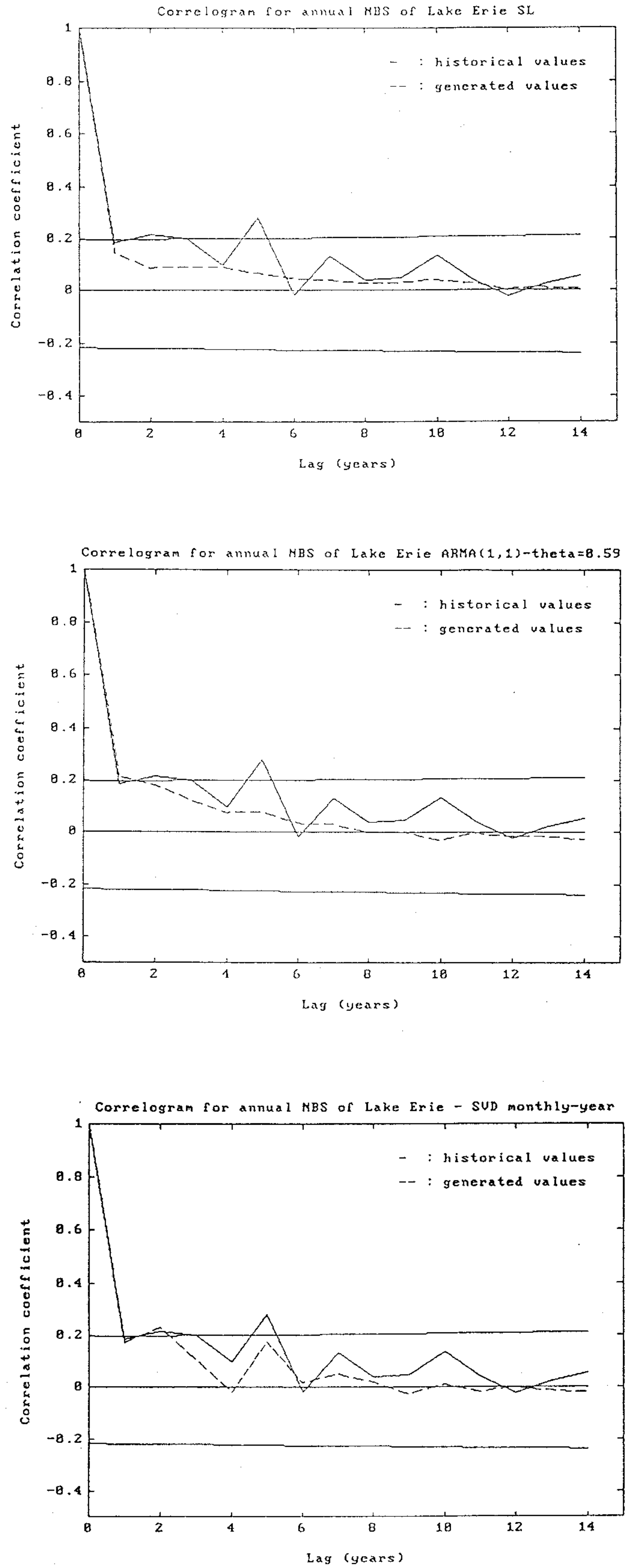


Figure 3.1: Correlogram of annual NBS (Lake Erie) for a) SL, b) ARMA, and c) SVD.

a)

b)

c)

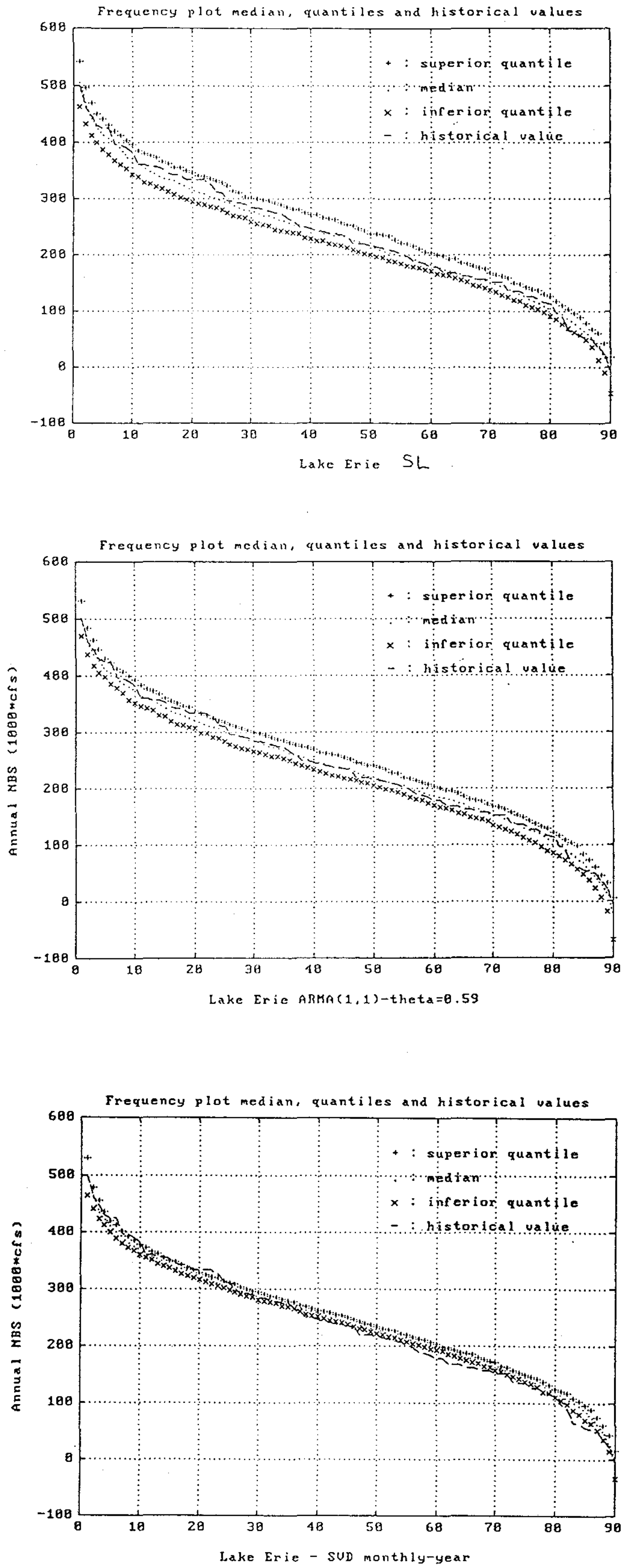
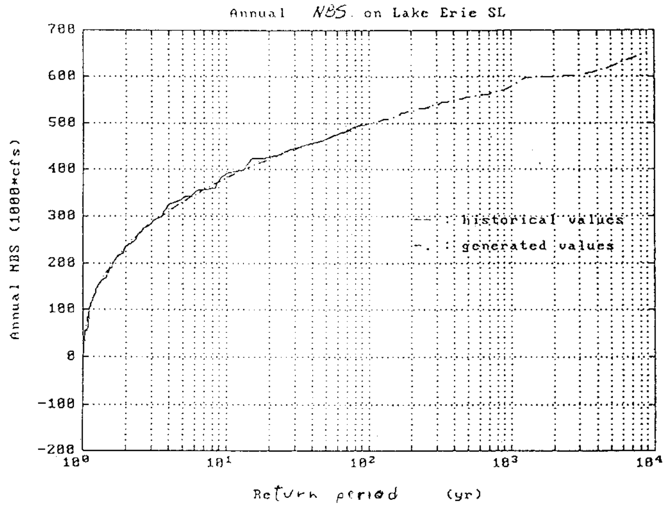
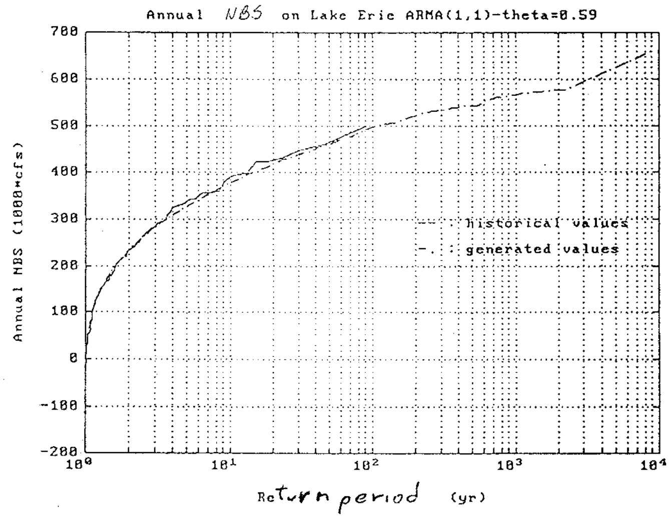


Figure 3.2: Frequency plot of annual NBS (Lake Erie) for a) SL, b) ARMA, and c) SVD.

a)



b)



c)

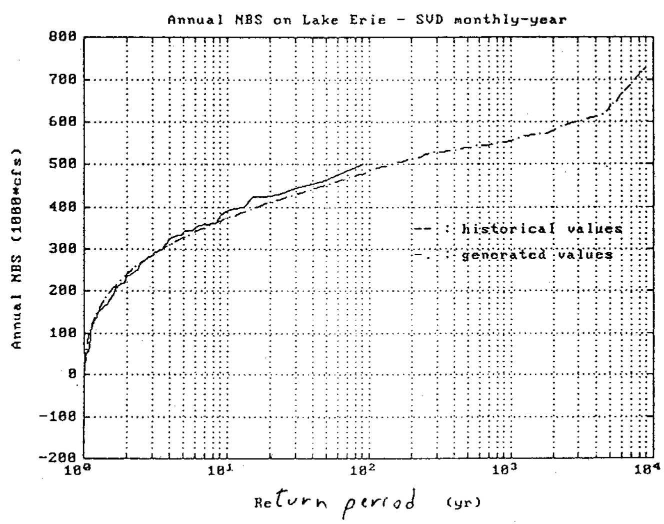


Figure 3.3: Return period (Weibull) of annual NBS (Lake Erie) for a) SL, b) ARMA, and c) SVD.

Characteristic #7: Surplus run length (RL)

The 100 series of 90 years of generated data were used to determine the highest run length (RL) for a truncation level equal to the historical mean. The mean, standard deviation, maximum and minimum values of RL for the generated data using the three models are shown with the corresponding maximum historical values in Table 3.3. Both the traditional definition and the 4-year running average run criteria were used.

Using the 4-year running average as a criterion, it could be seen that the historical maximum RL falls within the mean + standard deviation of the generated RL for all three models. It could be therefore stated that they all perform equally well with a little edge given to the SL model.

Characteristic #8: Surplus run volume (RS)

Same as #7 for the highest surplus run volume (RS) (Table 3.3). The same conclusion could be generally drawn for the RS values as for the RL values, although it should be noticed, this time, that the SVD model fail to reproduce a run volume whose mean + standard deviation is at least equal to the historical maximum on both 1-year and 4-year running average cirteria.

Characteristic #9: Partial record runs

The 9 000 years of generated data were also used to determine the partial record (25, 50 and 75- year) historical and generated largest surplus (RL and RS). As an example, for the 25-year historical surplus, the process consisted of subdividing the 90 years of records in partial record of 25 years each using a running length of 25 year records that generated 65 samples. The 65 values of each statistics (RL and RS) were averaged. For the 25-year generated surplus, the 9000 years are divided in 360 traces of 25 years each. The mean, standard deviation, maximum and minimum values of RL and RS for the generated data using the three models were computed with the partial record historical ones. The results of the partial record runs are not shown here because they were similar to the run analysis (Table 3.3). Here again, it could be stated that all three-models perform equally well.

