

AD-A141 685

EXTREME VALUE ESTIMATES FOR ARBITRARY BANDWIDTH
GAUSSIAN PROCESSES USING T..(U) DAVID W TAYLOR NAVAL
SHIP RESEARCH AND DEVELOPMENT CENTER BET.. R D PIERCE
MAR 84 DTNSRDC/CID-84/1

1/1

UNCLASSIFIED

F/G 12/1

NL

END
DATE
FILED
7 84
DTIC

12

DTNSRDC/CID-84/1

EXTREME VALUE ESTIMATES FOR ARBITRARY BANDWIDTH GAUSSIAN PROCESSES
USING THE ANALYTIC ENVELOPE

DAVID W. TAYLOR NAVAL SHIP RESEARCH AND DEVELOPMENT CENTER



Bethesda, Maryland 20084

AD-A141 685

EXTREME VALUE ESTIMATES FOR ARBITRARY BANDWIDTH
GAUSSIAN PROCESSES USING THE ANALYTIC ENVELOPE

by
Robert D. Pierce

APPROVED FOR PUBLIC RELEASE: DISTRIBUTION UNLIMITED

CENTRAL INSTRUMENTATION DEPARTMENT
DEPARTMENTAL REPORT

DTIC FILE COPY

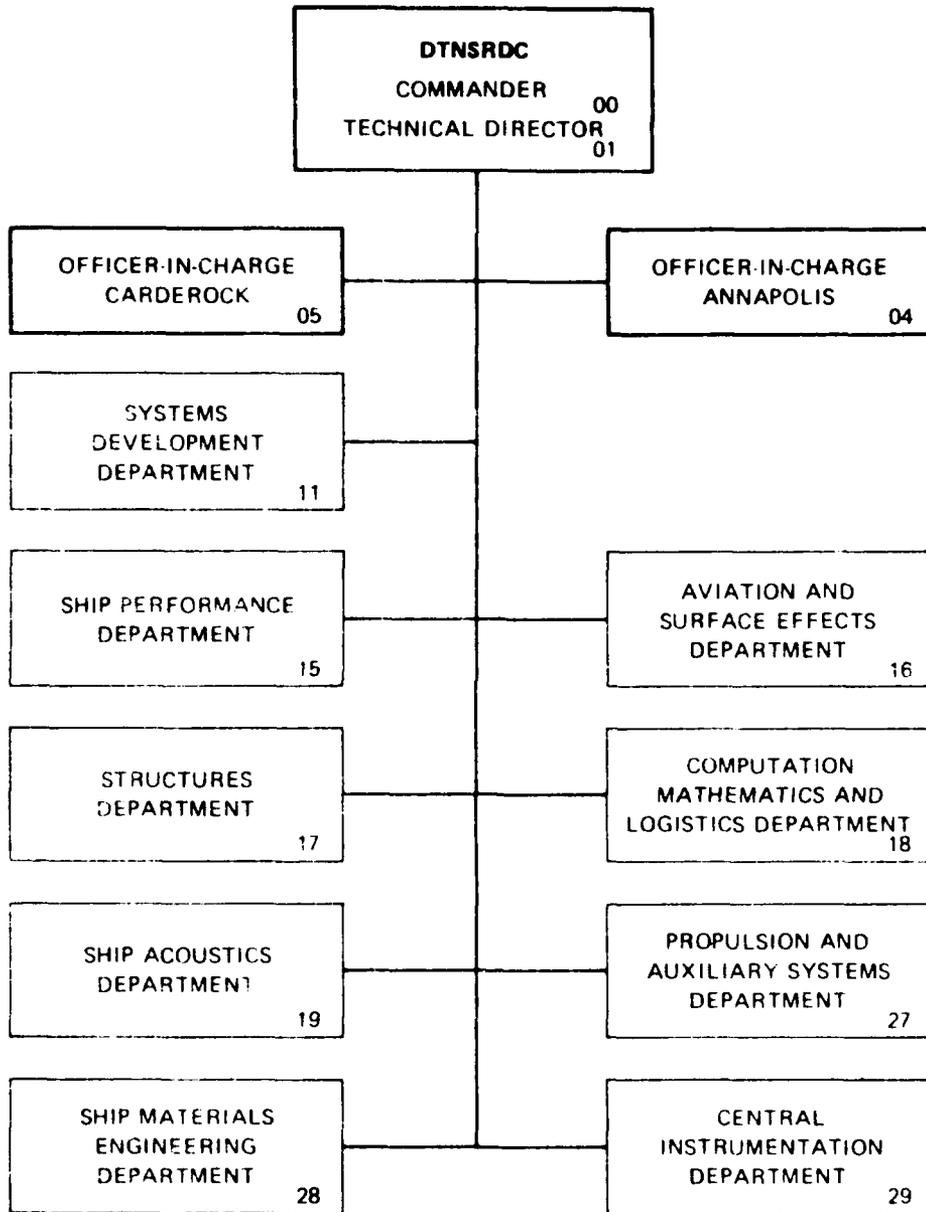
DTIC
ELECTE
S JUN 1 1984 D
A

March 1984

DTNSRDC/CID-84/1

84 06 01 006

MAJOR DTNSRDC ORGANIZATIONAL COMPONENTS



UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

(Block 20 continued)

