



US Army Corps
of Engineers

MISCELLANEOUS PAPER CERC-90-5

SIMULATION OF NONSTATIONARY, NON-GAUSSIAN WATER LEVELS ON THE GREAT LAKES

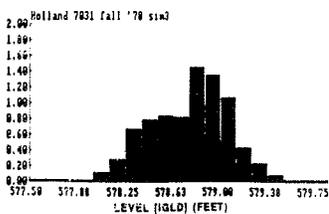
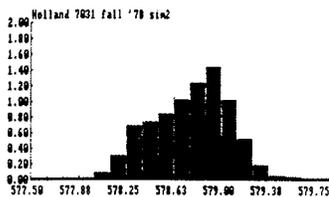
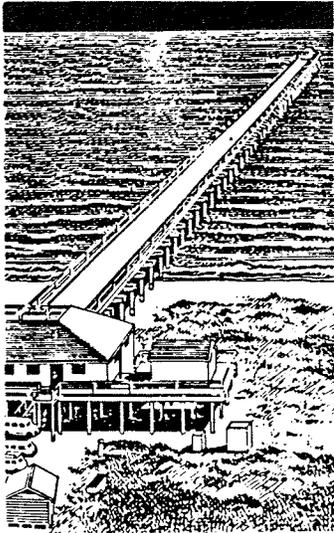
by

Todd L. Walton, Jr., Leon E. Borgman

Coastal Engineering Research Center

DEPARTMENT OF THE ARMY

Waterways Experiment Station, Corps of Engineers
3909 Halls Ferry Road, Vicksburg, Mississippi 39180-6199



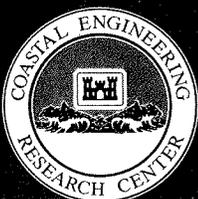
July 1990

Final Report

Approved For Public Release; Distribution Unlimited

Prepared for DEPARTMENT OF THE ARMY
US Army Corps of Engineers
Washington, DC 20314-1000

Under Work Unit No. 32487



When this report is no longer needed return it to
the originator.

The findings in this report are not to be construed as an
official Department of the Army position unless so
designated by other authorized documents.

The contents of this report are not to be used for
advertising, publication, or promotional purposes.
Citation of trade names does not constitute an
official endorsement or approval of the use of such
commercial products.

REPORT DOCUMENTATION PAGE				Form Approved OMB No. 0704-0188	
1a. REPORT SECURITY CLASSIFICATION Unclassified		1b. RESTRICTIVE MARKINGS			
2a. SECURITY CLASSIFICATION AUTHORITY		3. DISTRIBUTION / AVAILABILITY OF REPORT Approved for public release; distribution unlimited.			
2b. DECLASSIFICATION / DOWNGRADING SCHEDULE					
4. PERFORMING ORGANIZATION REPORT NUMBER(S) Miscellaneous Paper CERC-90-5		5. MONITORING ORGANIZATION REPORT NUMBER(S)			
6a. NAME OF PERFORMING ORGANIZATION USAEWES, Coastal Engineering Research Center		6b. OFFICE SYMBOL (if applicable)	7a. NAME OF MONITORING ORGANIZATION		
6c. ADDRESS (City, State, and ZIP Code) 3909 Halls Ferry Road Vicksburg, MS 39180-6199		7b. ADDRESS (City, State, and ZIP Code)			
8a. NAME OF FUNDING / SPONSORING ORGANIZATION US Army Corps of Engineers		8b. OFFICE SYMBOL (if applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER		
8c. ADDRESS (City, State, and ZIP Code) Washington, DC 20314-1000		10. SOURCE OF FUNDING NUMBERS			
		PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT ACCESSION NO. 32487
11. TITLE (Include Security Classification) Simulation of Nonstationary, Non-Gaussian Water Levels on the Great Lakes					
12. PERSONAL AUTHOR(S) Walton, Todd L., Jr.; Borgman, Leon E.					
13a. TYPE OF REPORT Final report		13b. TIME COVERED FROM Feb 88 TO Jan 90	14. DATE OF REPORT (Year, Month, Day) July 1990		15. PAGE COUNT 49
16. SUPPLEMENTARY NOTATION Available from National Technical Information Service, 5285 Port Royal Road, Springfield, VA 22161					
17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)		
FIELD	GROUP	SUB-GROUP			
			Coastal engineering Storm erosion		
			Coastal flooding Water level simulation		
			Coastal planning		
19. ABSTRACT (Continue on reverse if necessary and identify by block number)					
<p>A methodology for simulating nonstationary, non-Gaussian water levels on the Great Lakes is presented for use in providing input scenarios for erosion modeling on the Great Lakes. The methodology utilizes a low pass filtering technique to reduce the data to a stationary time series after which the data is transformed to a Gaussian series via use of the empirical distribution function coupled with a specialized tail fitting procedure for the extreme values in the data. The reduced series can then be resimulated in the frequency or time domain. The present series is resimulated in the frequency domain via a target spectrum, the spectrum of the reduced series. Inverse procedures are then utilized to simulate the original series. Comparisons of statistical and time series properties of the original and simulated series are made for a number of sites on the Great Lakes and it is found that the methodology provides a reasonable simulation for the original data.</p>					
20. DISTRIBUTION / AVAILABILITY OF ABSTRACT <input checked="" type="checkbox"/> UNCLASSIFIED/UNLIMITED <input type="checkbox"/> SAME AS RPT. <input type="checkbox"/> DTIC USERS			21. ABSTRACT SECURITY CLASSIFICATION Unclassified		
22a. NAME OF RESPONSIBLE INDIVIDUAL			22b. TELEPHONE (Include Area Code)	22c. OFFICE SYMBOL	

PREFACE

The investigation described in this report was authorized as part of the Civil Works Research and Development Program by the Headquarters, US Army Corps of Engineers (HQUSACE). This study was conducted by the work unit entitled "Stochastic Coastal Models," Work Unit 32487, under the Shore Protection and Restoration Program at the Coastal Engineering Research Center (CERC) of the US Army Engineer Waterways Experiment Station (WES). Messrs. John H. Lockhart, Jr., John G. Housley, and James E. Crews were the HQUSACE Technical Monitors. Dr. C. Linwood Vincent is CERC Program Manager.

The study was conducted from February 1988 through January 1990 by Drs. Todd L. Walton, Jr., CERC, and Leon E. Borgman, Professor of Statistics, University of Wyoming, Laramie, Wyoming.

This study was done under the general supervision of Dr. James R. Houston and Mr. Charles C. Calhoun, Jr., Chief and Assistant Chief, CERC, respectively, and under the direct supervision of Mr. Thomas W. Richardson, Chief, Engineering Development Division, CERC.

Commander and Director of WES during publication of this report was COL Larry B. Fulton, EN. Dr. Robert W. Whalin was Technical Director.

CONTENTS

	<u>Page</u>
PREFACE	i
INTRODUCTION	1
METHODOLOGY	3
Stationary Procedures	4
Normalization Procedure	8
Standardization Procedure	8
Spectrum Calculation	9
Simulation Spectrum Generation	9
Destandardization Procedure	11
Distribution Function Transformation	12
Nonstationary Transformation	15
RESULTS	15
CONCLUSIONS	16
ACKNOWLEDGMENTS	17
APPENDIX I. REFERENCES	19
APPENDIX II. NOTATION	21
FIGURES 1-21	

Simulation of Non-stationary, Non-Gaussian Water Levels on the Great Lakes

Todd L. Walton Jr. ¹,M. ASCE and Leon E. Borgman ²,M. ASCE

INTRODUCTION

Non-stationary non-Gaussian time series having correlation structure are common to the coastal engineering field. Often one does not know much about the underlying process or processes causing the series to vary as it does and hence cannot predict its future values with any certainty. Future values of the series can run a very different course than past values and hence a direct application of past events to future scenerios is not always a safe course of action. To do proper engineering design within a probabilistic framework it is safer to extract as much pattern from the process as feasible including extremal behavior, and then to prepare various possible future scenerios of the process for design purposes by some type of stochastic simulation. It should be emphasized that these are not deterministic predictions of what will occur at a particular time in the future. Rather, the set of such scenerios represent typical future conditions which the design or operation must withstand. The design should be examined relative to each such scenario to establish a range of response from "worst" to "best".

This paper will present one approach to simulation for a non-stationary non-Gaussian time series of hourly water levels on the Great Lakes. An earlier paper (Walton (1989)) discussed an autoregressive simulation approach in the time do-

¹Hydraulic Engineer,U.S. Army Corps of Engineers, Coastal Engineering Research Center, U.S.A.E. Waterways Experiment Station, Vicksburg,MS.

²Professor, Statistics Department, University of Wyoming, Laramie, Wyoming

main of the pseudocyclic behavior of the monthly mean water levels in the Great Lakes. One underlying need for dynamic simulation of water levels is to drive a cross shore sediment transport numerical model (see, for example, Kraus and Larson(1988) or Kriebel and Dean(1985)) for assessment of erosion volume versus frequency-of-occurrence curves. The dynamical aspects of the water level records are important to determination of such erosion volume/frequency-of-occurrence curves. As an example, a storm having a large peak water level but short duration can produce less erosion than a storm having a lower peak water level but longer duration.

Examples of the processes to be simulated herein are provided in Figs.1 and 2 which represent the average hourly water levels for the 1978 fall season (October through December) at two sites: Holland, Michigan; and Erie, Pennsylvania. These hourly water levels were measured at National Oceanic and Atmospheric Administration (NOAA) Great Lake water level gage sites, numbers 7031 and 3038, respectively, and were provided courtesy of the NOAA Great Lakes Acquisition Unit in Rockville, Maryland.

It is obvious from the time series plots that the short term (hourly) lake level is not stationary as evidenced by the heteroscedastic nature of the series fluctuations about a changing mean level. This complex structure of water levels is due to the underlying non-stationary forcing functions of wind, barometric pressure, and precipitation, which drive the water levels from equilibrium. The complexity of the water level series is also due to the complex response of the Great Lakes basins to the forcing functions. One such basin response to extreme event forcing function is seiching which is apparent in the water level records after an initial extreme event water level shock.

The non-Gaussian nature of these short term water levels is best seen in histograms of the water levels Figs. 3 and 4. These histograms depart from the Gaussian distribution as might be expected considering the numerous complex forcing functions which drive the Great Lakes water levels.

METHODOLOGY

For simulation of the present water levels, it is apparent from the above discussion that the simulation methodology utilized will have to deal with the non-stationary character of the data. In the present procedure this is accomplished via filtering/detrending/normal scores transformation procedures to reduce the data to stationarity, followed then by Gaussian correlated simulation, and, subsequently, by inverse operations to recapture the non-stationary non-Gaussian trending characteristics of the original data. The procedure deals with non-Gaussian data distribution functions of a varying unknown nature with tails representative of the measured extremes through a robust tail fitting procedure coupled with use of the empirical distribution function and a bootstrapping procedure for simulation purposes.

As a basis for later comparison of the simulated data to the measured data, the following four plots are considered: the time series plot (shown in Figs. 1 and 2); the histogram (Figs. 3 and 4); a spectral density plot, and an autocorrelation plot. Autocorrelation plots for the 1978 fall hourly water levels for the Holland and Erie sites are provided in Figs. 5 and 6 respectively. Spectral density plots for the same data as smoothed via a Gaussian window function are provided in Figs. 7 and 8.