Table 3.3: Surplus run length (RL) and run volume (RS) (Lake Erie).

	Historical maximum		Generated							
			SL			ARMA		SVD		
	1-year	4-year		1-year	4-year	1-year	4-year	1-year	4-year
RL (year)	15	22	mean	9,34	20,78	8,36	17,18	8,39	17,30
			s.d.	5,59	16,29	3,56	6,94	2,69	6,26
			max	25	70	19	39	17	39
			min	3	4	2	6	4	8
RS (Hm^3) x 10^6	1,66	1,72	mean	0,95	1,59	0,85	1,16	0,80	1,18
			s.d.	0,63	1,41	0,44	0,57	0,34	0,47
			max	2,80	5,80	2,18	2,71	2,47	2,94
			min	0,20	0,19	0,26	0,26	0,38	0,46

3.2.2 Validation of monthly NBS series: comparison of historical and generated monthly NBS of Lake Ontario

The generated monthly NBS was arranged into 11 series of 90 years by 12 months from which generated monthly validation characteristics were computed. The mean, standard deviation, maximum and minimum of each validation characteristic were obtained and compared with similar ones obtained from the monthly historical record.

Characteristic #10: Basic monthly statistics

Figures 3.4a,b,c and 3.5a,b,c show a comparison of the monthly historical and generated NBS mean and standard deviation for the three models. The mean, standard deviation, maximum and minimum values of the mean and the standard deviation for the 11 generated series are indicated.

It could be said that the three models performs equally well for the reproduction of the NBS monthly means and that the SVD performs somewhat less satisfactory than the other models for the reproduction of the NBS monthly standard deviation.

Characteristic #11: Monthly cross-correlation

Figures 3.6a,b,c respectively show for the three models a comparison of the historical and generated monthly lag zero cross-correlations (Site-Site correlation) between Lake Erie and Lake Ontario. The mean, standard deviation, maximum and minimum values of the correlation coefficients for the 11 generated series are indicated. Since all the correlations among the sites show a similar pattern, the correlations between remaining sites are not presented here.

The SL and ARMA models perform fairly well in reproducing the monthly lag-zero cross-correlations between Lakes Erie and Ontario, while the SVD model fails to reproduce them satisfactory. In effect, in the latter model the maximum lag-zero cross-correlations for 8 consecutive months starting from January of generated data, are below the historical values.

Characteristic #12: Month to Month correlation

Figures 3.7a,b,c respectively show for the three models a graphical comparison of the historical and generated month to month lag one correlations. The mean, standard deviation, maximum and minimum values of the coefficients of correlation for the 11 generated series are indicated.

Here again, it could be stated that the first two models namely SL and ARMA perform satisfactory while the SVD model performs somewhat less satisfactory.

a)

b)

c)

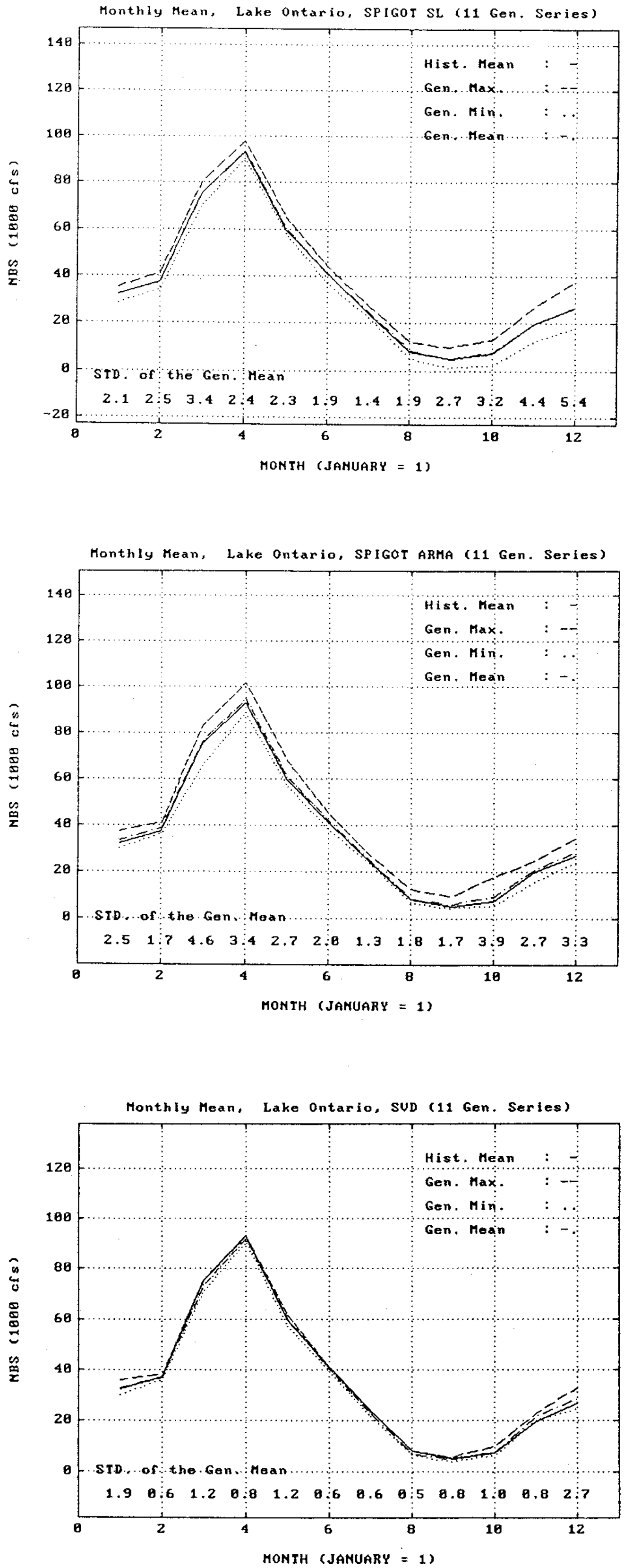


Figure 3.4: Mean of monthly NBS (Lake Ontario) for a) SL, b) ARMA, and c) SVD.

a)

b)

c)

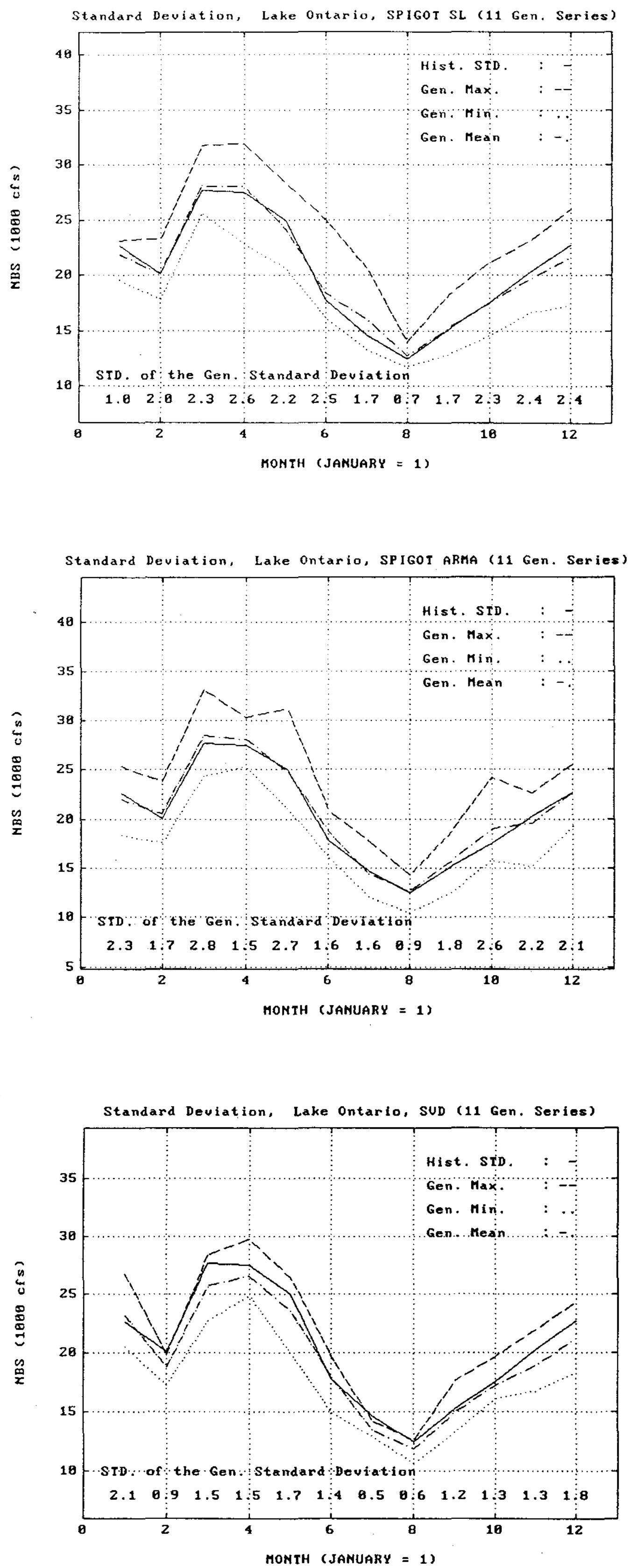


Figure 3.5: Standard deviation of monthly NBS (Lake Ontario) for a) SL, b) ARMA, and c) SVD.

a)

b)

c)