on a digital computer to produce samples of the envelope's time history. Next, the degree of correlation between these envelope samples is taken into account using a method developed from simulation studies of a series of synthetic Gaussian time histories with varying bandwidths. Once this correlation effect has been estimated, the standard methods of order statistics are applied to these samples using the Rayleigh probability density function. Examples of applying this procedure to experimentally derived data are presented.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

TABLE OF CONTENTS

	Page
LIST OF FIGURES	iii
LIST OF TABLES	v
LIST OF ABBREVIATIONS	vi
ABSTRACT	1
ADMINISTRATIVE INFORMATION	1
INTRODUCTION	1
BACKGROUND	3
THE ANALYTIC ENVELOPE AND ITS DISTRIBUTION	5
EXTREME VALUES OF THE ENVELOPE PROCESS AND ITS EFFECTIVE SAMPLE RATE	9
EFFECTIVE NUMBER OF INDEPENDENT SAMPLES FOR THE ENVELOPES FROM THE SIMULATED GAUSSIAN TIME HISTORIES	14
APPLICATION TO EXPERIMENTAL DATA	22
CONCLUSIONS	29
APPENDIX A - SIMULATED GAUSSIAN TIME HISTORIES AND THEIR ENVELOPES	31
APPENDIX B - RELATIONSHIP BETWEEN EQUIVALENT STATISTICAL BANDWIDTH AND THE AUTOCORRELATION FUNCTION	43
APPENDIX C - MEASURED WAVE HEIGHT DATA	47
REFERENCES	63

LIST OF FIGURES

1 - Example of an Envelope Time History	7
2 - Example for Finding the Effective Number of Statistically Independent Samples	13
3 - Mean Extreme Values for Simulated Gaussian Time Histories	16

	Page
4 - Effective Sample Ratio for Simulated Time Histories	19
5 - Chi-Square Test Statistic for Three of the Simulated Gaussian Time Histories	21
6 - Mean Extreme Values for Wave Height Data	24
7 - Effective Sample Ratio for Wave Height Data	27
8 - Chi-Square Test Statistic for Wave Height Data	28
A.1 - Time History Segments of Simulated Data and their Envelopes	35
A.2 - Autospectra for Simulated Time Histories and their Envelopes	38
A.3 - Autocorrelation Functions for Envelopes of Simulated Time Histories	39
A.4 - Histograms for Run 2202	40
A.5 - Histograms for Run 2203	41
A.6 - Histograms for Run 2206	42
C.1 - Time History Segment from Run 2410, Channel 1	51
C.2 - Autospectra of Time History and Envelope for Run 2410, Channel 1	51
C.3 - Histograms for Run 2410, Channel 1	52
C.4 - Time History Segment from Run 2410, Channel 2	53
C.5 - Autospectra of Time History and Envelope for Run 2410, Channel 2	53
C.6 - Histograms for Run 2410, Channel 2	54
C.7 - Autocorrelation Functions for Envelopes of Run 2410, Channels 1 and 2	55
C.8 - Time History Segment from Run 2721	56
C.9 - Autospectra of Time History and Envelope for Run 2721	56
C.10 - Histograms for Run 2721	57

	Page
C.11 - Time History Segment from Run 2724	58
C.12 - Autospectra of Time History and Envelope for Run 2724	58
C.13 - Histograms for Run 2724	59
C.14 - Time History Segment from Run 2731	60
C.15 - Autospectra of Time History and Envelope for Run 2731	60
C.16 - Histograms for Run 2731	61
C.17 - Autocorrelation Functions for Envelopes of Runs 2721, 2724 and 2731	62

LIST OF TABLES

1 - Hilbert Transform Weights for Non-Recursive Digital Filter	10
2 - Summary of Results from Simulated Time History at Five Hertz Sample Rate	18
3 - Average Overprediction of Envelope Maxima	18
4 - Summary of Results from Measured Wave Data at Highest Sample Rate	25
5 - Histograms of Envelope Maxima for Selected Measured Wave Data	25
6 - Average Overprediction of Envelope Maxima for Measured Wave Data	25
A.1 - Basic Statistics of the Simulated Time Histories and their Envelopes (over entire run)	34
A.2 - Chi-Square Goodness-of-Fit Test for Simulated Time Histories and their Envelopes	34
C.1 - Basic Statistics of the Measured Wave Data and their Envelopes (over entire run)	50
C.2 - Chi-Square Goodness-of-Fit Test for Measured Wave Data and their Envelopes	50