The entire procedure for simulating future scenerios of water levels consists of a pattern analysis procedure for reduction of data to stationary Gaussian correlated

random noise coupled with a simulation procedure for reconstructing future scenerios of water levels from simulated Gaussian random noise having the same underlying correlation structure as the original series. The two part analysis procedure is best described via block diagram as in Fig 9. A step by step discussion of the procedure follows.

Stationarity Procedures

The data is first low pass filtered to determine a time varying mean and time varying standard deviation using a Gaussian shaped smoothing or convolving function of the form:

$$w(\tau) = \frac{\exp\left(\frac{-\pi\tau^2}{c^2}\right)}{c} \quad (1)$$

where c is a constant to adjust the width of $w(\tau)$. The Fourier transform of $w(\tau)$ is given as:

$$W(f) = \int_{-\infty}^{\infty} w(\tau) \exp(-i2\pi f\tau) \delta\tau \quad (2)$$

$$= \exp(-\pi f^2 c^2) \quad (3)$$

hence $w(0) = 1/c$ and $W(0) = 1$.

The effective width of $w(\tau)$ is defined as the width of a square pulse which has height $w(0)$ and area equal to $W(0)$ (i.e., the area under $w(\tau)$). Thus, the value "c" may be identified as the effective width of $w(\tau)$.

The filtered trend $vtrend(t)$ is defined in the continuous time domain as a convolution of the signal $v(t)$ with the Gaussian smoothing function $w(\tau)$:

$$vtrend(t) = \int_{-\infty}^{\infty} w(\tau)v(t - \tau)\delta\tau \quad (4)$$

where $w(\tau)$ is given a chosen effective width. Let $Vtrend(f)$ be the Fourier transform of $vtrend(t)$, and $V(f)$ the Fourier transform of $v(t)$.

$$Vtrend(f) = \mathcal{F}\{vtrend(t)\} = \int_{-\infty}^{\infty} vtrend(\tau) \exp(-i2\pi f\tau) \delta\tau \quad (5)$$

$$V(f) = \mathcal{F}\{v(t)\} = \int_{-\infty}^{\infty} v(\tau) \exp(-i2\pi f\tau) \delta\tau \quad (6)$$

Since convolution in the time domain is equivalent to multiplication in the frequency domain, we then have

$$Vtrend(f) = W(f) \cdot V(f) \quad (7)$$

It is convenient to make these operations in frequency domain to gain large savings in computer processing time possible through the use of the fast Fourier transform algorithm.

In practice, only a finite piece of data is available and it is natural to revert to discrete Fourier transform (DFT) definitions. In the discrete time domain the filtered trend $vtrend(t)$ is defined as:

$$vtrend(n \Delta t) = \Delta t \sum_{k=0}^{N-1} w(k \Delta t) v((n-k) \Delta t) \quad (8)$$

where the sequence $v(n \Delta t); n=0, \dots, N-1$ is made circularly periodic, with period N for times outside the defined data. That is;

$$v((n-k) \Delta t) = v((N+n-k) \Delta t) \quad (9)$$

$$v(N \Delta t) = v(0) \quad (10)$$

This artificial periodicity effects $vtrend(n \Delta t)$ about one effective width at the beginning and end of the time series. The one effective width arises because the

standard deviation of the Gaussian shape of $w(\tau)$ is $\frac{c}{\sqrt{2\pi}}$, hence three standard deviations are approximately equal to one effective width, "c". The Gaussian curve approaches zero at three standard deviations from the mean.

Using discrete Fourier transforms with

$$V_{trend}(m \Delta f) = \Delta t \sum_{n=0}^{N-1} v_{trend}(n \Delta t) \exp\left(\frac{-i2\pi mn}{N}\right) \quad (11)$$

$$V(m \Delta f) = \Delta t \sum_{n=0}^{N-1} v(n \Delta t) \exp\left(\frac{-i2\pi mn}{N}\right) \quad (12)$$

$$W(m \Delta f) = \Delta t \sum_{n=0}^{N-1} w(n \Delta t) \exp\left(\frac{-i2\pi mn}{N}\right) \quad (13)$$

where $v_{trend}(n \Delta t)$, $v(n \Delta t)$ and $w(n \Delta t)$ are all made circularly periodic, the trend computation is made in the discrete frequency domain by the following equation:

$$V_{trend}(m \Delta f) = W(m \Delta f) \cdot V(m \Delta f) \quad (14)$$

where

$$\Delta f = \frac{1}{(N \Delta t)} \quad (15)$$

and the discrete time equivalent of Eq. 3 is

$$W(m \Delta f) \cong \exp(-\pi(m \Delta f)^2 c^2) \quad (16)$$

Letting the quantity EW be the number of Δt increments in the effective width c , so that

$$c = (EW) \Delta t \quad (17)$$

it follows that

$$W(m \Delta f) \cong \exp\left(\frac{-\pi m^2 (EW)^2}{N^2}\right) \quad (18)$$

since

$$(\Delta t)(\Delta f) = \frac{1}{N} \quad (19)$$

Calculations are made very rapidly in the frequency domain using fast Fourier transform (FFT) techniques (Blahut,(1985)) and the resulting time series $vtrend(n \Delta t)$; $n=0,\dots,N-1$ is found by inverse fast Fourier transform (IFFT). In actual simulations the $vtrend(n \Delta t)$ series should be made longer than needed such that one effective width can be removed at the beginning and the end of the series due to the imposed circular periodicity assumptions as noted previously.

The local variance can be defined in an analogous manner in the discrete time domain as a convolution of the instantaneous variance with the same smoothing filter $w(\tau)$:

$$vvar(n \Delta t) = \sum_{k=0}^{N-1} w(k \Delta t) ivar((n - k) \Delta t) \quad (20)$$

where the instantaneous variance is defined as

$$ivar(n \Delta t) = (v(n \Delta t) - vtrend(n \Delta t))^2 \quad (21)$$

The local variance can then be found as before in the frequency domain by a multiplication of the Fourier transform of $w(n \Delta t)$ by the Fourier transform of the instantaneous variance. A different effective width can be used for the variance filtering operation if so desired. On inverse Fourier transform, the time domain standard deviation series $vsd(n \Delta t)$; $n=0,\dots,N-1$ can be found as

$$vsd(n \Delta t) = \sqrt{vvar(n \Delta t)} \quad (22)$$

Again as in the case of the $vtrend$ series, in actual simulations the $vsd(n \Delta t)$ series should be made longer than needed such that at least one effective width can be

discarded at the beginning and end of the series. In this regard the longer of the two effective widths (trend and variance filter effective width) should be used as the basis for discarding information in order to keep the two series properly aligned.

The required stationary residuals are then found as;

$$vresid(n \Delta t) = \frac{v(n \Delta t) - vtrend(n \Delta t)}{vsd(n \Delta t)} \quad (23)$$

These residuals will locally have a mean of zero and standard deviation of unity.

Normalization Procedure

The residuals $vresid(n \Delta t)$ are then order ranked by size (smallest to largest) with the integer series $rank(n \Delta t); n=0, \dots, N-1$ being the rank number series where each value of $rank(n \Delta t)$ corresponds to the rank number of $vresid(n \Delta t)$. Hence the series $rank(n \Delta t); n=0, \dots, N-1$ ranges between 1 and N. An estimate of the distribution function of the detrended and normalized (mean zero, variance one) data after ranking may be stated as:

$$\hat{F}(vresid(n \Delta t)) = \frac{rank(n \Delta t)}{N + 1} \quad (24)$$

Each $vresid(n \Delta t)$ is thus replaced by a standardized normal value "zscore($n \Delta t$)" having the ranking associated with the empirical probability distribution fractile value $\frac{rank}{N+1}$. This transformation is accomplished via the following equation:

$$zscore(n \Delta t) = \Phi^{-1}\left(\frac{rank(n \Delta t)}{N + 1}\right) \quad (25)$$

where $\Phi^{-1}()$ is the inverse normal distribution function which can be solved for numerically (Zelen and Severo(1964) eq. 26.2.23).

Standardization Procedure

As a final, usually very small adjustment, the $zscore(n \Delta t)$; $n=0, \dots, N-1$ series is then restandardized to a zero mean and unit variance via

$$z(n \Delta t) = \frac{zscore(n \Delta t) - zbar}{zsd} \quad (26)$$

where $zbar$ and zsd are the mean and standard deviation of the $zscore$ series.

Spectrum Calculation

The "target" spectrum for simulation purposes is a smoothed version of the line spectrum of the series $z(n \Delta t)$; $n=0, \dots, N-1$. The discrete Fourier transform (DFT) of the $z(n \Delta t)$; $n=0, \dots, N-1$ series is:

$$Z(m \Delta f) = \Delta t \sum_{n=0}^{N-1} z(n \Delta t) \exp\left(\frac{-i2\pi mn}{N}\right) \quad (27)$$

and the spectral line at $f = \frac{m}{N\Delta t}$ is

$$Sz(m \Delta f) = \frac{|Z(m \Delta f)|^2}{N \Delta t} \quad (28)$$

The spectral lines are smoothed over frequency in order to obtain adequate spectral density estimates. This is achieved with Gaussian smoothing in the frequency domain via the convolution operation:

$$Sbar(m \Delta f) = \sum_{k=0}^{N-1} \left\{ \frac{\exp\left(\frac{-\pi(k\Delta f)^2}{c_f^2}\right)}{c_f} \right\} \cdot Sz((m-k) \Delta f) \quad (29)$$

For numerical efficiency this operation is actually performed in the lag time domain via the equation

$$sbar(n \Delta t) = sz(n \Delta t) \cdot \frac{\exp\left(\frac{-\pi n^2 (EW)^2}{N^2}\right)}{N} \quad (30)$$

where $sz(n \Delta t)$ is the inverse discrete Fourier transform (IDFT) of $Sz(m \Delta f)$, and EW is an effective width as before where $EW = \frac{ct}{\Delta f} = c_f N \Delta t$. Returning to the frequency domain, the smoothed "target" spectrum to be used in the simulation process, $Sbar(m \Delta f)$, is the DFT of $sbar(n \Delta t)$.