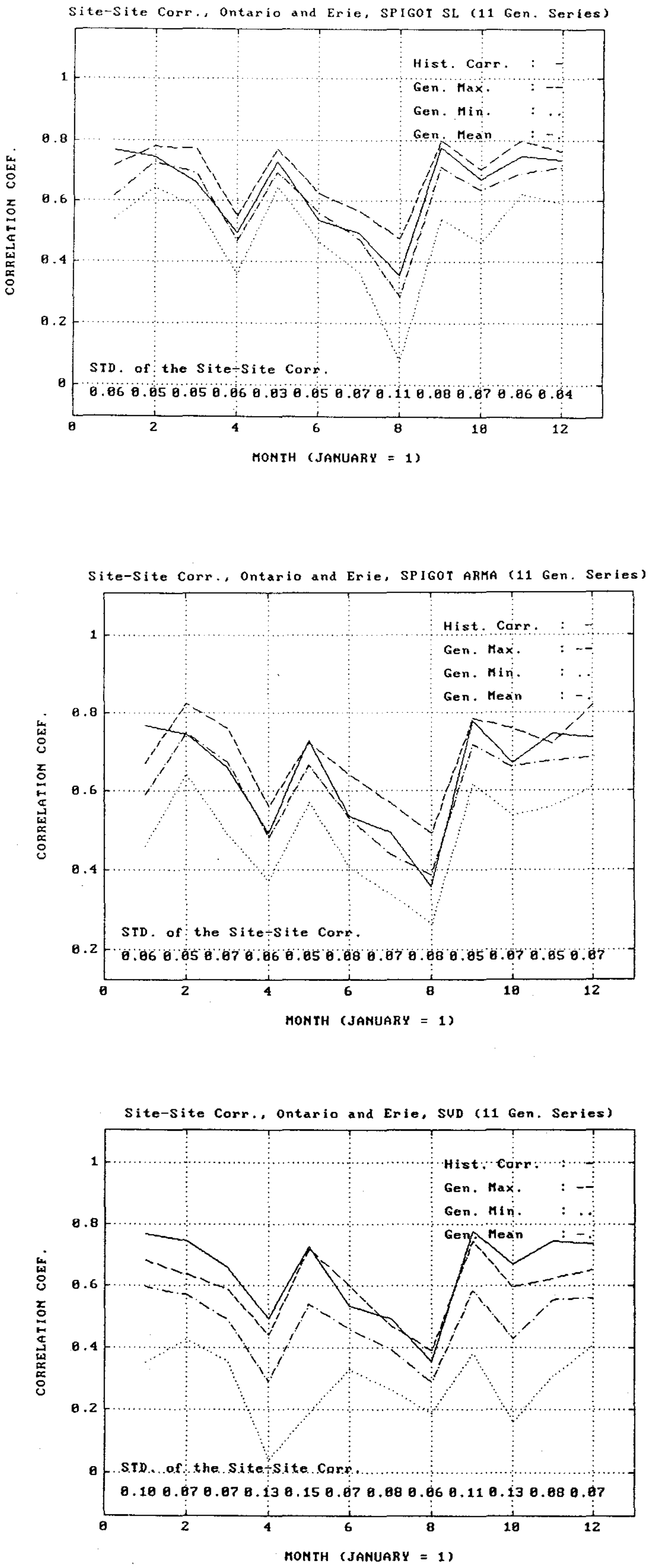
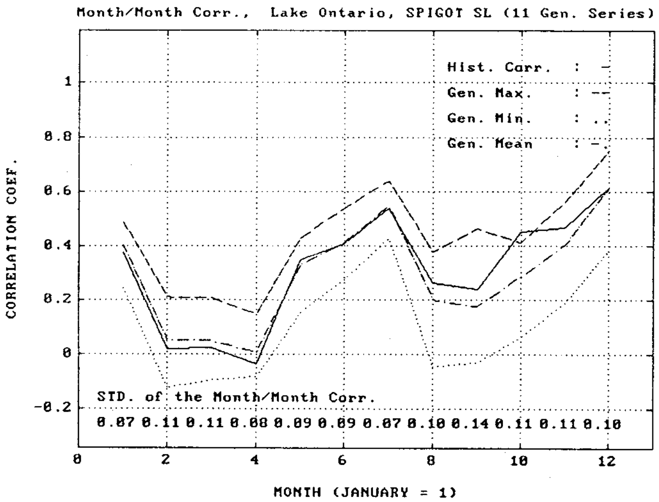
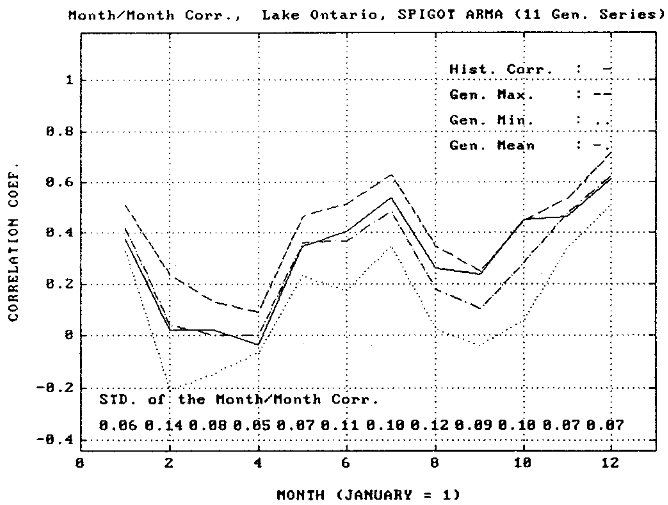


Figure 3.6: Monthly NBS lag zero cross-correlations between Lake Erie and Lake Ontario for a) SL, b) ARMA, and c) SVD.

a)



b)



c)

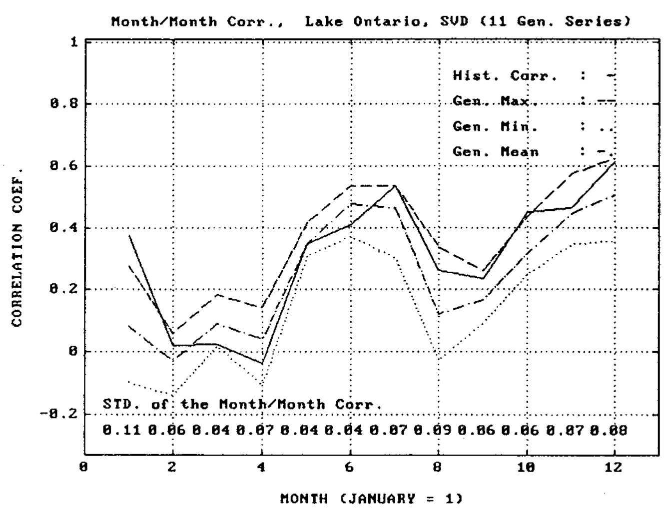


Figure 3.7: Month to month correlations (Lake Ontario) for a) SL, b) ARMA, and c) SVD.

3.2.3 Validation of quarter monthly levels: comparison of historical and simulated level of Lake Ontario.

Water level data have been simulated at the GLERL using monthly historical and generated NBS. A modification of the present regulation plan (1958-D) has been used for the simulation. As a consequence of this modification, the historical level data used in this study are different from the usual basis of comparison level data for Lakes Ontario and Erie that are shown in studies done by the International Joint Commission. The simulated quarter monthly level data were arranged in 11 series of 90 years by 48 quarter months, from which simulated level validation characteristics were computed. The mean, maximum and minimum of each validation characteristic were obtained and compared with similar ones obtained from the historical level record.

Characteristic #13 a: Basic level statistics, mean

Figure 3.8a,b,c shows a comparison of the quarter monthly historical and simulated mean levels for the three models. The mean, maximum and minimum values of the mean for the 11 simulated series are indicated.

For some reason, the three models fail to adequately reproduce the mean of the monthly means. Nevertheless, the historical mean falls within the generated minimum and maximum quater monthly values for the SL model, while it falls outside this marge for the first and last quarter months for the ARMA model, and while it falls completely below this marge for all quarter months for the SVD model.

Characteristic #13 b: Basic level statistics, maximum

Figure 3.9a,b,c shows a comparison of he quarter monthly historical and simulated maximum levels for the three models. The mean, maximum and minimum values of the maximum for the 11 simulated series are indicated.

The same remark as the one presented for the mean (Characteristic #13 a) could be done for the maximum of the quarter monthly levels.

a)

b)

c)

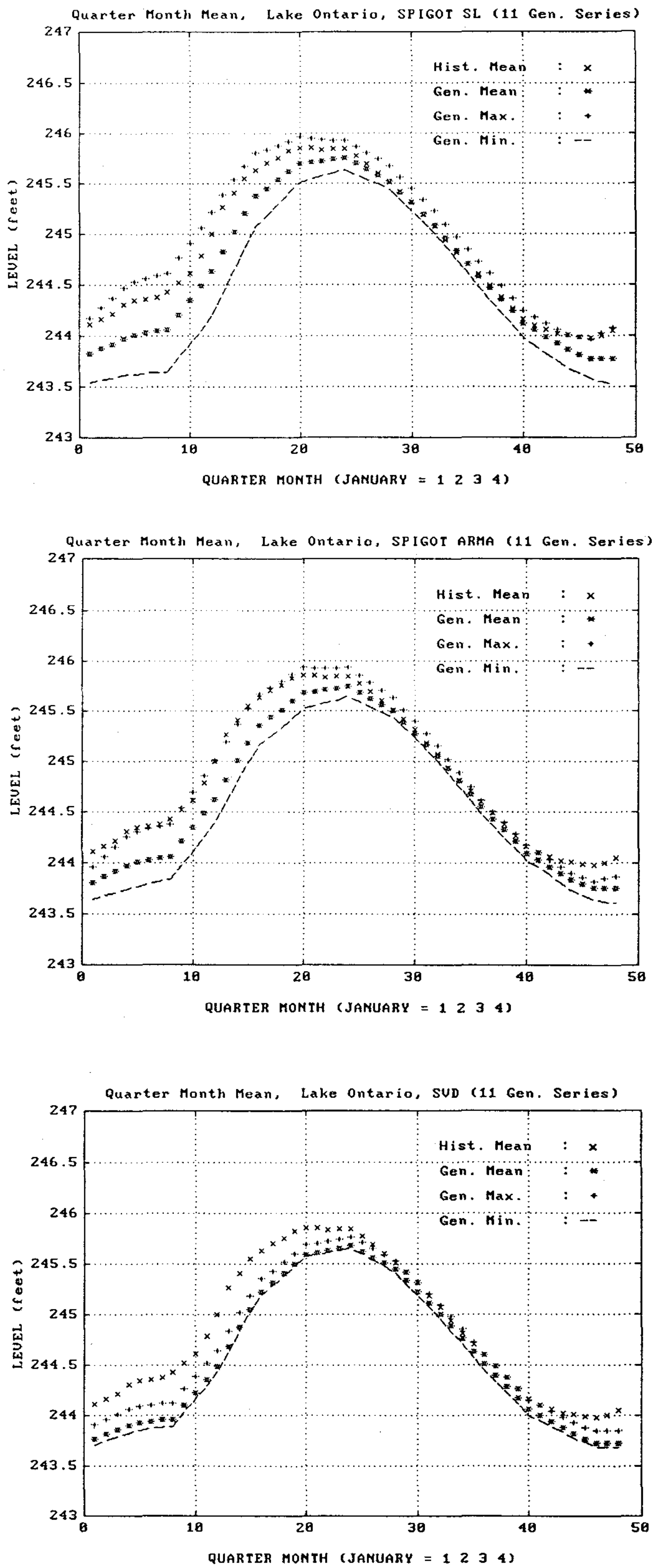
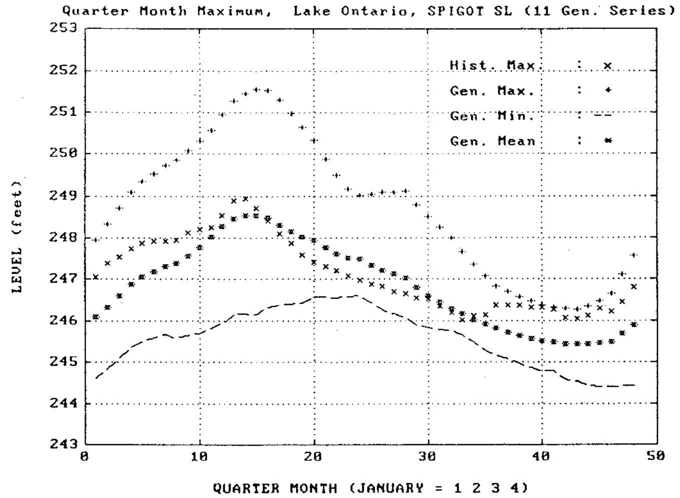
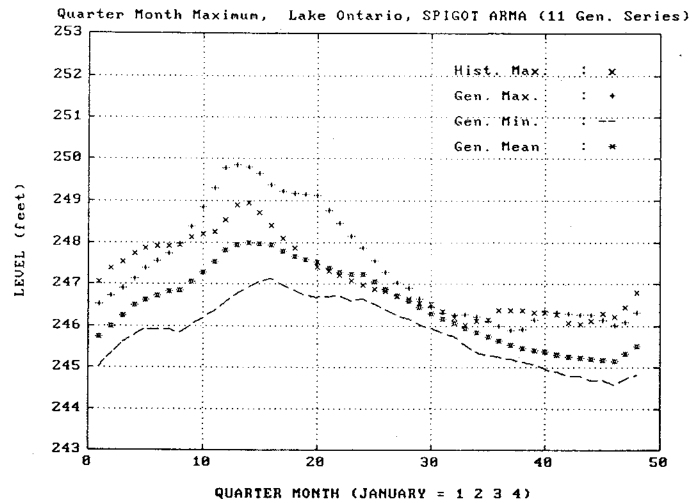


Figure 3.8: Mean of quarter monthly levels (Lake Ontario) for a) SL, b) ARMA, and c) SVD.

a)



b)



c)

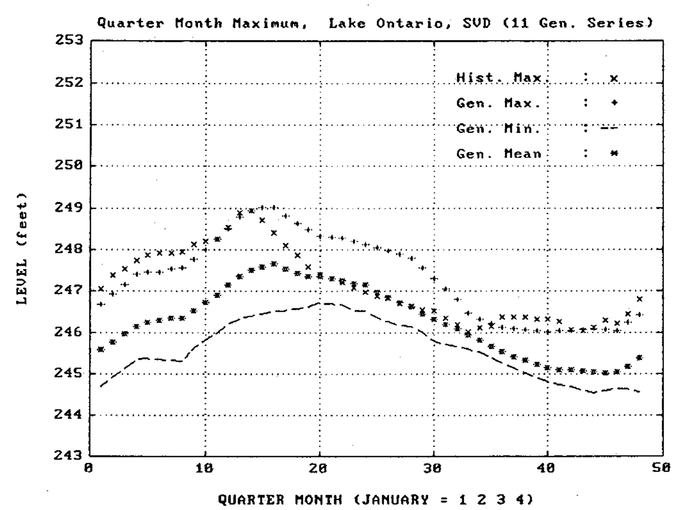


Figure 3.9: Maximum of quarter monthly levels (Lake Ontario) for a) SL, b) ARMA, and c) SVD.