LIST OF ABBREVIATIONS

B_{st}	Equivalent Statistical Bandwidth
$C(kT)$	Sampled Autocorrelation Function Normalized by Variance
DTNSRDC	David Taylor Naval Ship Research and Development Center
$E[\cdot]$	Expected Value Operator
$g(u_n)$	Probability Density Function for Largest Value in Ordered Sample u_n
$G(u_n)$	Cumulative Distribution for Largest Value in Ordered Sample u_n
$G(\omega)$	Autospectrum
$h(k)$	Filter Weights
$H(\omega)$	Transfer Function
i	Square Root of -1
M	Number of Segments
N	Number of Samples
N_a	Number of Samples Obtained from Measured Data
N_e	Effective Number of Statistically Independent Samples
$p(u)$	Probability Density Function
$P(u)$	Cumulative Distribution
$r(t)$	Envelope
$R(kT)$	Sampled Autocorrelation Function
$s(t)$	Time History
$\hat{s}(t)$	Hilbert Transform of Time History
$\text{sgn}(\omega)$	Signum Function
t	Time
$t_{n,\alpha}$	Student t distribution
T	Sample Period (reciprocal of sample rate)

u Normalized Amplitude, $u = r/\sigma$
 u_n n^{th} Value in Ordered Sample; Extreme Value
 \bar{u}_n Expected or Mean Extreme Value
 W_m Confidence Interval Width for Mean Value Estimate
 W_n Confidence Interval Width for Number of Statistically Independent Samples
 $x(nT)$ Sampled Time History
 $y(nT)$ Sampled Time History
 $\theta(t)$ Phase Angle
 σ Standard Deviation
 σ^2 Variance
 ω Frequency in Radians per Second
 ω_0 One Half the Sample Rate Frequency in Radians per Second



Accession For	
NTIS GRA&I	<input checked="" type="checkbox"/>
DTIC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
Distribution/	
Availability Codes	
Dist. Statement	
AI	

ABSTRACT

A procedure is presented for the estimation of extreme values of stationary Gaussian random processes with arbitrary bandwidths. This approach is based on the analytic envelope defined by the Hilbert Transform; this envelope is Rayleigh distributed regardless of bandwidth. For experimentally derived data that has been converted into digital form, the Hilbert Transform is approximated using algorithms implemented on a digital computer to produce samples of the envelope's time history. Next, the degree of correlation between these envelope samples is taken into account using a method developed from simulation studies of a series of synthetic Gaussian time histories with varying bandwidths. Once this correlation effect has been estimated, the standard methods of order statistics are applied to these samples using the Rayleigh probability density function. Examples of applying this procedure to experimentally derived data are presented.

ADMINISTRATIVE INFORMATION

This work was sponsored by the Mathematical Statistics program under Task SR0140501 and Element 61153N (Job order 1-1561-108) and administered by NAVSEA Code 05R24.

INTRODUCTION

At the David Taylor Naval Ship Research and Development Center (DTNSRDC), random ship motion or wave height data from full scale trials or model basin experiments is routinely subjected to various forms of analysis. Spectral analysis is performed to obtain autospectra, cross spectra and response amplitude operators; and time domain analysis is performed to obtain means, standard deviations, data point maxima and minima, data point histograms, peak histograms and peak-to-trough histograms. These different types of analysis reduce the time histories into statistical parameters that characterize the test conditions and the performance of the test vehicle. In particular, the time history statistics that describe the peaks or extremes of these random

data are used to predict the largest value (extreme value) that is expected to occur in a given period of time. The length of time available for an experiment is limited, so these results must be extrapolated to longer time intervals. The predicted extremes in motion or wave induced force are then applied to the design of ships and other marine structures.

Despite the number of studies undertaken in this area, two major problems remain. A consistent, easy to apply method is required for specifying peak occurrences for random processes with arbitrary bandwidths. And, second, a general method for accounting for statistical dependence between these events is necessary when making extreme value estimates. This report presents an approach for solving these problems under the assumption that the random process is Gaussian.