Simulation Spectrum Generation

The smoothed "target" spectrum is then standardized to unit area to require

$$\sum_{m=0}^{N-1} Sbar(m \Delta f) \Delta f = 1.0 \quad (31)$$

Simulation is based on the random behavior of a real Gaussian covariance stationary periodic (period = $N \Delta t$) sequence $zsim(n \Delta t); n=0, \dots, N-1$ with a zero mean and variance equal to one. The DFT of such a random series is

$$Zsim(m \Delta f) = U_m - iV_m = \Delta t \sum_{n=0}^{N-1} zsim(n \Delta t) \exp\left(\frac{-i2\pi mn}{N}\right) \quad \text{for } 0 \leq m \leq \frac{N}{2} \quad (32)$$

where U_m and V_m are independent, normally distributed random variables with zero mean and variance given by Borgman(1973,1982,1990 in press) and Miller and Borgman(1985) as:

$$Variance(U_m) = Variance(V_m) = \frac{N \Delta t S zsim(m \Delta f)}{2} \quad \text{if } 0 < m < \frac{N}{2} \quad (33)$$

$$Variance(U_m) = N \Delta t S zsim(0) \quad \text{if } m = 0$$

$$Variance(U_m) = N \Delta t S zsim\left(\frac{N \Delta f}{2}\right) \quad \text{if } m = \frac{N}{2}$$

$$Variance(V_m) = 0 \quad \text{if } m = 0, \frac{N}{2}$$

where $S zsim(m \Delta f)$ is the desired line spectrum to simulate. For $\frac{N}{2} < m < N$,

$$U_m = U_{N-m} \quad \text{and} \quad V_m = -V_{N-m} \quad (34)$$

The frequency series $U_0, U_1, V_1, \dots, U_{\frac{N}{2}-1}, V_{\frac{N}{2}-1}, U_{\frac{N}{2}}$ is thus generated by generating N independent standard Gaussian variates Z_1, Z_2, \dots, Z_N and then calculating the frequency series in accordance with $Szsim(m \Delta f) = Sbar(m \Delta f)$:

$$U_0 = \sqrt{N \Delta t Sbar(0)} Z_1 \quad (35)$$

$$U_1 = \sqrt{N \Delta t Sbar(\Delta f)/2} Z_2$$

$$V_1 = \sqrt{N \Delta t Sbar(\Delta f)/2} Z_3$$

$$U_2 = \sqrt{N \Delta t Sbar(2 \Delta f)/2} Z_4$$

$$V_2 = \sqrt{N \Delta t Sbar(2 \Delta f)/2} Z_5$$

etc.

$$U_{\frac{N}{2}-1} = \sqrt{N \Delta t Sbar((\frac{N}{2} - 1) \Delta f)/2} Z_{N-2}$$

$$V_{\frac{N}{2}-1} = \sqrt{N \Delta t Sbar((\frac{N}{2} - 1) \Delta f)/2} Z_{N-1}$$

$$U_{\frac{N}{2}} = \sqrt{N \Delta t Sbar(\frac{N}{2} \Delta f)} Z_N$$

The simulated Gaussian time series with zero mean and unit variance $zsim(n \Delta t); n=0, \dots, N-1$ is then generated via the IDFT of the series $Zsim(m \Delta f) = U_m - iV_m; m=0, \dots, N-1$ with the equation:

$$zsim(n \Delta t) = \Delta f \sum_{m=0}^{N-1} (U_m - iV_m) \exp\left(\frac{i2\pi mn}{N}\right) \quad (36)$$

Destandardization Procedure

After generating the zero mean, unit variance time series $zsim(n \Delta t); n=0, \dots, N-1$, it is rescaled with the $zscore(n \Delta t); n=0, \dots, N-1$ series mean "zbar" and standard

deviation "zsd". This is done via the equation

$$zscosim(n \Delta t) = zsd \cdot zsim(n \Delta t) + zbar \quad (37)$$

to create a simulated Gaussian time series $zscosim(n \Delta t); n=0, \dots, N-1$ having the same correlation structure, mean, and standard deviation, as the $zscore(n \Delta t); n=0, \dots, N-1$ series.

Distribution Function Transformation

At this point the Gaussian simulated series $zscosim(n \Delta t); n=0, \dots, N-1$ is transformed back to the probability distribution function of the detrended series to obtain the resulting non-Gaussian trended simulated time series $vsimres(n \Delta t); n=0, \dots, N-1$. This is done via back interpolation using the empirical distribution function in the central portion of the distribution along with a functional form for the tails of the distribution to make the distribution function more "robust". The back interpolation from the simulated normal score time series $zscosim(n \Delta t)$ is achieved from inverting

$$\hat{F}(vsimres(n \Delta t)) = \Phi(zscosim(n \Delta t)) \quad (38)$$

where $\Phi(z)$ is the distribution function for a standard normal probability law. This may be written as:

$$vsimres(n \Delta t) = \hat{F}^{-1}(\Phi(zscosim(n \Delta t))) \quad (39)$$

where $\hat{F}^{-1}()$ is the inverse empirical probability distribution function. In practice, the step

$$u = \Phi(zscosim(n \Delta t)) \quad (40)$$

is performed with an algorithm from Zelen and Severo(1964) (eq. 26.2.17) and then the operation

$$v_{simres}(n \Delta t) = \hat{F}^{-1}(u) \quad (41)$$

is achieved by direct interpolation within the ranked detrended original data, at least for $1/(N+1) \leq u \leq N/(N+1)$.

If $u < 1/(N+1)$ (lower tail) or $u > N/(N+1)$ (upper tail), another procedure was selected. It was found that the upper and lower tails of many common probability laws such as the normal, exponential, gamma, and lognormal could be fitted either exactly or with high accuracy of approximation by a representation of the form:

$$v = z_1 + (a + bT)^c \quad \text{for upper tail} \quad (42)$$

or

$$v = z_1 - (a + bT)^c \quad \text{for lower tail}$$

where

$$T = \sqrt{-2\ln(1-U)} \quad \text{upper tail unbounded above} \quad (43)$$

$$T = U - U_1 \quad \text{upper tail bounded above}$$

$$T = \sqrt{-2\ln(U)} \quad \text{lower tail unbounded below}$$

$$T = U_1 - U \quad \text{lower tail bounded below}$$

The constants a,b,c, and z_1 are based on extremal values of the $\hat{F}(v_{resid}(n \Delta t))$ empirical distribution function estimate.

For the upper tail unbounded above, let $z_1 < z_2 < z_3$ be selected values within the upper tail of the ranked $v_{resid}(n \Delta t)$ values, and $u_1 < u_2 < u_3$ be the corresponding values of $\text{rank}/(N+1)$.

If the upper tail is bounded by some value z_{top} , then set $z_3 = z_{top}$ and $u_3 = 1.0$. The other two sets of (z,u) are obtained as before.

For the lower tail unbounded below, the procedure is very similar. Let $z_3 < z_2 < z_1$ be selected within the lower ranked values of $vresid(n \Delta t)$ and $u_3 < u_2 < u_1$ be the $rank/(N+1)$ values associated with them.

If the lower tail is bounded below by z_{bottom} , then set $z_3 = z_{bottom}$ and $u_3 = 0.0$. The other two pairs of (z, u) are picked as for the unbounded below case.

The z_1 in the previous tail representation formula coincides exactly with the z_1 values assigned for each tail here. The u_1 values needed for the bounded cases in each tail are the corresponding u_1 values selected here.

This leaves the a,b,and c values which can be computed by:

$$c = \frac{\ln\left(\frac{z_2 - z_1}{z_3 - z_1}\right)}{\ln\left(\frac{T_2 - T_1}{T_3 - T_1}\right)} \quad (44)$$

$$\ln(b) = \frac{\ln |z_3 - z_1|}{c} - \ln |T_3 - T_1| \quad (45)$$

$$a = -bT_1 \quad (46)$$

where T_1 is the value of T computed from u_1 in the previous definition of T.

The direct empirical interpolation is performed for $lowertail u_1 \leq u \leq uppertail u_1$, while the tail approximation formulas are used for values outside this interval.

Although the upper and lower tail represent a very small fraction of the simulated values, they are important in many engineering problems where extremes are significant. In most published "bootstrap" procedures, the simulations are restricted to lying between the maximum and minimum of the original data. In predicting the future, these methods appear artificial, since it seems obvious that extreme values

outside the historical data base may occur eventually unless there is some sort of finite bounds due to the physical constraints.

The procedure presented represents a type of tail-extended bootstrap simulation generalized with a normal score transformation. A great deal of research and comparison with common population probability laws was involved in selecting the tail formulation recommended above (Borgman,(1989)). It appears reasonable under circumstances examined to date.

Nonstationarity Transformation

The last transformation necessary to produce a simulated series $vsim(n\Delta t); n=0, \dots, N-1$ is an inverse transformation using the computed $vtrend(n\Delta t)$ and $vsd(n\Delta t)$ series. The final simulated non-stationary, non-Gaussian time series is found as:

$$vsim(n \Delta t) = vsd(n \Delta t) \cdot vsimres(n \Delta t) + vtrend(n \Delta t) \quad (47)$$

This operation is performed in the time domain.

RESULTS

Results of the simulation procedure for one simulation run for the Erie site is presented in time series plot, Fig. 10, histogram Fig. 11, autocorrelation Fig.12, and smoothed spectrum (via a Gaussian window) Fig.13. These figures can be compared to the actual series plots in Figs.2,4,6, and 8. The low pass trend and standard deviation series for the Erie site are shown in Fig.14 and Fig.15 respectively. The effective filter width utilized for the trend series was 20 hours while the effective filter width utilized for the standard deviation series was 50 hours. As the effective width

chosen is somewhat of an ad hoc procedure, no justification for these choices of filter width is given. A natural extension of these procedures is to analyze and synthesize the smoothing trend and local standard deviation series in a similar manner rather than use the deterministic output directly from the data analysis. This would not involve any new theory but rather would introduce a hierarchy of simulations. It is not being done here because it would introduce unnecessary complexity in explaining the method.

Results of the simulation procedure for three simulation runs at the Holland site are provided in Figs. 16 thru 19 for the time series, histogram, autocorrelation and smoothed spectra (smoothed via a Gaussian window). The low pass trend and standard deviation series for this site were computed with the same effective filter width as the Erie site. Results are shown in Figs. 20 and 21.

CONCLUSIONS

The results of the procedure seem to "mimic" the original data in a reasonable fashion while still incorporating process noise consistent with the original series. The method can analyze and simulate a large time series extremely fast. The method allows for exact duplication of the univariate probability law contained in the data behavior and is thus capable of producing non-gaussian process behavior related to skewness and higher order moments. It is understood that some of the higher order bivariate moments may not be properly reproduced by this transformation to gaussian (and the inverse transformation). Applications for which the higher order bivariate (trivariate, etc.) moments are important should consider other techniques, perhaps based on Volterra series. The method also allows for the treatment of

non-stationarity in mean and variance either through including the bursts of non-stationarity of data in deterministic fashion or through a hierarchy of simulations as suggested in the text.

ACKNOWLEDGMENTS

The authors would like to especially thank Mr. Jeff Oyler and Mr. Harry Lippincott, Great Lakes Acquisition Unit, National Oceanic and Atmospheric Administration, for their help in obtaining the water level data.

The work presented herein was conducted under the Stochastic Coastal Models work unit of the Shore Protection and Restoration Program, United States Army Corps of Engineers, Coastal Engineering Research Center, Waterways Experiment Station. Permission was granted by the Chief of Engineers to publish this information.

APPENDIX I. REFERENCES

- Bendat, J.S. and Piersol, A.G. (1986). *Random Data: Analysis and Measurement Procedures*, Second Ed., John Wiley and Sons Inc., New York.
- Blahut, R.E. (1985). *Fast Algorithms for Digital Signal Processing*, Addison-Wesley Publishing Co., Reading, MA.
- Borgman, L.E. (1973). "Statistical Properties of Fast Fourier Transform Coefficients Computed from Real-Valued, Covariance-Stationary, Periodic Random Sequences", Research Paper No. 23, University of Wyoming, Laramie, Wyoming.
- Borgman, L.E. (1982). "Techniques for Computer Simulation of Ocean Waves", *Topics in Ocean Physics*, Italian Physical Society, 387-416.
- Borgman, L.E. (). "An Empirical Tail-Extended, Bootstrap Procedure for Nonlinear Simulations." Submitted (1989) to *Computers and Geoscience*.
- Borgman, L.E. (1990). "Irregular Ocean Waves: Kinematics and Forces", *The Sea*, Vol. 9, Ocean Engineering Science.
- Kraus, N. and Larson, M. (1988). "Prediction of Initial Profile Adjustment of Nourished Beaches to Wave Action", *Proceedings, Beach Preservation Technology Conference, Florida Shore and Beach Preservation Association*, Gainesville, Florida.
- Kriebel, D.L. and Dean, R.G. (1985). "Numerical Simulation of Time-Dependent Beach and Dune Erosion", *Coastal Engineering*, Vol. 9, No. 3, Sept. 1985, 221- 245.