Characteristic #14: Maximum level frequency curve

The frequency plot of the sorted quarter monthly maximum annual historical level versus the sorted maximum and minimum of the simulated ones are shown for the three models in Figures 3.10a,b,c.

While the quarter monthly maximum annual historical level frequency plot falls within the range of the maximum maximum and minimum maximum of the simulated values for the SL and ARMA models, it could be observed that the SVD fails to generate extreme maxima (i.e. in the upper tail) that are larger than historical ones.

Characteristic #15: Global maximum level frequency curve

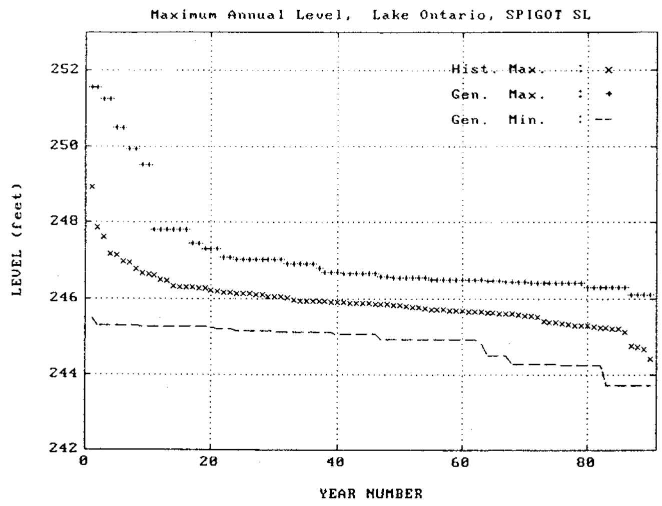
The global frequency plot of the sorted maximum quarter monthly annual historical and simulated level for the three models are shown in Figures 3.11a,b,c.

It would be seen that, in this respect, the SL model outperforms the two others. While the SVD model generated only one (1) value of Lake Ontario maximum quarter monthly annual level that is larger than the BOC, the ARMA generates two (2) and the SL 13.

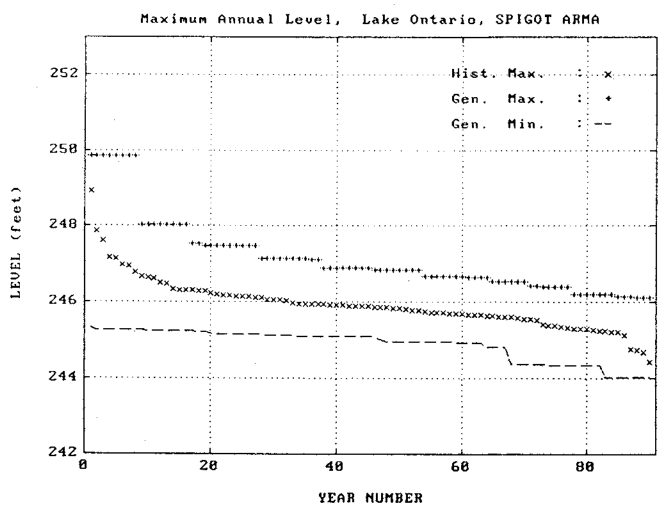
Characteristic #16: Global maximum level return period

The frequency plot with a return period scale using the Weibull formula of the quarter monthly maximum annual historical and simulated levels for each model are shown in Figures 3.12a,b,c. Although this characteristic is similar to the previous one, it nevertheless shows on a different scale the satisfactory performance of the SL model to simulate the maximum level distribution of Lake Ontario. On this basis, the SL model is the only acceptable one.

a)



b)



c)

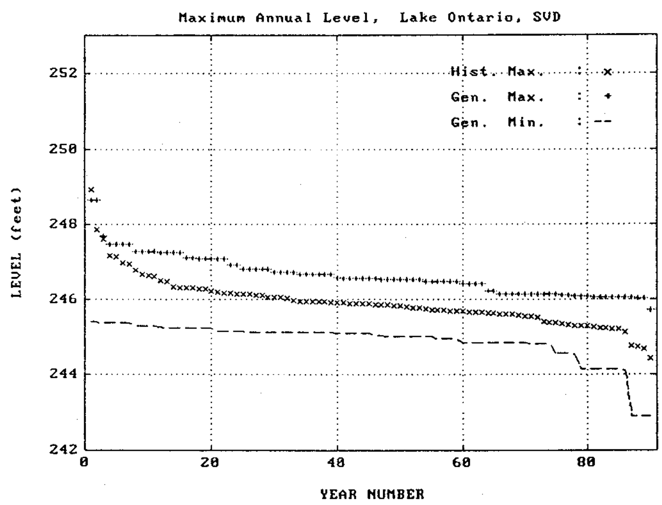
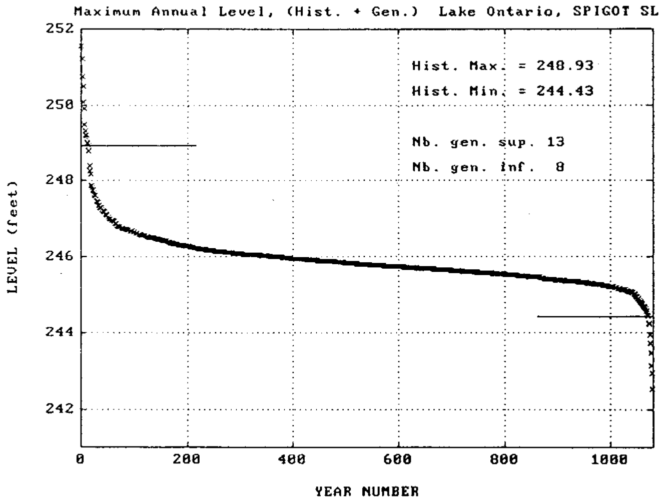
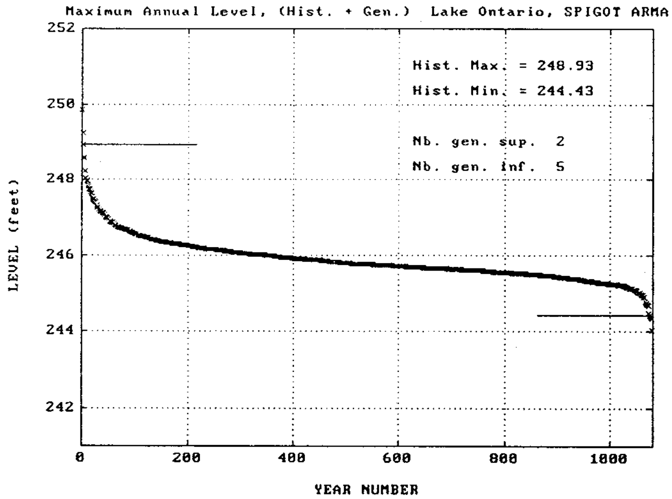


Figure 3.10: Frequency plot of the sorted quarter monthly maximun annual level and min. max. and max. max. of simulated values (Lake Ontario) for a) SL, b) ARMA, and c) SVD.

a)



b)



c)

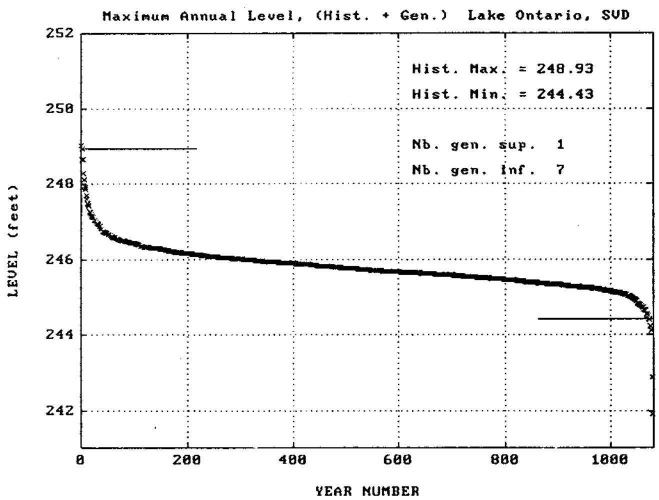


Figure 3.11: Global frequency plot of the sorted maximun quarter monthly annual level (Lake Ontario) for a) SL, b) ARMA, and c) SVD.

a)

b)

c)

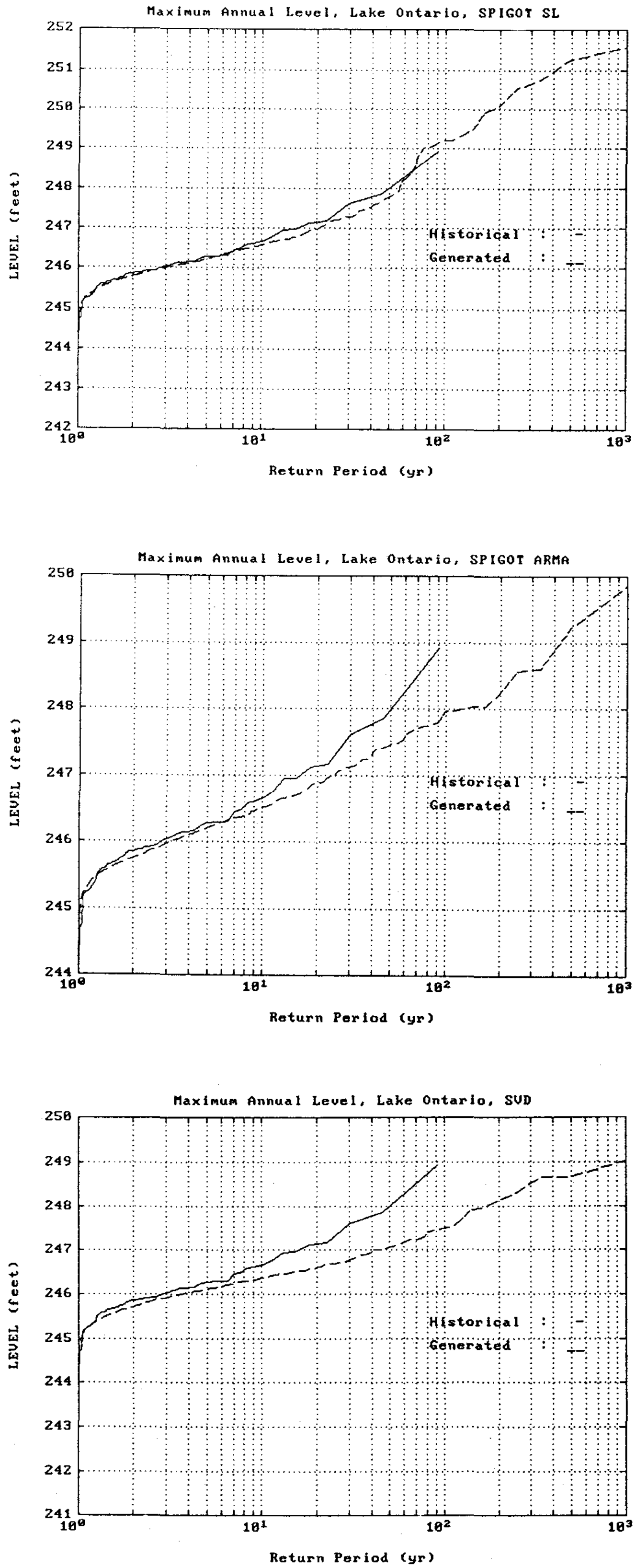


Figure 3.12: Frequency plot with a return period scale (Weibull) of the maximun quarter monthly annual level (Lake Ontario) for a) SL, b) ARMA, and c) SVD.