The proposed solution to these problems is based on applying a procedure that produces the time history of the envelope of the actual test data; this new time history is then analysed to produce the information necessary for extreme value predictions. The envelope time history is generated by approximating the Hilbert Transform with a digital filter. For a Gaussian random process, the statistics of this envelope time history are known; the envelope random process is Rayleigh distributed for wideband as well as narrowbanded time histories. The envelope time history is then analysed to obtain all standard statistics including a data point histogram (For narrowbanded time histories, this histogram is equivalent to the peak-to-trough histogram.) A method developed for estimating or measuring the time between statistically independent samples of the envelope can then be applied. Once the number of statistically independent samples is known for a given time, extreme value predictions are then made.

For narrowbanded random processes, each maximum and minimum does touch the envelope; however, for random processes with wide bandwidths, these maxima and minima will generally be lower than the envelope. In this sense, the envelope approach can overpredict the extreme value. In the most widebanded simulation, the average overprediction was 6.6%. This drawback is balanced by the method's ability to determine if the measured data is non-Gaussian by performing a Chi-Square goodness-of-fit test on the histogram of the envelope time history. The envelopes of Gaussian random processes with arbitrary bandwidths are still

Rayleigh distributed. Since the results of this test are not confused by the bandwidth of the process, the characteristics of non-Gaussian processes are more easily studied.

The method is implemented using digital computer algorithms; simulated, Gaussian distributed time histories with varying bandwidths are used to develop the procedures. This approach is verified by using Chi-Square goodness-of-fit tests to demonstrate that the envelope time history is Rayleigh distributed and that the extreme values follow the distribution predicted by order statistics. This method is then applied to experimentally obtained time histories.

BACKGROUND

The methods described in the literature for making extreme value predictions are based on one or more of the following assumptions about the random time history: The random process is stationary, zero mean and Gaussian distributed; its autospectrum is narrowbanded; and the peaks or maxima of the time history are statistically independent. The narrowbanded assumption implies that only one maximum and minimum (peak-to-trough) occur between two successive zero crossings of the random process. The method described by Longuet-Higgins^{1*} and by Cartwright and Longuet-Higgins² is the classical approach. With this approach, the peak-to-trough amplitudes are distributed according to a Rayleigh distribution for narrowbanded, Gaussian random processes; order statistics methods predict the extreme value provided these amplitudes are statistically independent of each other. This method requires only one parameter for the Rayleigh distribution, either the estimated mean square of the peak-to-trough amplitudes or the estimated variance of the time history. The number of peak-to-trough occurrences divided by the record length determines the encounter rate; this rate is then used with order statistics methods to estimate (extrapolate) the exposure time necessary for an extreme value to exceed a specified threshold with a given probability of occurrence.

To avoid the narrowbanded assumption, other approaches are described by

* A complete listing of references is given on page 63.

Cartwright and Longuet-Higgins² and by Ochi³. These methods are based on the study presented by Rice⁴. Since a non-narrowbanded random process can have many maxima or minima between successive zero crossings, these maxima or minima are shown to be distributed according to the distribution (Rice distribution) given by Rice⁴. When applying order statistics, statistical independence of these events is still assumed. This method requires the estimation of the first three non-zero spectral moments; these moments are equivalent to the variance of the random process and the variance of the first two derivatives of the random process. These moments are estimated from the time history and they are used to form a bandwidth parameter; the Rice distribution is described by this bandwidth parameter and the variance estimate. The occurrence rate of these events establishes the time base for extreme value estimates.

In practice, the peak-to-trough approach proposed by Longuet-Higgins¹ is still generally applied to experimental data. It is uncomplicated; this method only requires one parameter, the variance, to specify the distribution. The average of the highest one-third peak-to-trough amplitudes is given by four times the standard deviation. Measurement noise in the test data will add to the variance estimate, and computation of the peak-to-trough histogram may require a deadband zone such that only the "significant" zero crossing cycles are used as events; however, this approach is fairly tolerant of noise. In comparison, the method using the Rice distribution also requires the variance as well as the bandwidth parameter. This method is more sensitive to measurement noise in the test data; noise will change the shape of the histogram as well as inflate the estimated values of the second and fourth spectral moments that are then used to compute the bandwidth parameter. Estimation of these spectral moments involves single and double differentiation of the time history (These calculations are usually performed in the frequency domain.); the level of the noise is increased by the differentiation process.

The peak-to-trough approach is also applied to cases where the experimental data is non-Gaussian or non-narrowbanded. For Gaussian random processes with moderate bandwidths, Tayfun⁵ proposed a peak-to-trough probability density function that is based on sampling the Rayleigh distributed envelope. In some cases, empirical distributions such as those given by Ochi⁶ or Forristall⁷ are fitted to the measured peak-to-trough histograms. In other

cases, however, wideband data is simply assumed to be narrowbanded. Between two "significant" zero crossings, the difference between the largest and the smallest data values is used to form a "peak-to-trough" amplitude. Attempts are made to treat the wideband nature of the data as if it were noise. Deadband regions are set to determine if a zero crossing is "significant". Defining what constitutes a "significant" zero crossing (or alternately a "significant" peak to trough occurrence) is a major drawback with this approach; the required criteria cannot be specified in a consistent manner.