Miller,S.M. and Borgman,L.E.(1985)."Spectral Type Simulation of Spatially Correlated Fracture Set Properties",Mathematical Geology, Vol.17, No.1, 41-52.

Walton,T.L. Jr.(1989). "Simulating Great Lakes Water Levels For Erosion Prediction," Journal of Coastal Research, Vol.5, Issue 3, 377-389.

Zelen,M. and Severo,N.C.(1964)."Probability Functions", Chapter 26, Handbook of Mathematical Functions, Ed. M.Abramowitz and I.A.Stegun, Dover Publications Inc., New York, N.Y.

APPENDIX II. NOTATION

a, b = empirically fit constants;

c = Gaussian time domain weighting function constant, also empirically fit constant;

c_f = Gaussian frequency domain weighting function constant;

EW = effective width of filter;

f = frequency;

$ivar(n \Delta t)$ = discrete time instantaneous variance series;

m, n = index counters;

N = number of signal samples;

$rank(n \Delta t)$ = discrete time integer ranking series;

$sbar(n \Delta t)$ = inverse discrete Fourier transform of $Sbar(m \Delta f)$ series;

$Sbar(m \Delta f)$ = smoothed spectral line function of z series;

$Sz(m \Delta f)$ = spectral line function of z series;

$Szsim(m \Delta f)$ = desired spectral line function series for simulation;

t = time;

T = tail function value;

$u_i; i = 1, N$ = fractile value

U_m = real part of $Zsim(m \Delta f)$ series;

$v(t)$ = signal function, time representation;

$V(f)$ = signal function, frequency representation;

V_m = imaginary part of $Zsim(m \Delta f)$ series;

$vresid(n \Delta t)$ = discrete time stationary signal series;

$vsd(n \Delta t)$ = discrete time standard deviation series;

$vsim(n \Delta t)$ = discrete time simulated signal series;

$vsimres(n \Delta t)$ = discrete time simulated stationary signal series;

$vtrend(t)$ = filtered trend function, time representation;

$vtrend(n \Delta t)$ = discrete time trend series;

$Vtrend(f)$ = filtered trend function, frequency representation;

$Vtrend(m \Delta f)$ = discrete time Fourier transform of $vtrend(n \Delta t)$ series;

$vvar(n \Delta t)$ = discrete time smoothed variance series;
 $w(\tau)$ = weighting function, time domain representation;
 $w(n \Delta t)$ = discrete time weighting series;
 $W(f)$ = weighting function, frequency domain representation;
 $W(m \Delta f)$ = discrete time Fourier transform of $w(n \Delta t)$ series;
 $z(n \Delta t)$ = discrete time standardized zscore series;
 $Z(m \Delta f)$ = discrete Fourier transform of z series;
 $Z_i; i = 1, N$ = independent standard Gaussian random variable;
 $zbar$ = mean of the zscore series;
 $zscore(n \Delta t)$ = discrete time standardized normal score series;
 zsd = standard deviation of the zscore series;
 $zscosim(n \Delta t)$ = discrete time restandardized simulated Gaussian series;
 $zsim(n \Delta t)$ = discrete time simulated signal series;
 $Zsim(m \Delta f)$ = discrete time Fourier transform of $zsim(n \Delta t)$ series;
 $\hat{F}()$ = empirical distribution function;
 $\hat{F}^{-1}()$ = inverse empirical distribution function;
 $\mathcal{F}()$ = Fourier transform notation;
 $\Phi()$ = standard normal distribution function;
 $\Phi^{-1}()$ = inverse normal distribution function;
 Δf = frequency step;
 Δt = time step;

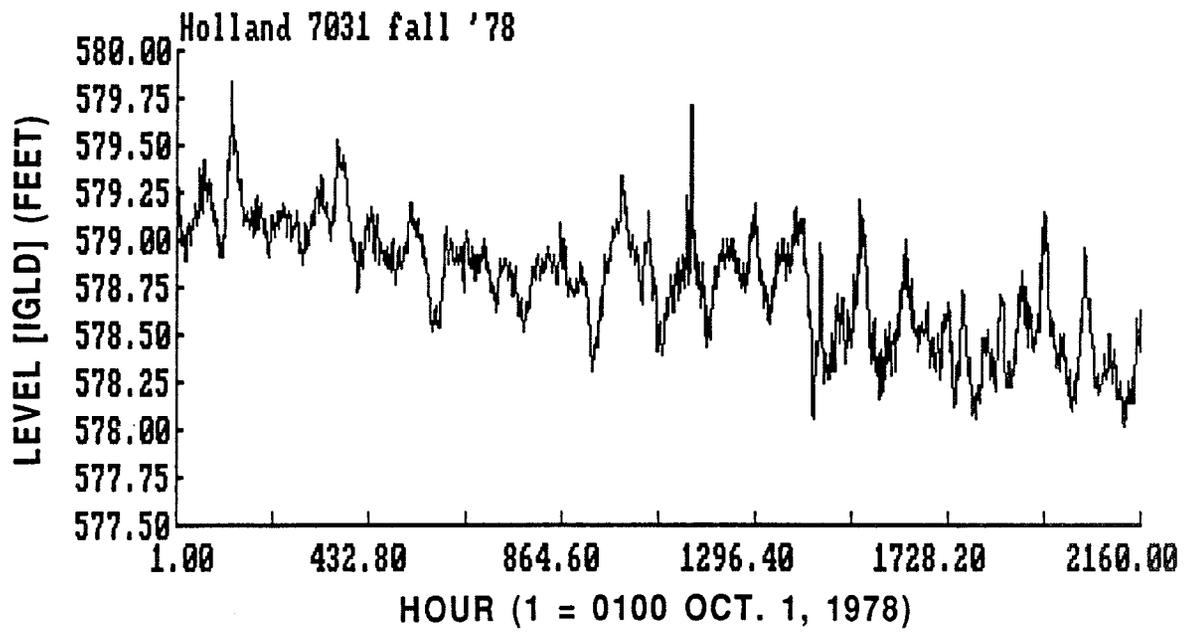


Figure 1. Hourly Water Level, Oct. 1 - Dec. 31, 1978, Holland, MI.

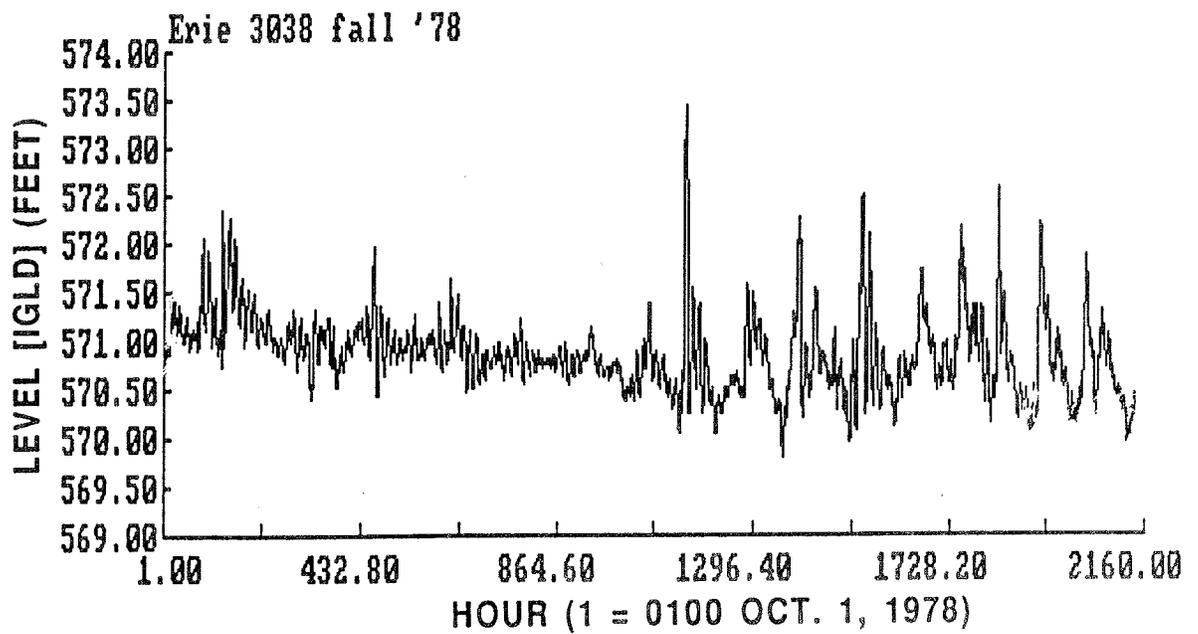


Figure 2. Hourly Water Level, Oct. 1 - Dec. 31, 1978, Erie, PA.

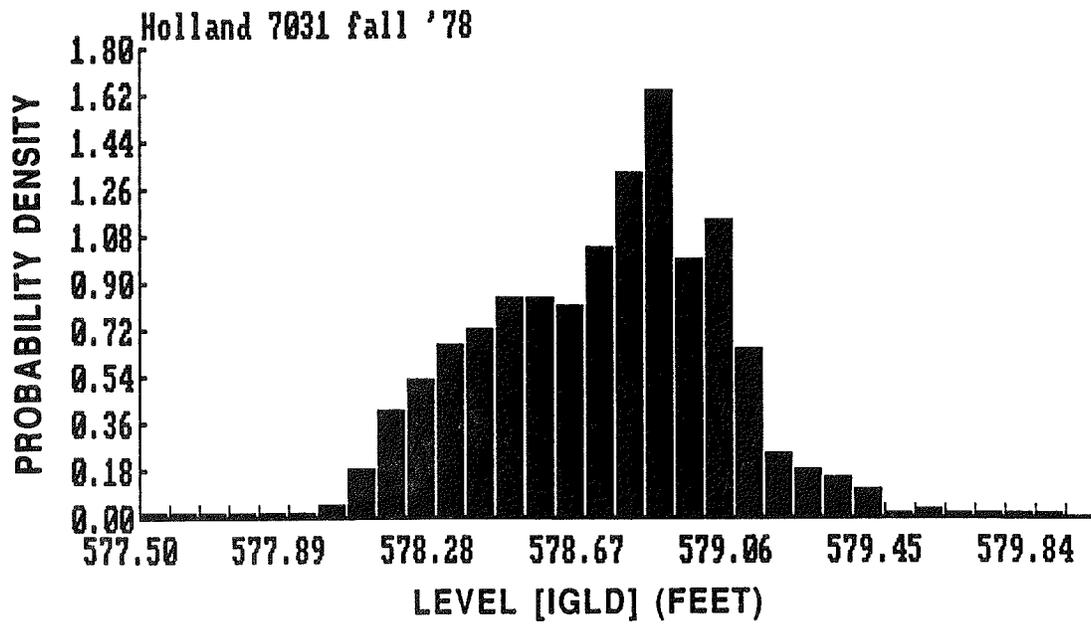


Figure 3. Histogram of Hourly Water Levels, Oct. - Dec. 1978, Holland, MI.