3.3 Final model selection

The global performance of the three models were scored according to 17 criteria, each correspondind to one of the above mentionned characteristics, given that characteristic 13 was further divided into 2 criteria: 13-a and 13-b. The scores were given according to the following 3-point scale:

- -1: bad performance,
- 0: good performance,
- 1: excellent performance.

The score was attributed subjectively by unanimous decision of various hydrologists from Hydro-Québec and INRS-Eau. Table 3.4 shows the unanimous appreciation code assigned to the three models by the participants. These appreciation codes are used to choose the model showing the overall best performance in regard to the different validation criteria.

In Table 3.4 the appreciation code of charateristics #1, 3, 9, 10 and 11 is attributed globally for all the statistics considered.

For the annual validation, best results have been achieved using the SL model, followed by the ARMA model. The SVD model does not adequately reproduce the annual basic statistics (Table 3.1, Fig. 3.2c). Figure 3.3c also shows that the concordance between return periods for the historical and generated annual NBS is poor.

For the monthly validation, SL and ARMA models give good results. The SVD model fails to adequately reproduce basic statistics (Figs 3.4c and 3.5c) and month to month correlations (Fig. 3.7c).

In general, validation results for the annual and monthly NBS for they other Lakes are the same as the one shown in the previous sections, and there is no inconsistancy in the results.

Table 3.4: Comparison of the three models.

ANNUAL VALIDATION (Erie)	**SL**			**ARMA**			**SVD**		
1. Basic annual statistics (Tab. 3.1)		1			1			-1	
2. Hurst coefficient (Tab. 3.1)		1			1			1	
3. Annual cross-correlation (Tab. 3.2)		0			0			0	
4. Autocorrelation (Fig. 3.1)		0			0			1	
5. NBS Frequency curve (Fig. 3.2)		1			0			-1	
6. NBS Return period (Fig. 3.3)		1			0			-1	
7. Surplus run length, RL (Tab. 3.3)		1			0			0	
8. Surplus run volume, RS (Tab. 3.3)		1			0			0	
9. Partial record runs (not shown)		0			0			0	
MONTHLY VALIDATION (Ontario)									
10. Basic monthly statistics (Figs 3.4, 3.5)		1			1			-1	
11. Monthly cross-correlation (Fig. 3.6)		0			0			0	
12. Month-to-month correlation (Fig. 3.7)		0			0			-1	
QUARTER MONTHLY LEVEL VALIDATION (Ontario)									
13 a). Basic level statistics, mean (Fig. 3.8)		-1			-1			-1	
13 b). Basic level statistics, max. (Fig. 3.9)		0			0			0	
14. Max. level frequency curve (Fig. 3.10)		0			0			-1	
15. Global max. level freq. curve (Fig. 3.11)		1			-1			-1	
16. Global max. level return period (Fig. 3.12)		1			-1			-1	
SCORE	1	0	-1	1	0	-1	1	0	-1
TOTAL	9	7	1	3	11	3	2	6	9

Finally, for the level validation the three models have difficulties in exactly reproducing the historical mean (Figures 3.8a,b,c). The mean of the quarter month historical levels are systematically higher then the simulated ones. In some cases the historical means are equal or greater than the simulated maximum mean levels. For the other characteristics considered, the SL model performances are excellent.

In the light of the results presented in Table 3.4, all the participants have selected the multivariate AR(1) model with shifting levels using SPIGOT for disaggregation (SL) as the best model for the final simulation of the Great Lakes NBS.

4. NBS data generation

The multivariate AR(1) model with shifting levels is used to generate 555 series of 90 years of annual NBS for the four Lakes. The annual values are disaggregated into monthly NBS using SPIGOT.

4.1 Simulation

The final data generation contain a total of 555 series of 90 years of monthly NBS. The series have been generated using SPIGOT software in 15 distinct runs. Each run containing 37 series of 90 years of monthly NBS. Thus, for each lake we have 555 files of generated monthly NBS, for a total of 2 220 files for the four lakes. In addition, for each lake three files respectively contain the 49 950 values of:

- annual NBS;
- minimun monthly NBS;
- maximun monthly NBS.

4.2 Validation of the generated annual NBS

The generated annual NBS data was arranged in 555 series of 90 years from which generated annual validation characteristics were computed. The comparison of historical and generated data is based on the:

- basic statistical parameters (mean, standard deviation, skewness);
- time series properties (correlogram, Hurst coefficient, surplus run length and volume);
- spatial properties (lag zero cross-correlation);
- frequential properties (annual NBS return period).

The mean, standard deviation, maximum and minimum of each validation characteristic at the annual scale are obtained and compared with those obtained from the historical record.

Since the validation of the generated monthly NBS previously shown has been performed on a sample of the final simulation (first 11 series of 90 years by 12 months), the readers are refered to section 3.2 for further details on monthly NBS validation.

4.2.1 Basic annual statistics

Table 4.1 shows a comparison of the mean, standard deviation and skewness coefficient for the historical and generated data. The mean, standard deviation, maximum and minimum values for each statistic are indicated.

4.2.2 Time series properties

Correlogram

Figures 4.1a,b,c,d respectively show for the four Lakes a graphical comparison of the historical and generated (average of the 555 series) lag 1 to lag 14 autocorrelation coefficients.

Table 4.1: Basic statistical parameters.

Lake		Historical	Generated data			
			mean	s.d.	max	min
SU	Mean (tcfs)	870,1	874,5	27,7	962,6	792,4
	Stand.Dev. (tcfs)	204,0	202,1	18,9	261,9	157,5
	Skewness	0,026	0,082	0,246	1,091	-0,639
MH	Mean (tcfs)	1 344,3	1 351,6	52,4	1 537,9	1 226,3
	Stand.Dev. (tcfs)	312,1	301,6	24,7	375,1	230,6
	Skewness	-0,052	-0,016	0,255	0,874	-0,918
ER	Mean (tcfs)	236,4	237,3	27,4	333,5	186,6
	Stand.Dev. (tcfs)	110,3	104,6	9,4	135,7	80,2
	Skewness	0,093	0,04	0,25	1,02	-0,75
ON	Mean (tcfs)	430,1	431,5	26,4	530,6	375,6
	Stand.Dev. (tcfs)	98,1	96,3	11,9	131,0	68,2
	Skewness	0,471	0,545	0,304	1,52	-0,187

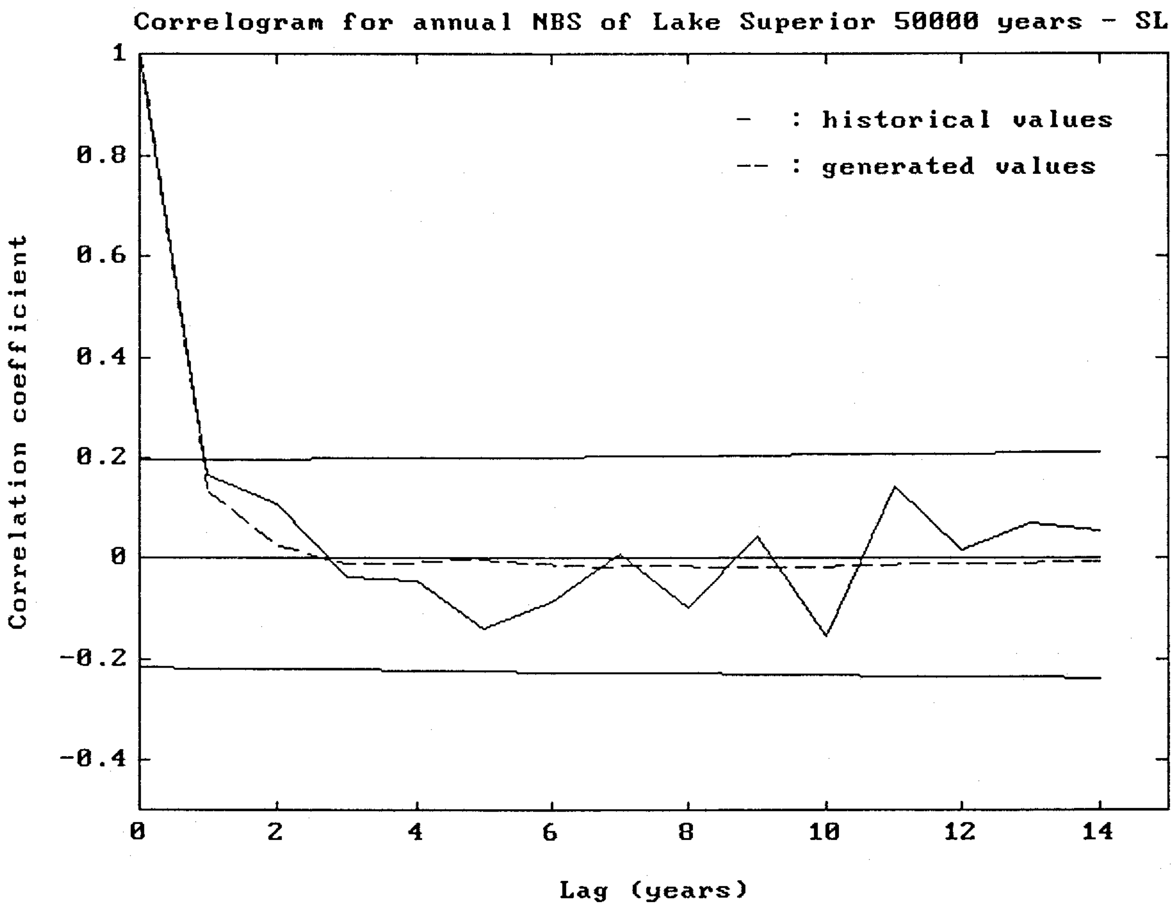


Figure 4.1 a: Correlogram of annual NBS for Lake Superior.