These previous methods do not offer any procedures for taking into account the correlation between successive events when making extreme value predictions. Epstein⁸, however, has proposed such a procedure using a Markov chain condition. The Markov chain condition requires a certain kind of dependence between successive events. Recently Ochi⁹ and Naess¹⁰ suggest that this approach may account for the effects of this correlation for random processes with moderate bandwidths; however, they do not demonstrate if the Markov chain condition is appropriate for arbitrary bandwidth Gaussian random processes. Naess¹⁰ disagrees with Ochi's⁹ results. Also, Epstein⁸ states that a clean-cut analytic solution is expected for *only certain kinds of stochastic dependence*. This approach appears to be subject to continuing development.

A procedure for estimating extreme values using the envelope associated with the time history is proposed by Naess¹¹; a method for taking the effects of correlation into account is also presented. The average rate at which the envelope crosses the standard deviation level from below is used as part of the criteria for determining the effect of dependent samples. This procedure is shown to be limited to narrowbanded Gaussian processes.

THE ANALYTIC ENVELOPE AND ITS DISTRIBUTION

Regardless of the bandwidth of a Gaussian distributed random process, the envelope of this process is Rayleigh distributed. (See Middleton¹², pg 351 and pg 396; also, see Arens¹³ and Dugundji¹⁴.) This envelope is defined by using the Hilbert transform. In a recent paper by Rice¹⁵, this envelope is called the analytic envelope.

The envelope, $r(t)$, is defined by

$$r(t) = [s(t)^2 + \hat{s}(t)^2]^{1/2} \quad (1)$$

and, the phase angle, θ , is defined by

$$\theta(t) = \arctan(\hat{s}(t)/s(t)) \quad (2)$$

such that

$$s(t) = r(t) \cos \theta(t) \quad (3)$$

where $s(t)$ is the input time history (either random or deterministic), and

$$\hat{s}(t) = r(t) \sin \theta(t) \quad (4)$$

where $\hat{s}(t)$ is the Hilbert Transform of $s(t)$. An example of the envelope of a random time history with moderately wide bandwidth is given in Figure 1. For a real-valued function, $s(t)$, on $-\infty < t < \infty$, the Hilbert transform, $\hat{s}(t)$, is defined by

$$\hat{s}(t) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{s(\tau)}{t - \tau} d\tau \quad (5)$$

where the principal value of the integral is used. The Hilbert transform shifts all the frequency components in the input time history by 90 degrees. In the frequency domain, the Hilbert transform is equivalent to a filter with the following transfer function, $H(\omega)$:

$$H(\omega) = -i \operatorname{sgn}(\omega) \quad (6)$$

where i is the square root of minus one and $\operatorname{sgn}(\omega)$ is the signum function [$\operatorname{sgn}(\omega) = 1$ for $\omega > 0$, 0 for $\omega = 0$, and -1 for $\omega < 0$].

For a zero mean Gaussian distributed random process, several relationships exist. $s(t_0)$ and $\hat{s}(t_0)$ are statistically independent at any arbitrarily selected time, t_0 . Since the envelope is the square root of the sum of two statistically independent random variables at each instant of time, the