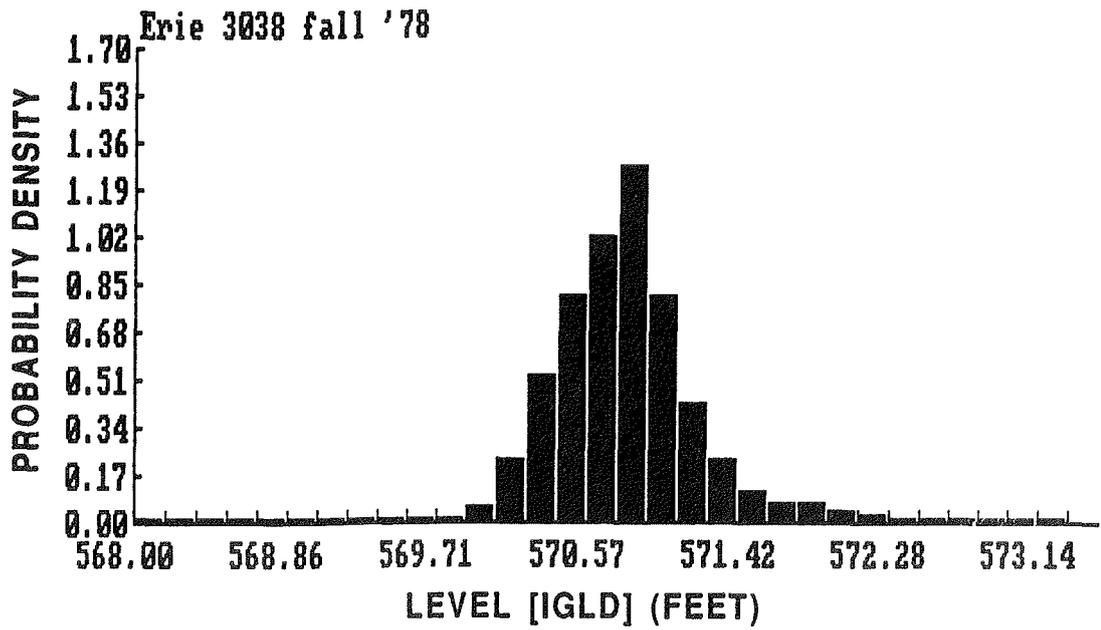


Figure 4. Histogram of Hourly Water Levels, Oct. - Dec. 1978, Erie, PA.

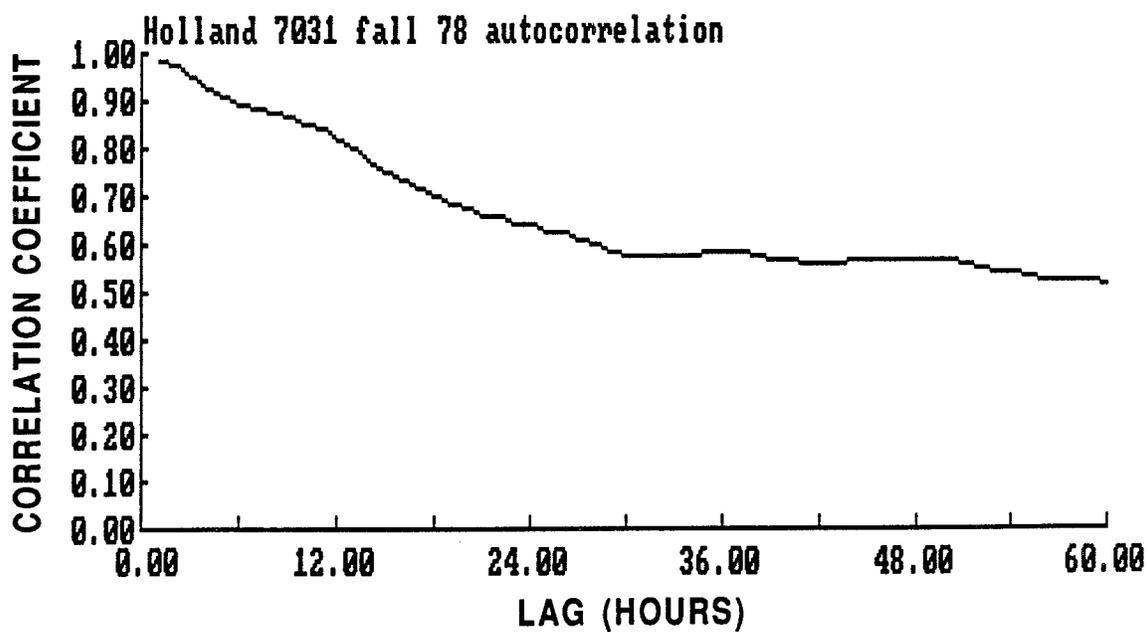


Figure 5. Autocorrelation, Hourly Water Levels, Oct. - Dec. 1978, Holland, MI.

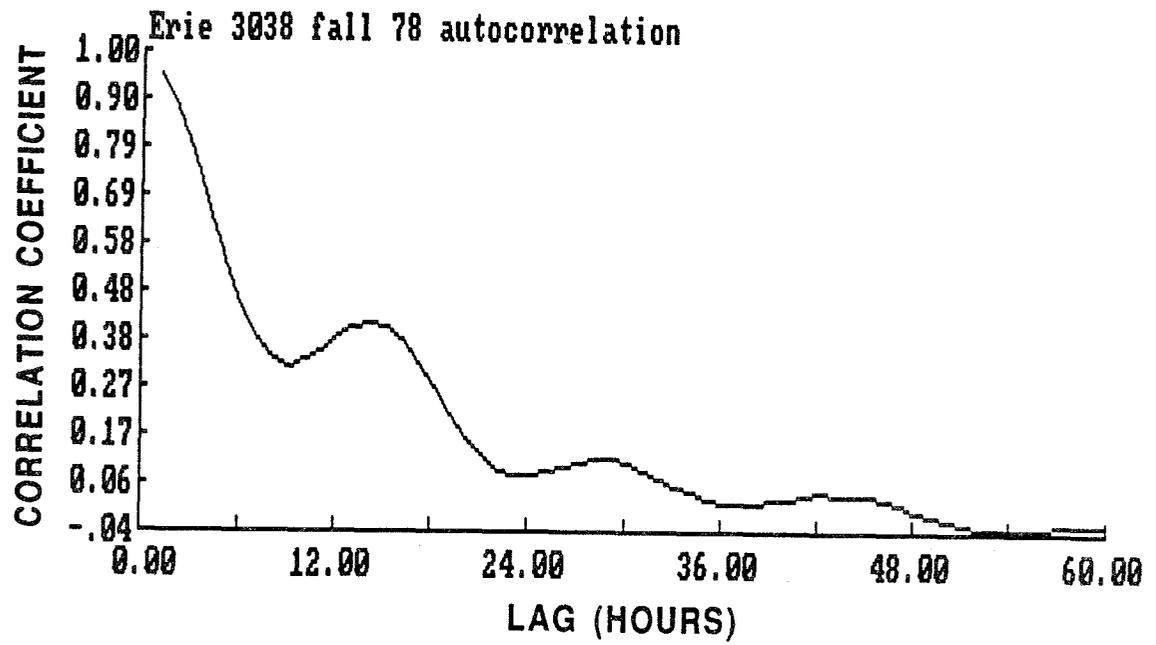


Figure 6. Autocorrelation, Hourly Water Levels, Oct. - Dec. 1978, Erie, PA.

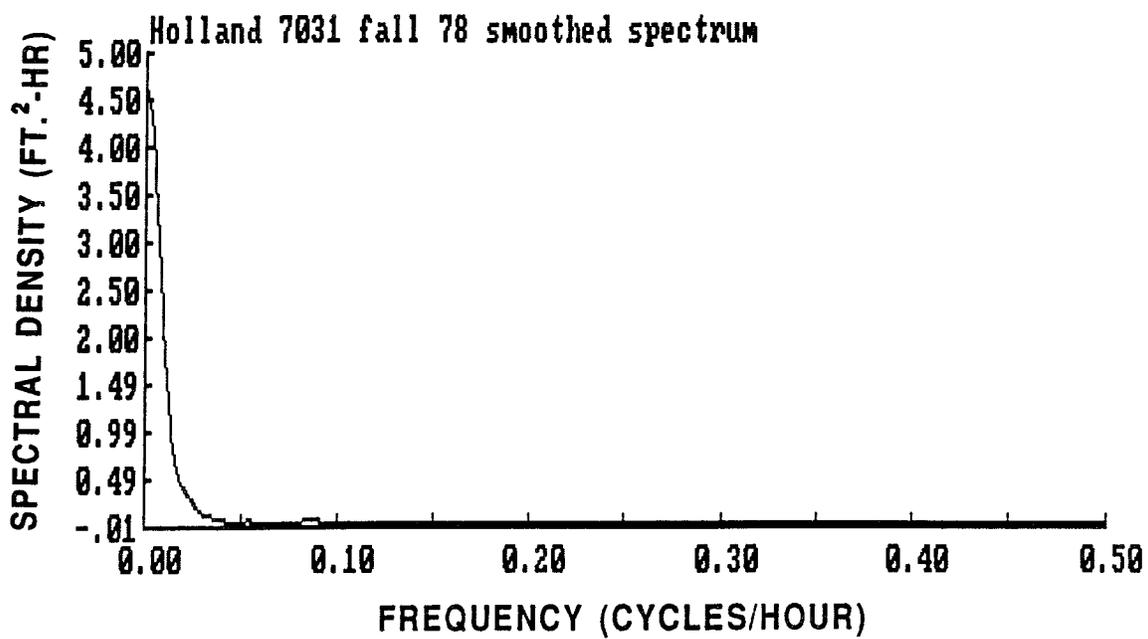


Figure 7. Smoothed Spectra, Hourly Water Levels, Oct. - Dec. 1978, Holland, MI.

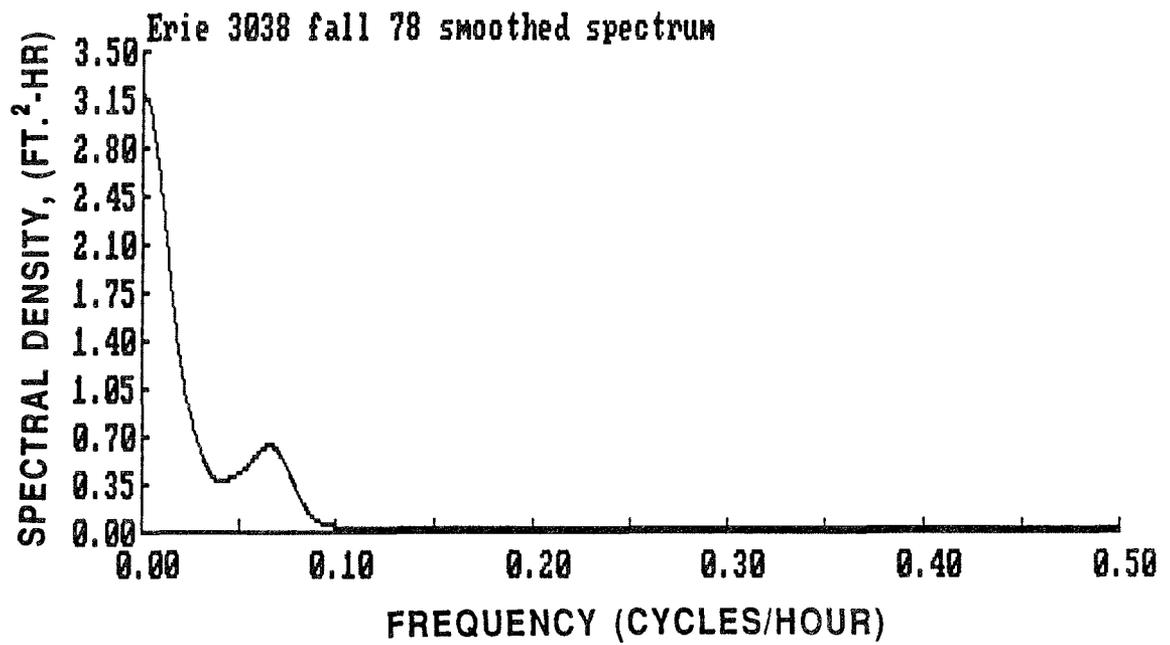


Figure 8. Smoothed Spectra, Hourly Water Levels, Oct. - Dec. 1978, Erie, PA.