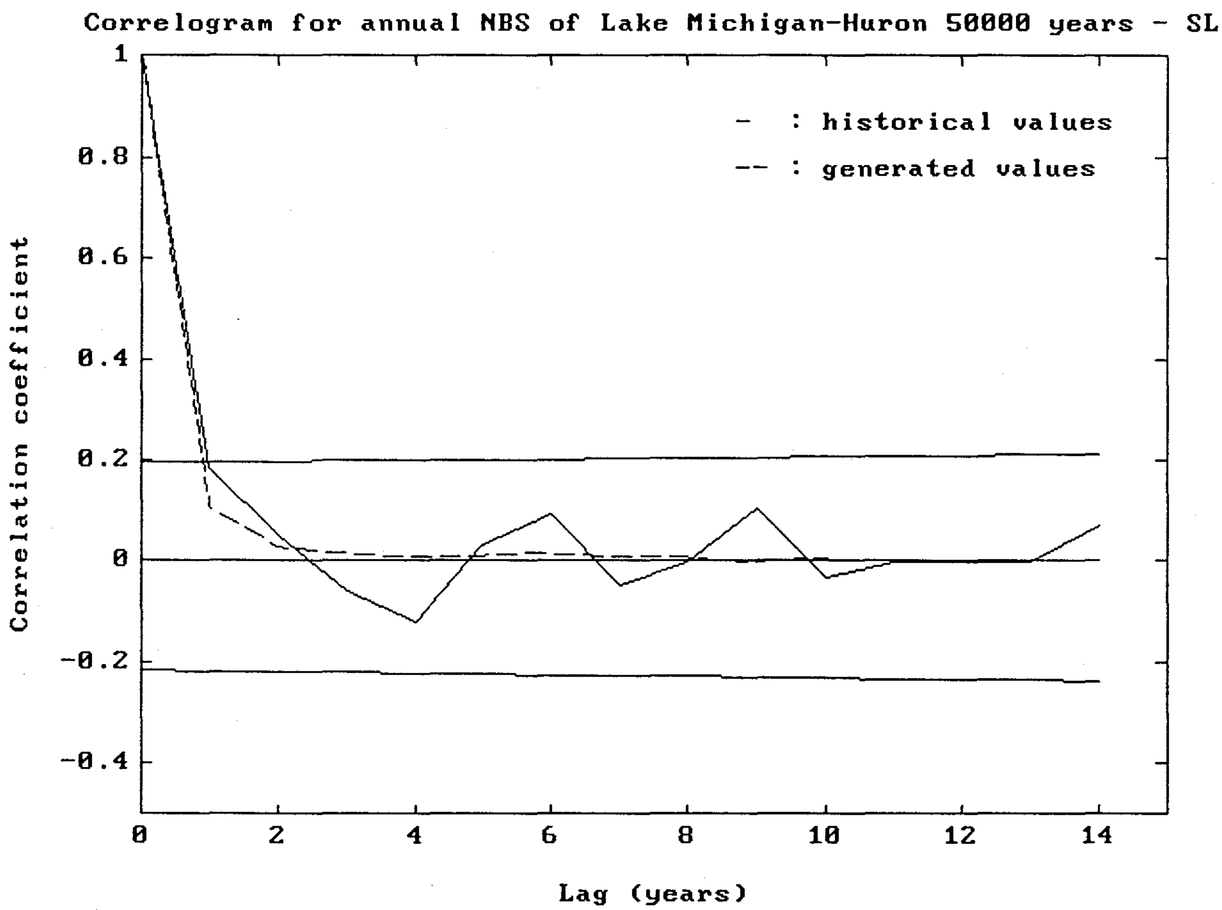


Figure 4.1 b: Correlogram of annual NBS for Lake Michigan-Huron.

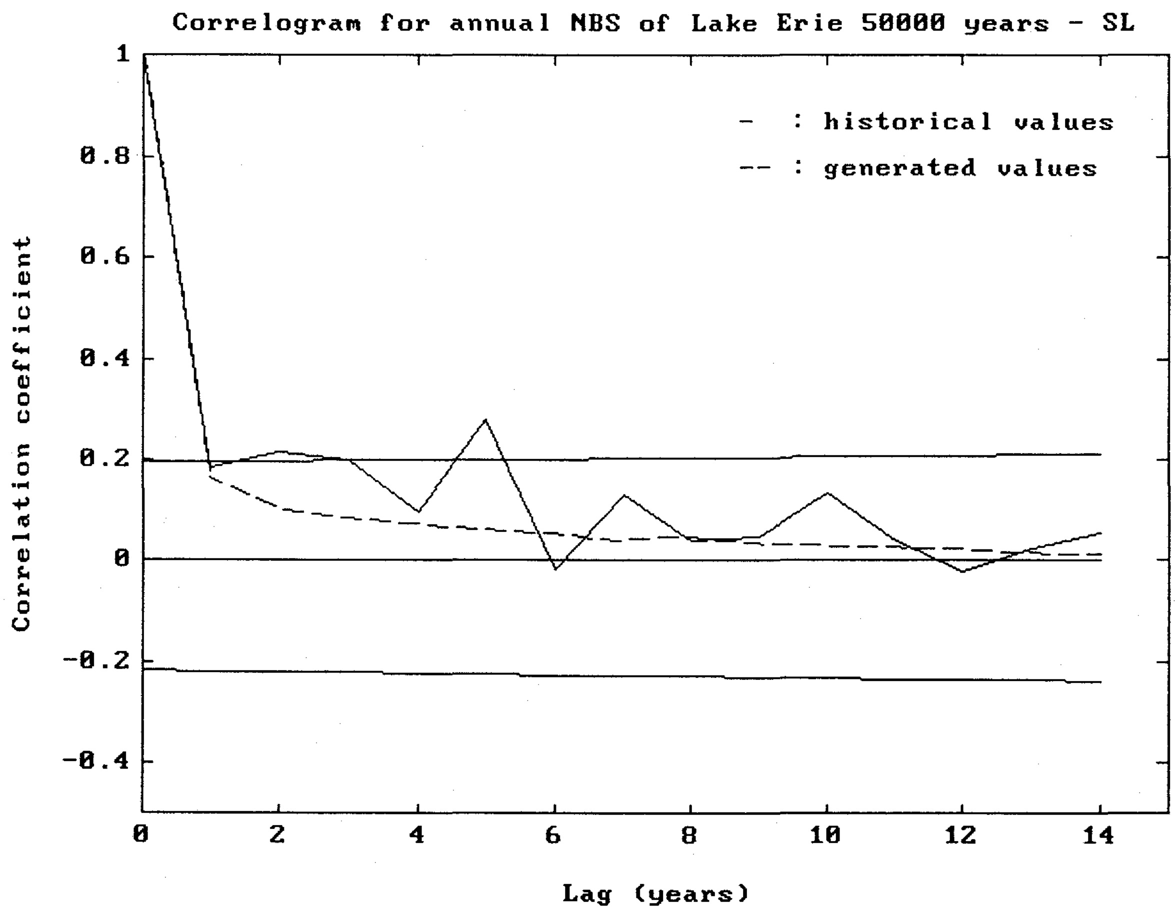


Figure 4.1 c: Correlogram of annual NBS for Lake Erie.

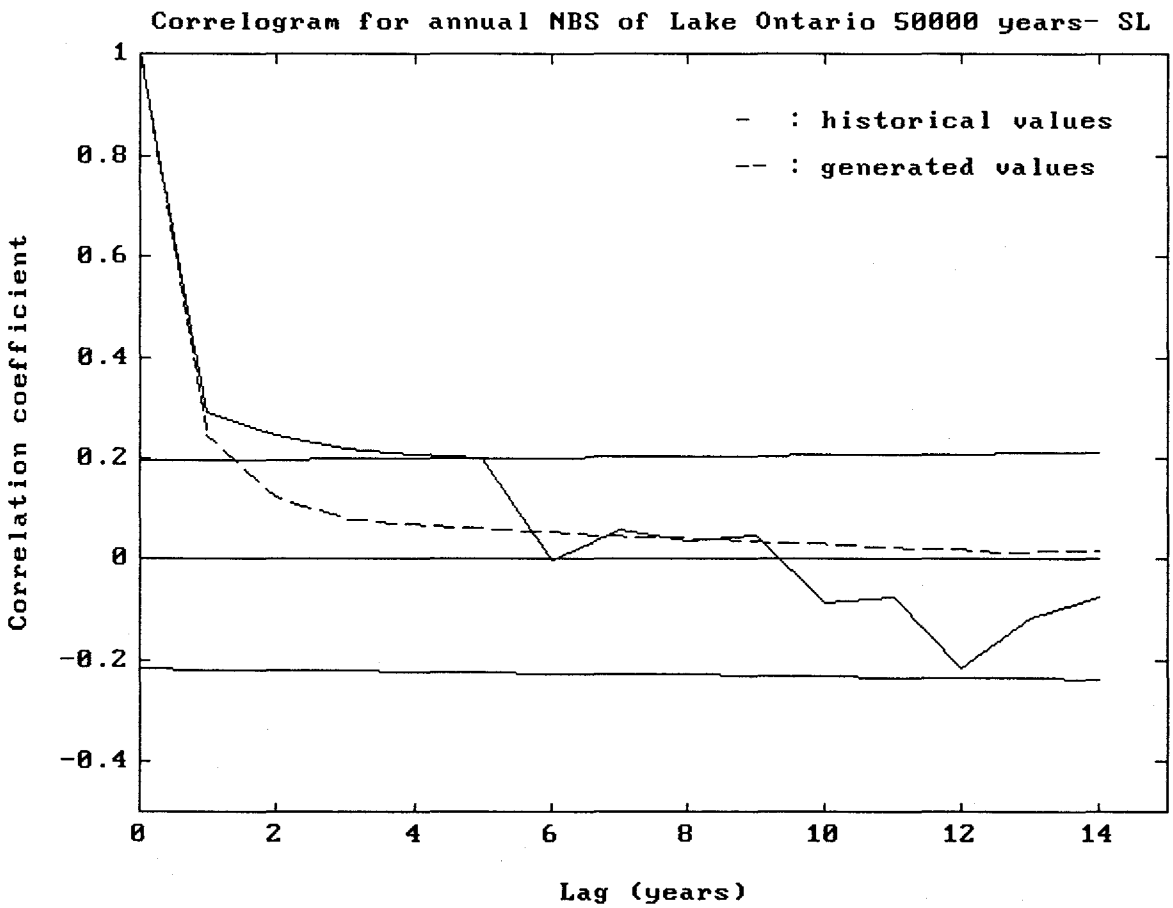


Figure 4.1 d: Correlogram of annual NBS Lake Ontario.

Hurst coefficient

Table 4.2 shows a comparison of the historical and generated Hurst coefficients. The mean, standard deviation, maximum and minimum values for the coefficient are indicated.

Table 4.2: Hurst coefficients.

Lake	Historical	Generated data			
	Hurst	mean	s.d.	max	min
Superior	0,647	0,657	0,064	0,831	0,457
Michigan-Huron	0,731	0,672	0,068	0,845	0,491
Erie	0,757	0,718	0,087	0,91	0,489
Ontario	0,752	0,724	0,085	0,905	0,504

Surplus run length (RL)

The 555 series of 90 years of generated data were used to determine the highest run length (RL) for a truncation level equal to the historical mean. The mean, standard deviation, maximum and minimum values of RL for the generated data at the four sites are shown with the corresponding historical values in Table 4.3. Only the 4-year running average run criteria were used.

Surplus run volume (RS)

Same as the surplus run length for the highest surplus run volume (RS) (Table 4.3).

Table 4.3: Surplus run analysis (4-year running average).