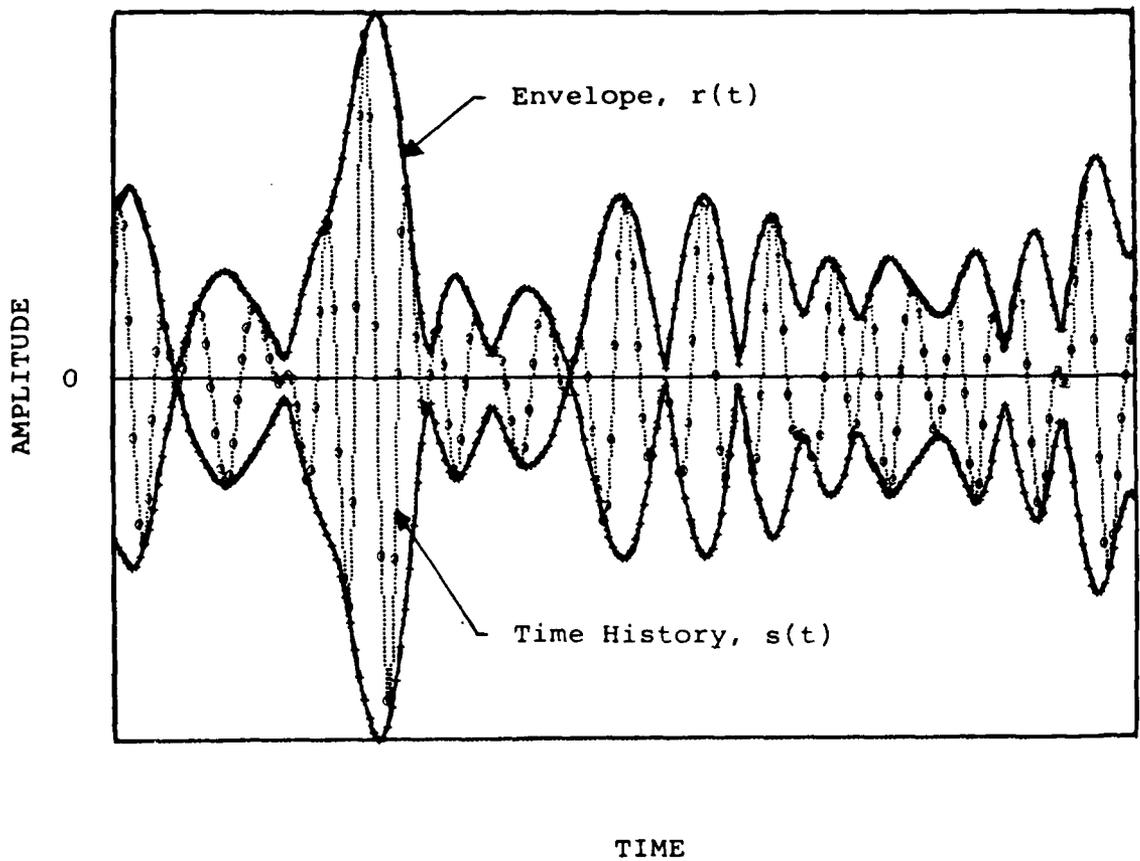


Figure 1 - Example of an Envelope Time History

envelope is Rayleigh distributed with the following probability density function

$$p(u) = u e^{-u^2/2} \quad (7)$$

and cumulative distribution,

$$P(u) = 1 - e^{-u^2/2} \quad (8)$$

where $u = r/\sigma$. σ is the standard deviation of $s(t)$. Also, $s(t)$ and $\hat{s}(t)$ have the same autospectrum and autocorrelation functions. The envelope and phase are statistically independent, and the phase angle is uniformly distributed between $+\pi$ and $-\pi$. The mean of the envelope is

$$E[u] = (\pi/2)^{1/2} \quad (9)$$

where $E[.]$ is the expected value operator. The mean square of the envelope (variance plus square of the mean) is

$$E[u^2] = 2 \quad (10)$$

Since the envelope is Rayleigh distributed, the average of the highest one-third envelope samples is two times the standard deviation of $s(t)$. These relationships hold regardless of the bandwidth of the random process.

Except for a narrowband random process, the peaks of the process do not necessarily touch its associated envelope. See Figure 1 for an example of a moderately wide bandwidth process. From Equation (1), the absolute value of the random process, $|s(t)|$, is equal to the envelope whenever $\hat{s}(t)$ is zero. Since $s(t)$ and $\hat{s}(t)$ have the same autospectrum and hence the same zero crossing statistics, this equality will, on the average, occur between each zero crossing. From Equation (3), this equality will occur whenever the phase is either 0 or π ; and since phase is statistically independent of the envelope, a joint occurrence at an envelope maxima is therefore not expected. This characteristic of the envelope leads to overestimation of the extreme value of

the random process; the significance of the overprediction is discussed in the section titled **Effective Number of Independent Samples for the Envelopes from the Simulated Gaussian Time Histories.**

The Hilbert transform is equivalent to a filter. Methods for approximating this filter are described by Tayfun⁵ and Rabiner and Schafer¹⁶. Given a time history that has been digitized for analysis by a computer, an approximate Hilbert transform of this time history can be produced. Thus a time history of the envelope can be created and standard data analysis techniques (power spectra, correlation, and histograms) can be applied to this time history.

The method used to approximate the Hilbert transform is given by Rabiner and Schafer¹⁶. With this method, a non-recursive digital filter (finite impulse response, FIR) is formed, so the Hilbert transform of time histories with arbitrary lengths can be calculated using a computer. From this reference, the filter weights (impulse response) given in Table 1 for a 95 element filter were used; these filter weights, $h(k)$, are reproduced in Table 1. This filter has a peak approximation error of 2.2% over the frequency interval from 0.01 to 0.49 times the sample rate. Odd-indexed filter weights are zero and the last half of the filter weights are given by:

$$h(k) = -h(K-1-k) \quad (11)$$

for $k = 0, 1, \dots, K-1$ where K is 95. The non-recursive digital filter has the following form:

$$\hat{s}(jT) = \sum_{k=0}^{K-1} h(k) s\left[\left(\frac{K+1}{2} + j - k\right)T\right] \quad (12)$$

where $s(jT)$ is the filter's input and $\hat{s}(jT)$ the output.