Original Data
(Nonstationary, nongaussian)



Stationarity Procedures
(calculates moving average trend,
variance; converts to stationary data)



Normalization Procedure
(transformation to gaussian data)

Analysis



Standardization Procedure
(converts to zero mean, unit variance)



Spectrum Calculation
(target simulation "colored" spectrum)

Simulation Spectrum Generation
(simulated "colored" spectrum
of normal-score time series)



Destandardization Procedure
(reintroduce mean and standard deviation
from standardization procedure)



Distribution Function Transformation
(convert simulated gaussian data to data
with correct probability distribution)

Synthesis



Nonstationarity Transformation
(reintroduce trend and variance
from stationarity procedure)



Simulated Data
(nonstationary, nongaussian)

Figure 9. Block Diagram of Simulation Procedure.

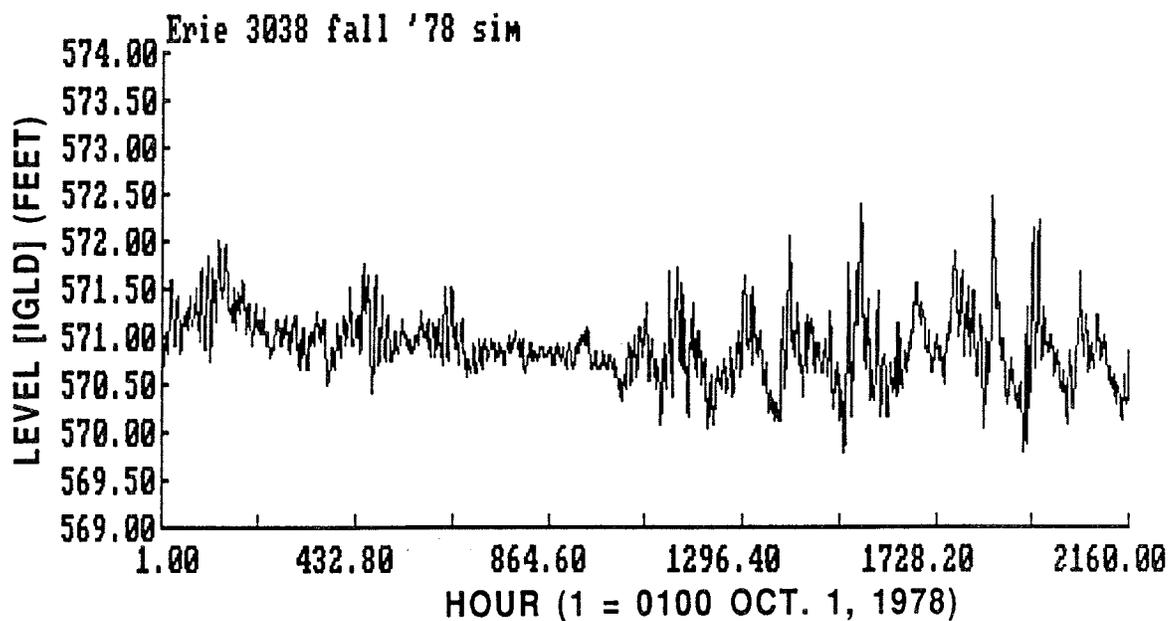


Figure 10. Simulated Hourly Water Levels, Oct. - Dec., Erie, PA.

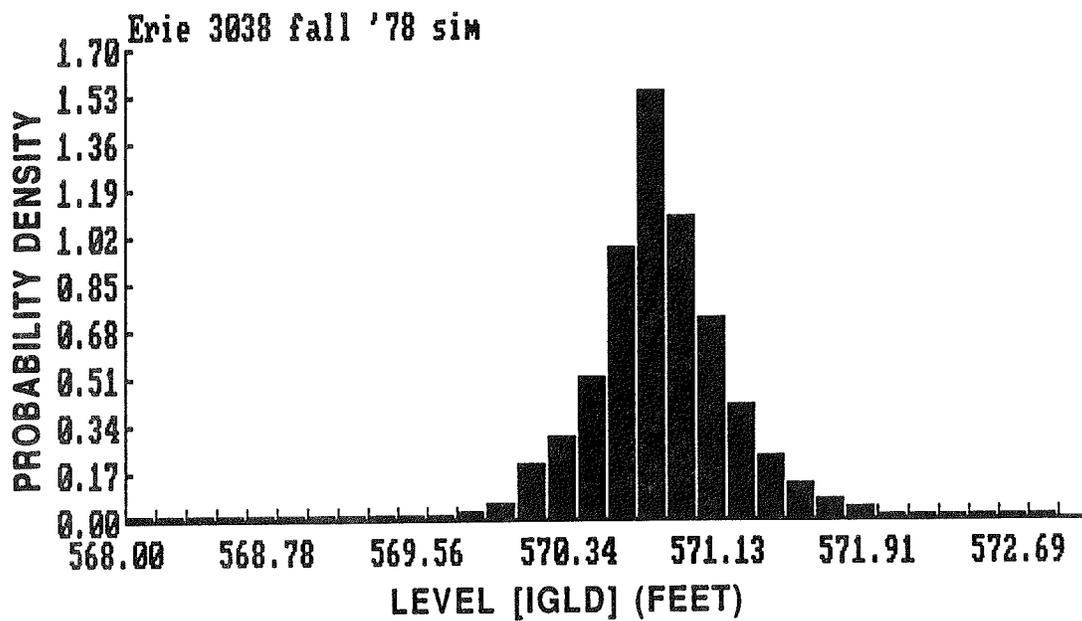


Figure 11. Histogram of Simulated Hourly Water Levels, Oct. - Dec., Erie, PA.

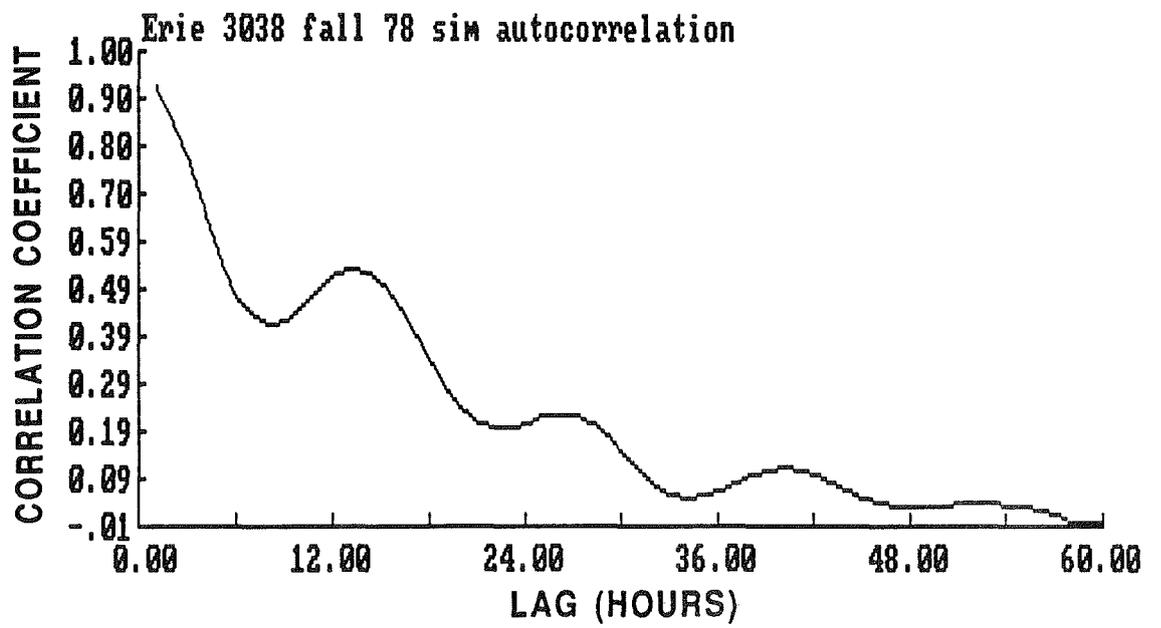


Figure 12. Autocorrelation, Simulated Hourly Water Levels,
Oct. - Dec., Erie, PA.

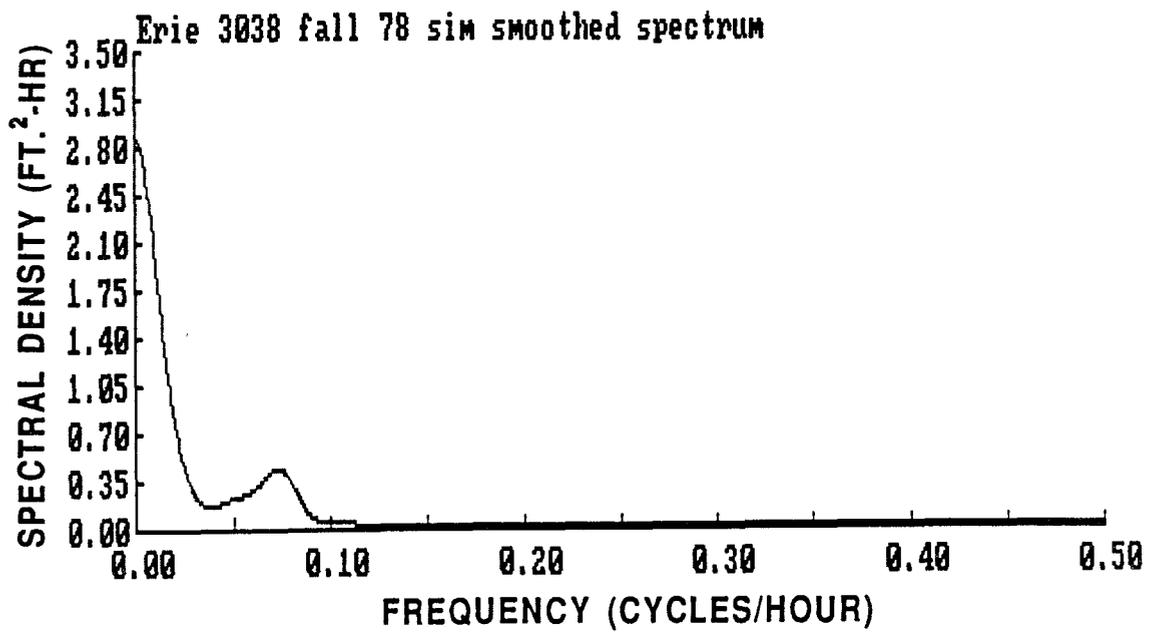


Figure 13. Smoothed Spectra, Simulated Hourly Water Levels,
Oct. - Dec., Erie, PA.