Lake		Historical	Generated data			
			mean	s.d.	max	min
SU	RL (year)	18	14,45	5,67	58	5
	RS ($Hm^3 \times 10^6$)	1,414	1,646	0,726	7,407	0,368
MH	RL (year)	22	16,45	9,01	63	4
	RS ($Hm^3 \times 10^6$)	3,531	2,904	1,849	14,206	0,739
ER	RL (year)	22	20,26	14,69	90	3
	RS ($Hm^3 \times 10^6$)	1,724	1,532	1,338	7,887	0,156
ON	RL (year)	18	18,1	12,6	90	2
	RS ($Hm^3 \times 10^6$)	1,525	1,402	1,192	8,082	0,126

4.2.3 Spatial properties

Annual cross-correlation

Table 4.4 shows a comparison of the historical and generated annual lag-zero cross-correlations among the sites. The mean, standard deviation, maximum and minimum values for the coefficients are indicated.

Table 4.4: Annual NBS lag-0 cross-correlations between the four Lakes.

	SU-MH	SU-ER	SU-ON	MH-ER	MH-ON	ER-ON
Historical	0,54	0,30	0,27	0,50	0,62	0,66
Generated						
mean	0,48	0,23	0,19	0,53	0,60	0,65
s.d.	0,12	0,12	0,11	0,10	0,08	0,09
max	0,82	0,55	0,49	0,78	0,80	0,84
min	0,003	-0,17	-0,18	0,17	0,31	0,36

4.2.4 Frequential properties

NBS return period

The return period (using the Weibull formula) of the historical and generated annual NBS for each lake is shown in Figures 4.2a,b,c,d.

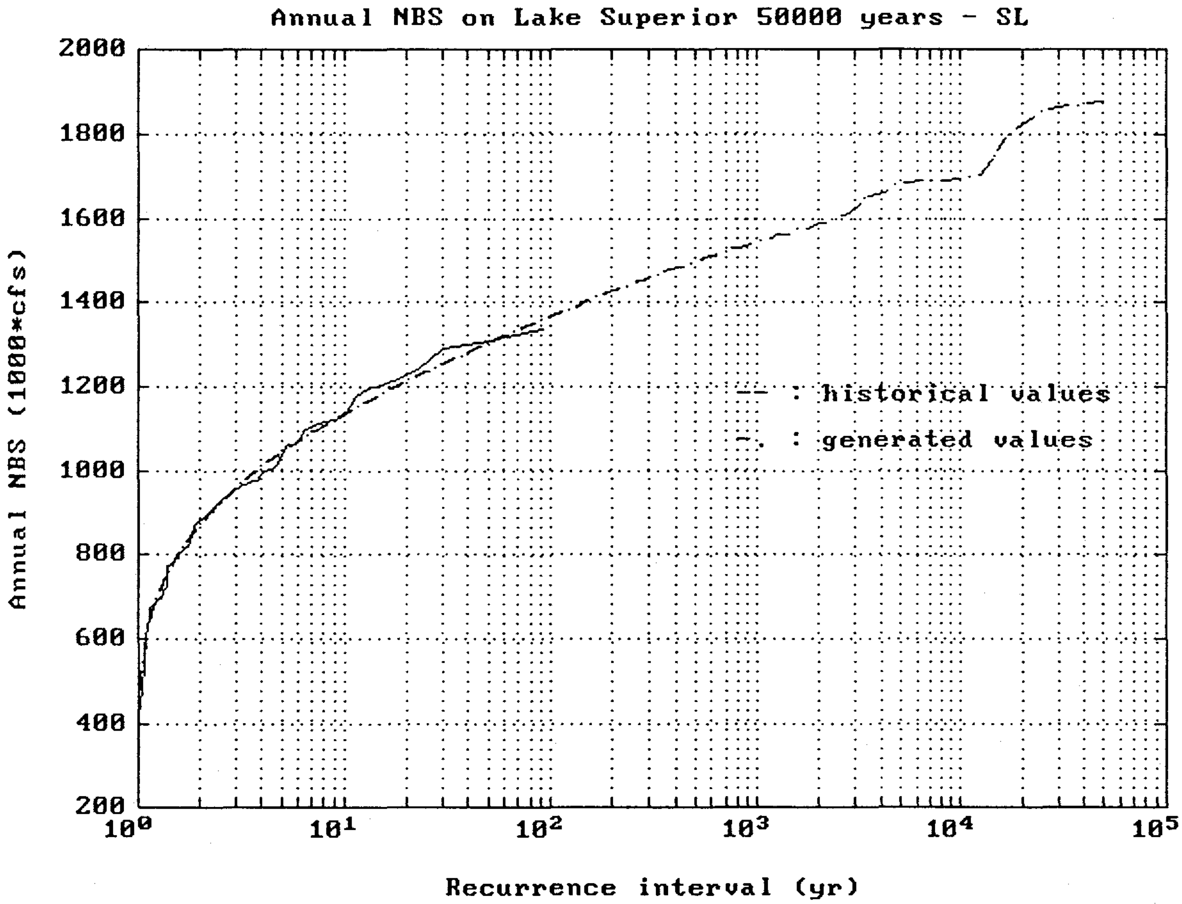


Figure 4.2 a: Annual NBS return period (using the Weibull formula) for Lake Superior.

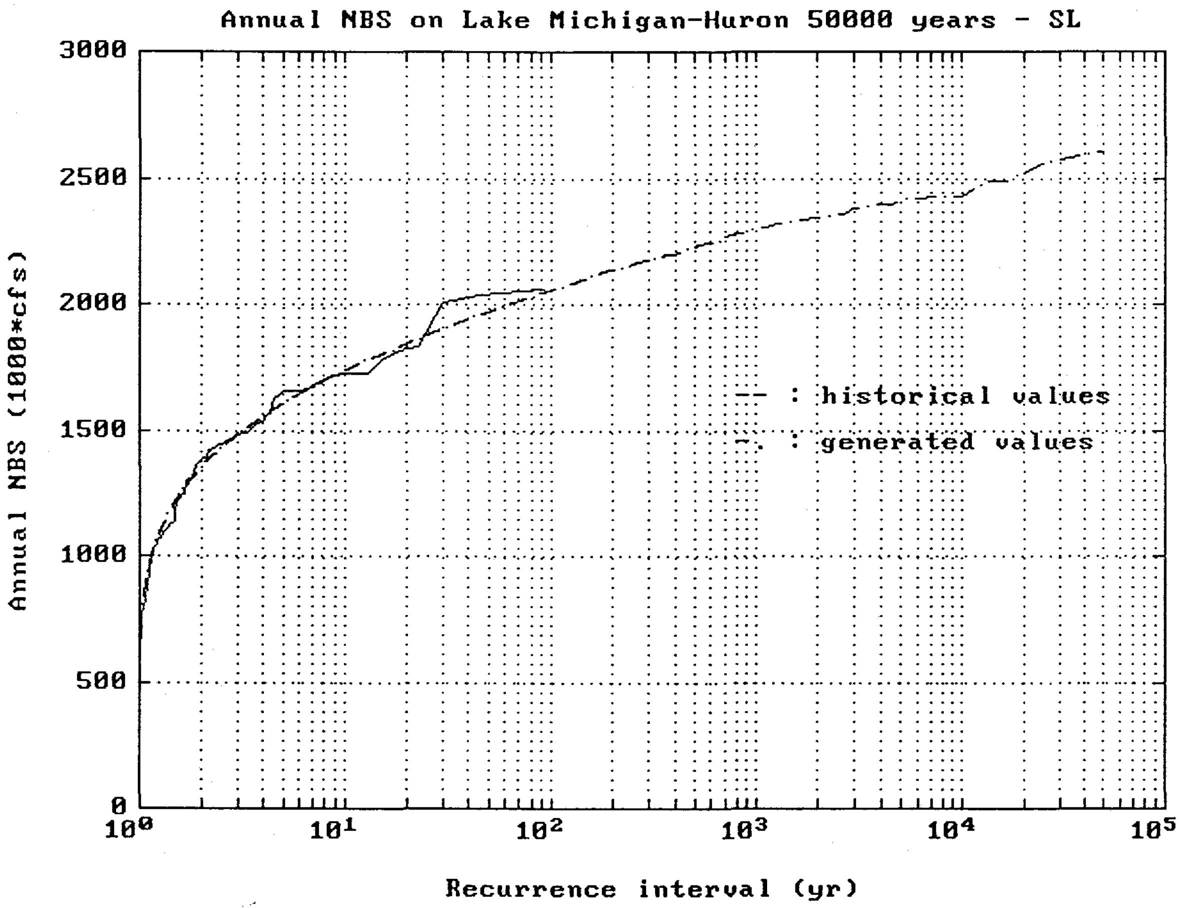


Figure 4.2 b: Annual NBS return period (using the Weibull formula) for Lake Michigan-Huron.

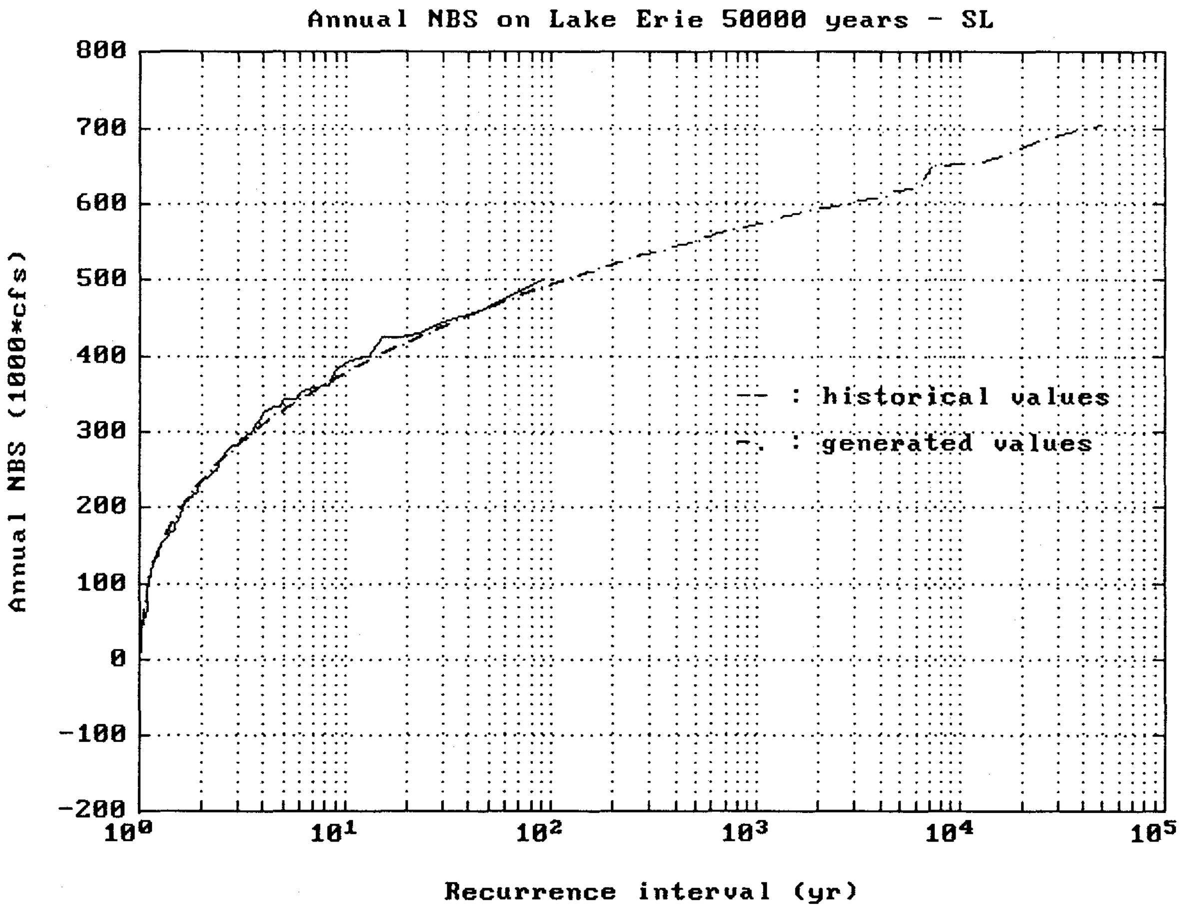


Figure 4.2 c: Annual NBS return period (using the Weibull formula) for Lake Erie.