EXTREME VALUES OF THE ENVELOPE PROCESS AND ITS EFFECTIVE SAMPLE RATE

The extreme value of the envelope is derived by applying order statistics

TABLE 1 - HILBERT TRANSFORM WEIGHTS FOR NON-RECURSIVE DIGITAL FILTER
 (FROM RABINER AND SCHAFER¹⁶, TABLE I, K=95)

k	h(k)
0	-.0130099
2	-.0045718
4	-.0053689
6	-.0062800
8	-.0072616
10	-.0083873
12	-.0096455
14	-.0110350
16	-.0126022
18	-.0143770
20	-.0163895
22	-.0186922
24	-.0213465
26	-.0244424
28	-.0281175
30	-.0325665
32	-.0380864
34	-.0451608
36	-.0546233
38	-.0680547
40	-.0888468
42	-.1258168
44	-.2112989
46	-.6363167

to the sampled envelope. Given that the envelope is sampled N times such that these samples are statistically independent, the extreme value is defined as the largest value of the envelope that will occur in these N samples. Once these samples are arranged in ascending order, the largest normalized envelope value in the ordered sample, u_n , has the following probability density function:

$$g(u_n) = N p(u_n) [P(u_n)]^{N-1} \quad (13)$$

Since the envelope is Rayleigh distributed, $p(u_n)$ is given by Equation (7) and $P(u_n)$ is given by Equation (8). The cumulative distribution is given by:

$$G(u_n) = \int_0^{u_n} g(v) dv = \{ P(u_n) \}^N \quad (14)$$

The expected or mean extreme value is given by

$$\overline{u_n} = E[u_n] = \int_0^{\infty} u_n g(u_n) du_n \quad (15)$$

To this point, the application of order statistics to the sampled envelope is equivalent to the treatment of order statistics to wave amplitudes by Longuet-Higgins¹.

When the envelope of the time history is approximated using algorithms implemented on a digital computer, the sample rate is the same for both the envelope and the time history. For a time history of fixed length, N_a samples are obtained. In general, these envelope samples will be highly correlated. A measure of the effective sample rate is required that will produce N_e statistically independent samples, where N_e is the effective number of statistically independent samples. Middleton¹² shows that uncorrelated envelope samples are also statistically independent, so one procedure for estimating an effective sample rate is to compute the autocorrelation function of the envelope time history and then sample the envelope at time intervals such that

this function decays below a set threshold. The use of this procedure, however, can result in underestimation of extreme events because of the reduced sample rate; sampling may not occur at the peaks of the envelope. A approach that uses all of the available samples is required.

Two approaches are proposed for estimating or measuring the effective sample rate. With either method, the available time history is first divided into M segments where each segment contains N_a actual samples. The largest envelope value in each segment is then found; from one segment to another, these extreme envelope values are assumed to be statistically independent. With the first approach, a histogram of these extreme envelope values is formed; this histogram is then matched to the hypothetical distribution given by Equation (13) by varying the number of independent samples, N . The width of each class interval is selected to give an equal number of occurrences for the expected distribution (Equation (14)); the last class interval has infinite width. Next, the Chi-Square goodness-of-fit test is applied to this histogram using the procedures described by Bendat and Piersol¹⁷. The number of independent samples that minimizes the Chi-Square statistic is the estimate of the effective number of independent samples, N_e . The effective sample rate is N_e divided by the segment length in seconds. This approach has the advantage that the measured distribution of extreme values is compared to the hypothetical distribution.

The second approach for estimating the effective number of independent samples is based on matching the measured mean extreme value to the hypothetical mean given by Equation (15). First, the mean is estimated for the extreme envelope values from the M segments. Next, the number of independent samples that give the same mean value is found by searching a look-up table produced by numerically integrating Equation (15). Graphically, this procedure is shown in Figure 2; here, each of the M segments contains N_a (actual) samples, and the effective number of independent samples is given by N_e . This approach has several advantages. First, a confidence interval can be placed around the measured mean extreme value that can then be transformed into a confidence interval about the effective number of independent samples. Second, this procedure can be repeated after sampling the envelope at different rates to determine the effects of undersampling the envelope time history.