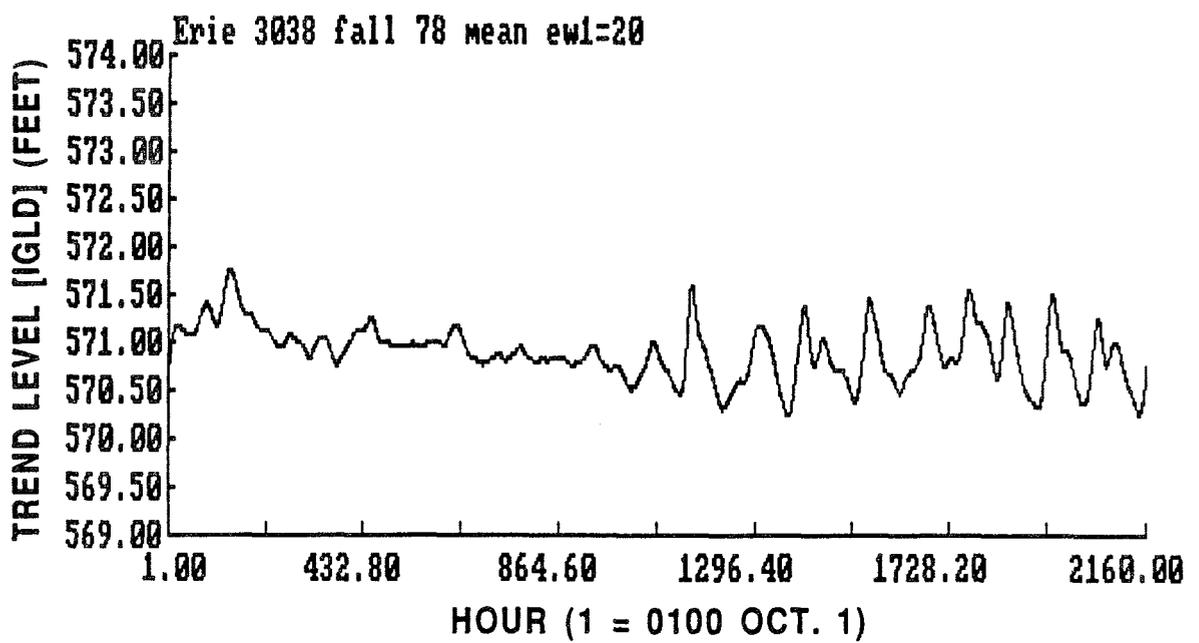


Figure 14. Trend, Hourly Water Levels, Oct. - Dec. 1978, Erie, PA.

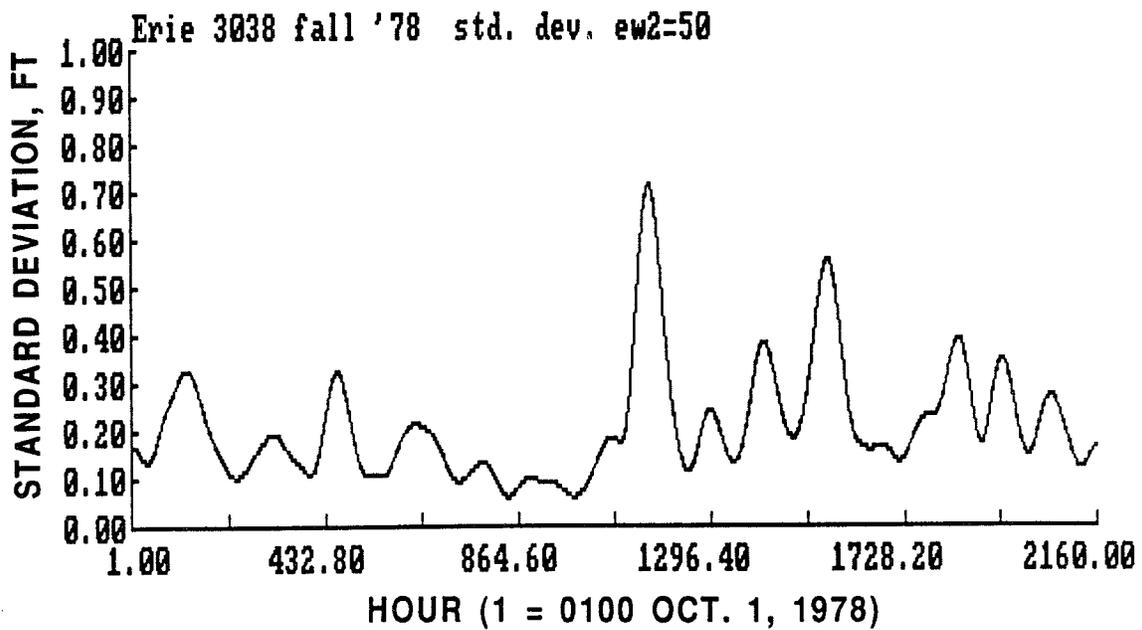


Figure 15. Local Standard Deviation, Hourly Water Levels, Oct. - Dec. 1978, Erie, PA.

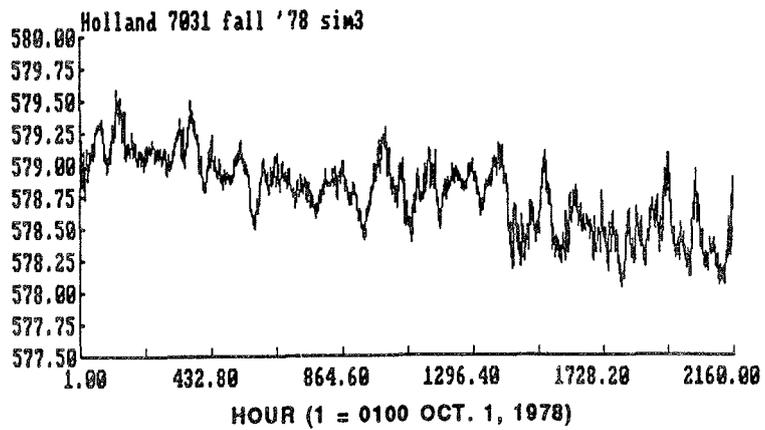
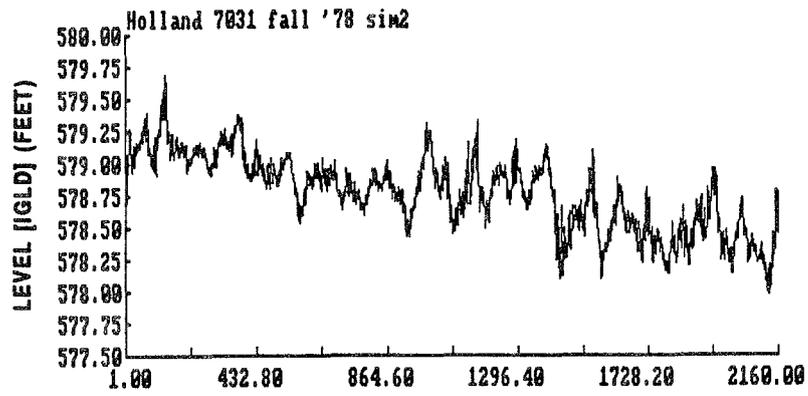
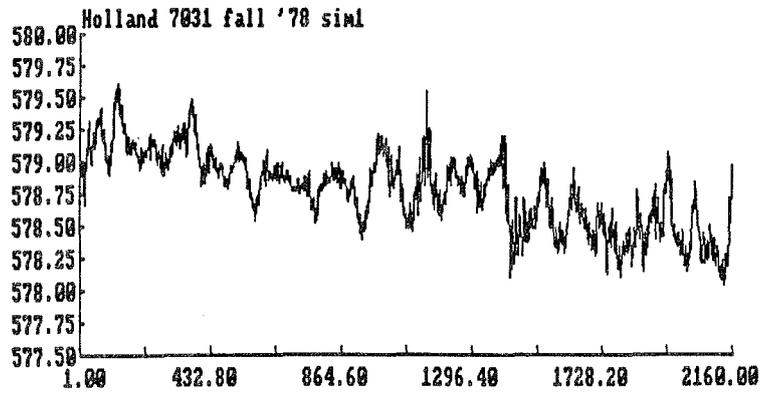


Figure 16. Simulated Hourly Water Levels, Oct. - Dec., (3 Runs), Holland, MI.

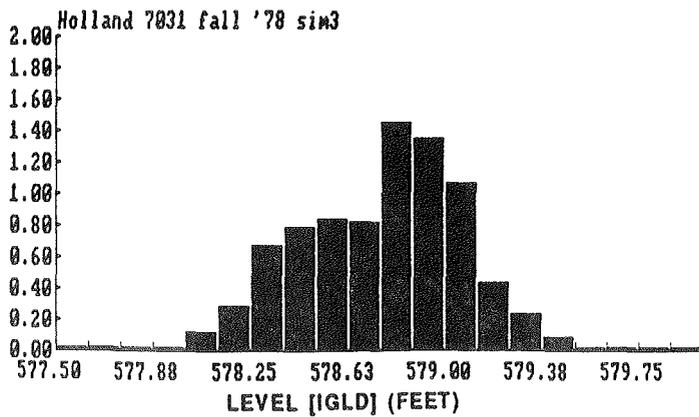
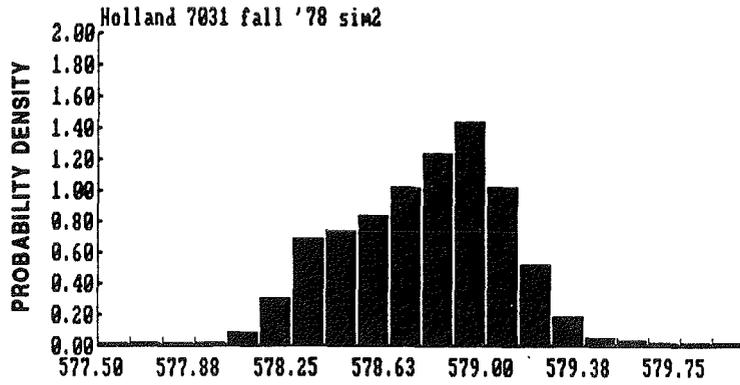
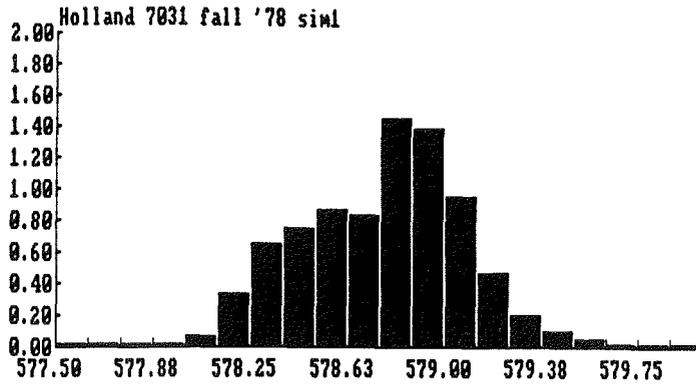


Figure 17. Histograms, Simulated Hourly Water Levels, Oct. - Dec., (3 Runs), Holland, MI.