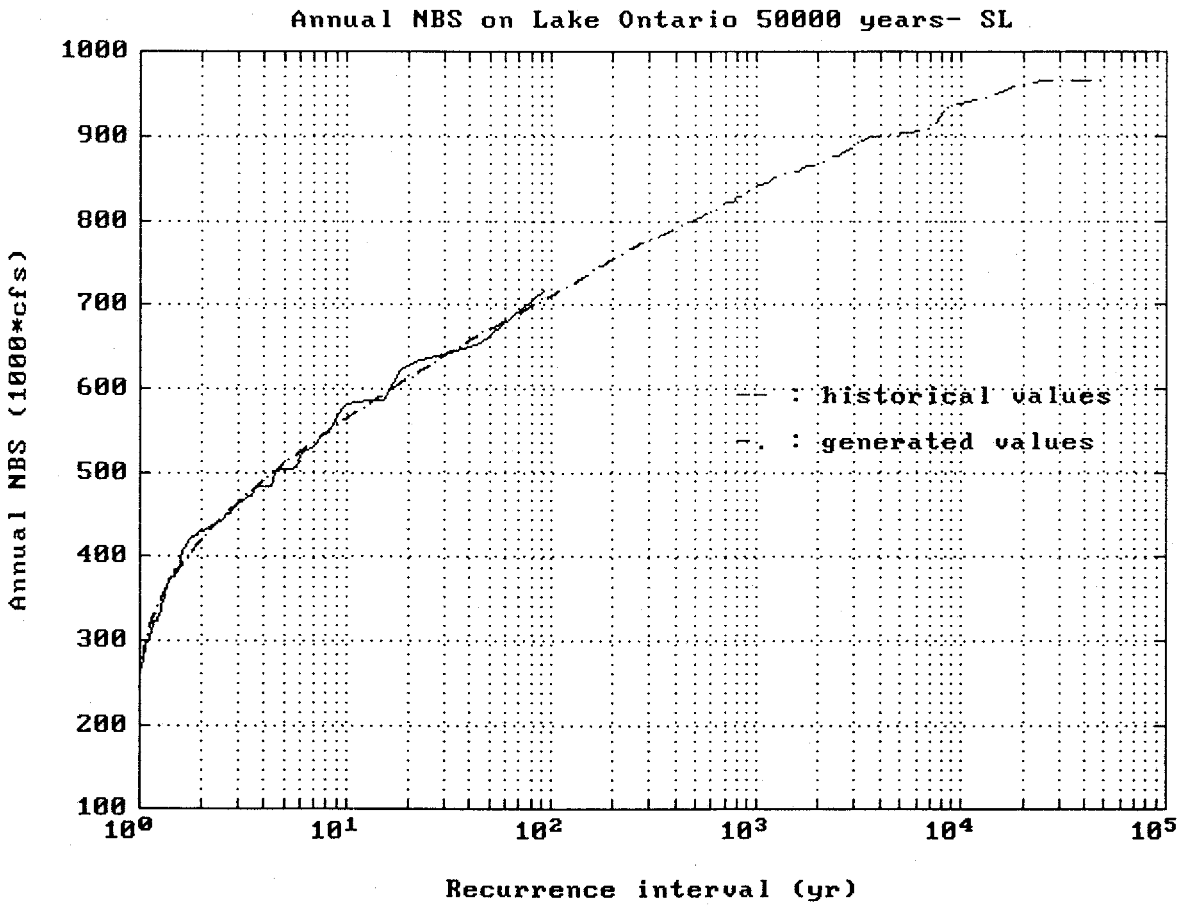


Figure 4.2 d: Annual NBS return period (using the Weibull formula) for Lake Ontario.

5. CONCLUSION

This report presents the first successful attemp made to simulate the monthly NBS using an indirect approach with a model that explicitly reproduces the shifts observed in the annual NBS of Lakes Erie and Ontario.

The extensive validation of the simulated data (NBS and levels) based on several criteria at different time scales ensures the good quality of the generated data set. The shifts, persistence and spatial correlations observed in the annual historical records are adequately reproduced by the simulation. The monthly NBS and levels characteristics validated over the generated samples are also well preserved by the model.

REFERENCES

Bernier, J. 1971. Modèles probabilistes à variables hydrologiques multiples et hydrologie synthétique. Proceedings of the International Symposium on Mathematical Models in Hydrology, Warsaw, Poland, 1: 619-635.

Boes, D.C. and Salas, J.D. 1978. Nonstationarity of the mean and the Hurst phenomenon. Water Resources Research, 14: 135.

Bruneau, P. and Rassam, J.-C. 1983. Application d'un modèle bayesien de détection de changements de moyennes dans une série. Hydrological Sciences Journal, 28: 341-354.

Cavadias, G.S. 1985. A multivariate seasonal model for streamflow simulation. Proceeding of the Colloque sur la recherche en hydrologie au Québec, Les Cahiers de l'ACFAS, No 35: 58-77.

C.E.A. (Canadian Electrical Association) 1990. Evaluation of monthly flow generation models DRAFT, Acres International Limited, Niagara Falls, Ontario, 134 p.

Fiering, M.B. 1964. Multivariate technique for synthetic hydrology. Journal of the Hydraulics Division, ASCE, 90: 43-60.

Fiering, M.B. and Jackson, B.B. 1971. Synthetic streamflows. American Geophysical Union, Water Resources Monograph, 1, Washington, D.C., 98 p.

Grygier, J.C. and Stedinger, J.R. 1988. Condensed disaggregation procedures and conseration corrections for stochastic hydrology, Water Resources Research, 10: 1574-1584.

Grygier, J.C. and Stedinger, J.R. 1990. SPIGOT a synthetic streamflow generation software package, Technical description, Version 2.6, Cornell University, Ithaca, New York 14853-3501.

Grygier, J.C. and Stedinger, J.R. 1991. SPIGOT a synthetic streamflow generation software package, User's Manual, Version 2.6, Cornell University, Ithaca, New York 14853-3501.

Hurst, H.E. 1951. Long-term storage capacity of reservoirs. Trans. ASCE, 116: 770-799.

I.G.L.L.B. (International Great Lakes Levels Board). 1973. Regulation of Great Lakes Water Levels. Appendix B, Lake Regulation, 3, 16 p.

Lane, W.L. 1979. Applied stochastic techniques, User's manual, Bureau of Reclamation, Engineering and Research Center, Denver, Co.

Lane, W.L. and Fervert, D.K. 1988. Applied stochastic techniques, User's manual, Bureau of Reclamation, Engineering and Research Center, Denver, Co., 5th Revision.

Lee, A.F.S. and Heghinian, S.M. 1977. A shift of the mean level in a sequence of independent normal random variables - a Bayesian approach. Technometrics, 19: 503-506.

Loucks, E.D. 1989. Modelling the Great Lakes Hydrologic-Hydraulic System, Ph. D. Thesis, University of Wisconsin, Madison.

Loucks, D.P., Stedinger, J.R. and Haith, D.A. 1981. Water Resources Planning and Analysis, Prentice-Hall, Englewood Cliffs, New Jersey.

Mandelbrot, B.B. and Wallis, J.R. 1968. Noah, Joseph, and operational hydrology, Water Resources Research, 4: 909-918.

Mann, H.B. and Whitney, D.R. 1947. On the test of whether one of two variables is stochastically larger than the other. Ann. Math. Stat. , 8: 50-60.

Matalas, N.C. 1967. Mathematical assessment of synthetic hydrology. Water Resources Research, 3: 937-945.

MATLAB. 1990. 386-MATLAB User's Guide. The MathWorks, Inc. , South Natick, MA.

Megerian, E. and Pentland, R. L. 1968. Simulation of Great Lakes basin water supplies. Water Resources Research, 4: 11-17.

Nathan, R.J., Adams, G.A. and McMahon, T.A. 1989. A case study of multi-site data generation applied to long term lake simulation, Proceedings of the Hydrology and Water Resources Symposium, Christchurch, N.Z., 48-52.

Quinn, F.H. 1982. Trends and extremes of Lake Erie water supplies. In Internat. Symp. Hydrometeo., Am. Water Resources Assoc., 267-270.

Quinn, F.H. and Guerra, B. 1986. Current Perspectives on the Lake Erie water balance. J. Great Lakes Res. 12 (2): 109-116.

Rassam, J.C., Mathier, L., Fagherazzi, L.D., Roy, R., Bobée, B. and Carballada, L. 1992. Beauharnois-Les Cèdres spillway design flood study with a stochastic approach. Hydro-Québec, Final report to the Experts committee, 105 p.

Salas, J.D. and Abdelmohsen, M.W. 1991. Streamflow drought statistics by stochastic simulation, Proceedings of the U.S.-P.R.C. Bilateral Symposium on Droughts and Arid-Region Hydrology, Tucson, Arizona.

Salas, J.D. and Boes, D.C. 1980. Shifting level modelling of hydrological series. Advances in Water Resources, 3: 59-63.

Salas, J.D., Delleur, J.W., Yevjevich, V. and Lane, W.L. 1980. Applied Modelling of Hydrologic Time Series. Water Resources Publications, Chelsea, 484 p.

Salas, J.D., Obeysekera, J.B. and Boes, D.C. 1981. Modeling of the equatorial lakes outflow, Statistical Analysis of Rainfall and Runoff, Water Resources Publications, Littleton, CO, 431-440.

Salas, J.D., Tabios, G. and Bartolini, P. 1985. Approaches to multivariate modeling of water resources time series, Water Resources Bulletin, 21: 683-708.

Santos, E. and Salas, J.D. 1991. Stepwise disaggregation scheme for synthethic hydrology, ASCE J. of Hydraulic Engineering, (Accepted).

Srikanthan, R., McMahon, T.A., Codner, G.P. and Mein, R.G. 1983. Practical aspects of multisite generation of streamflow data, Proceedings of the Hydrology and Water Resources Symposium, Hobart, 257-261.

Srikanthan, R., McMahon, T.A., Codner, G.P. and Mein, R.G. 1984. Theory and application of some multisite streamflow generation models, Institution of Engineers Australia, CE 26: 272-279.

Stedinger, J.R., Lettenmaier, D.P., and Vogel, R.M. 1985a. Multisite ARMA(1,1) and disaggregation models for annual streamflow generation, Water Resources Research, 4: 497-509.

Stedinger, J.R., Pei, D. and Cohn, T. 1985b. A condensed disaggregation model for incorporating parameter uncertainty into monthly reservoir simulations, Water Resources Research, 5: 665-675.

Valencia, D. and Schaake, Jr., J.C. 1973. Disaggregation processes in stochastic hydrology, Water Resources Research, 9: 580-585.

Yee, P., Edgett, R. and Eberhardt, A. 1990. Great Lakes-St. Lawrence River Regulation, What it Means and How it Works. Ontario Region of Environment Canada, Burlington, and the North Central Division of the U.S. Army Corps of Engineers, Chicago, 13 p.

Yevjevich, V. 1975. Generation of hydrologic samples case study of the Great Lakes. Hydrology Papers, Colorado State University, Fort Collins, Colorado, 72, 39 p.

Young, G.K. and Pisano, W.C. 1968. Operational hydrology using residuals. Journal of the Hydraulics Division, ASCE, 94: 909-923.