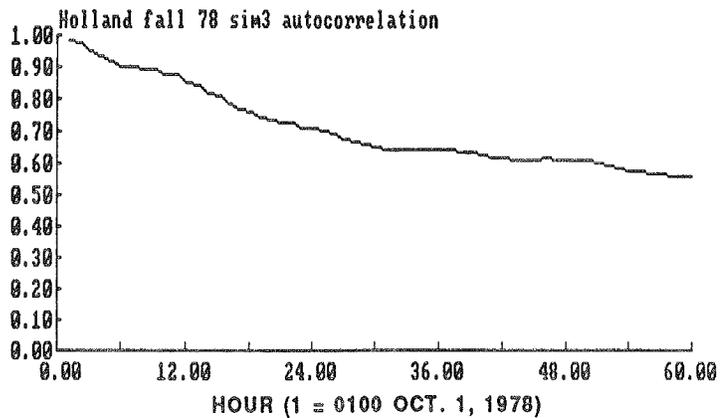
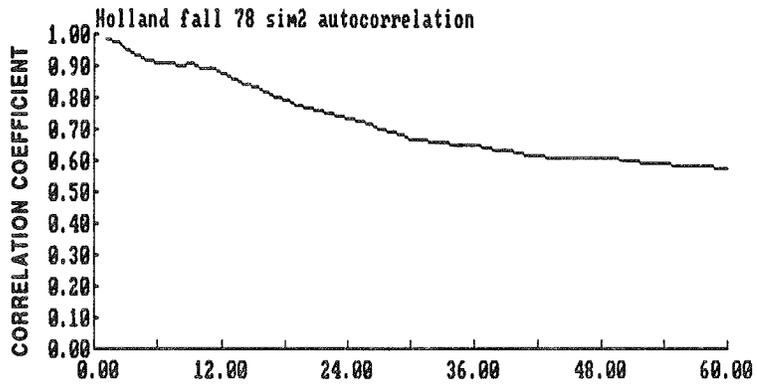
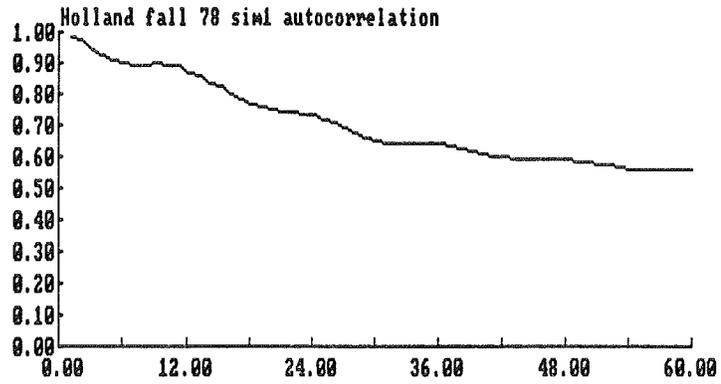


Figure 18. Autocorrelation Plots, Simulated Hourly Water Levels, Oct. - Dec., (3 Runs), Holland, MI.

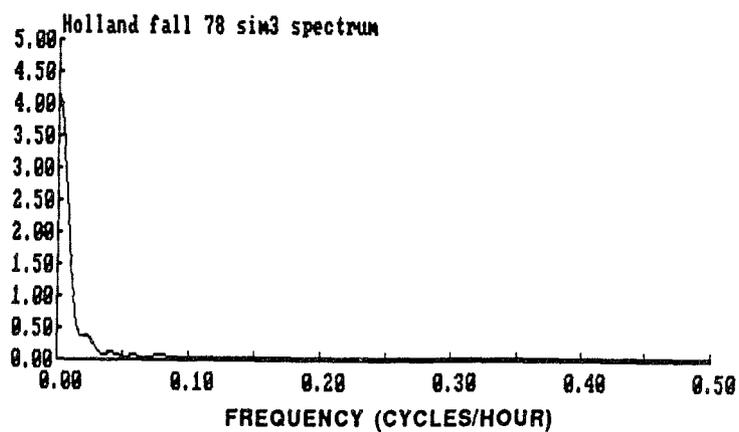
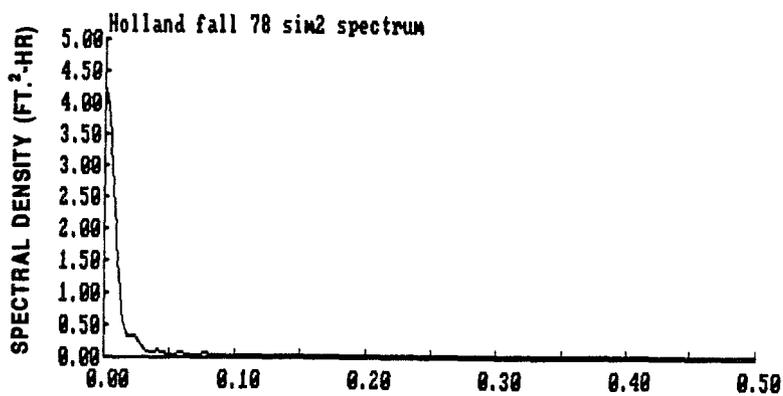
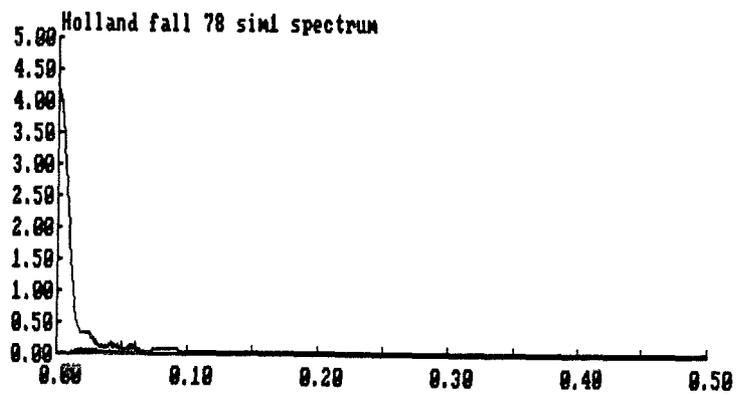


Figure 19. Smoothed Spectra, Simulated Hourly Water Levels, Oct. - Dec., (3 Runs), Holland, MI.

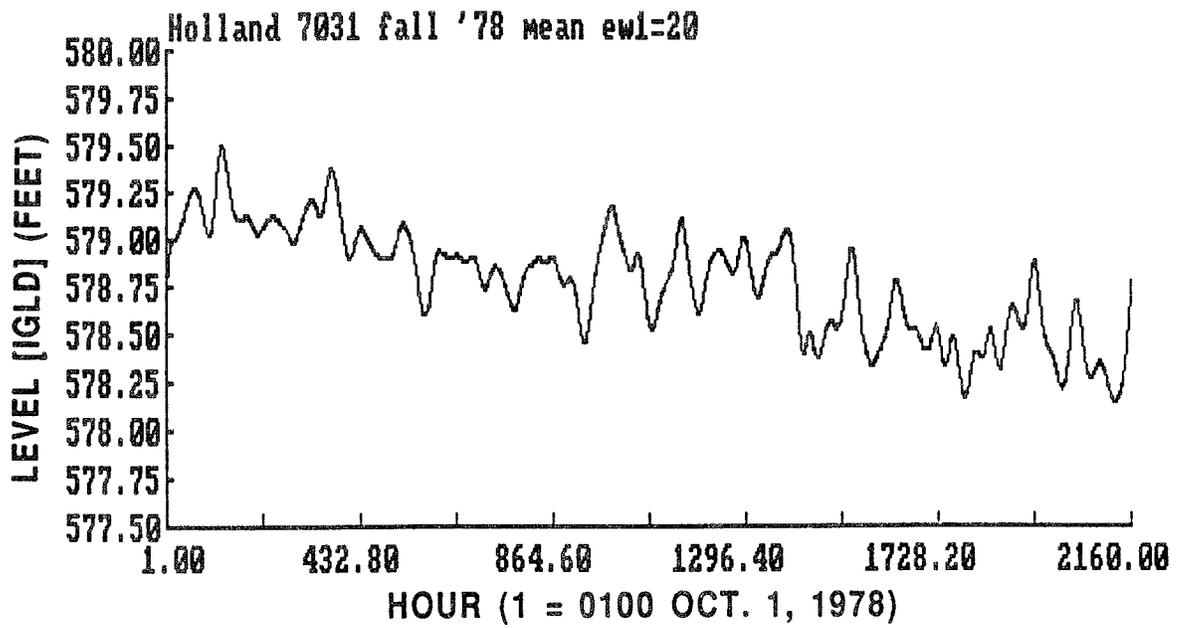


Figure 20. Trend, Hourly Water Levels, Oct. - Dec. 1978, Holland, MI.

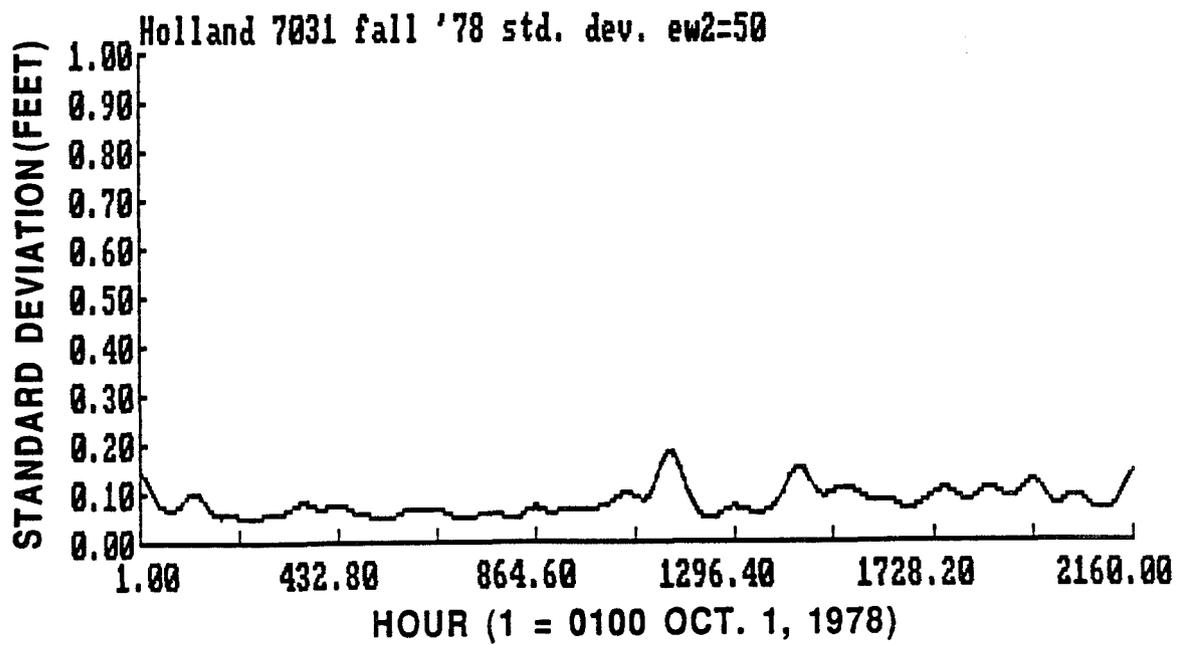


Figure 21. Local Standard Deviation, Hourly Water Levels, Oct. - Dec. 1978, Holland, MI.