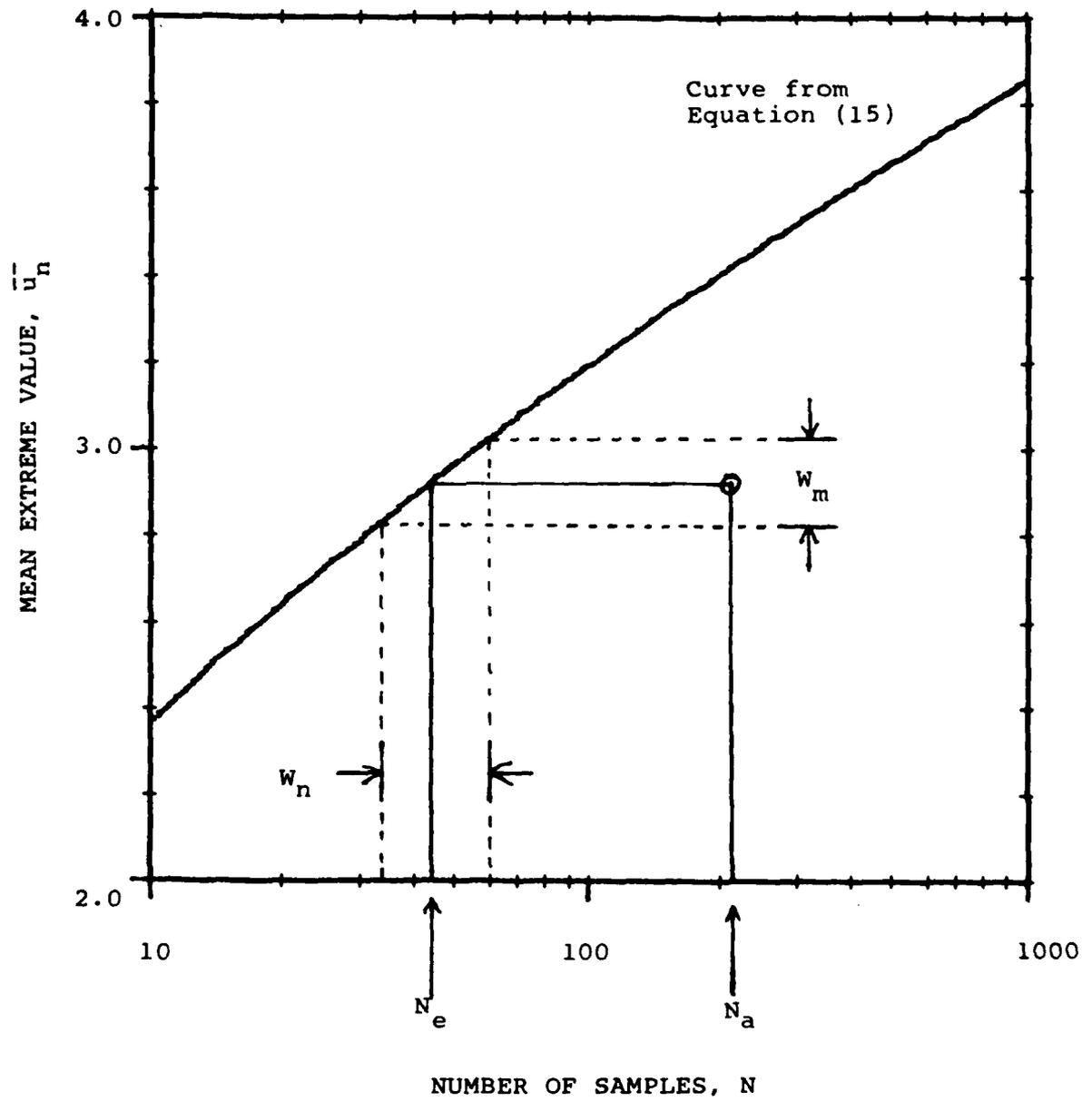


Figure 2 - An Example for Finding N_e , the Effective Number of Statistically Independent Samples

A combination of these two approaches was used; the effective number of independent samples, N_e , and its confidence interval was found from mean value estimates using the second method, and then the distribution of the extreme values was checked at this N_e value by using the first approach. The hypothetical distribution was accepted at a 5% level of significance provided a N_e value was found within the confidence interval that also passed the Chi-Square goodness-of-fit test. 90% confidence intervals were found for these mean value and N_e estimates. Since the mean of M samples (M large) from any distribution will approach a Gaussian distribution, the procedure for determining the mean value confidence interval given by Bendat and Piersol¹⁷ is used. For this application, the true mean value is known to be located within an interval of width, W_m , where:

$$W_m = 2 \frac{s t_{M-1;0.05}}{M^{1/2}} \quad (16)$$

such that W_m is centered about the estimated mean value where M is the number of segments, s is the estimated standard deviation, and $t_{n,\alpha}$ is the Student t distribution. These limits on the mean value estimate are then used to find the confidence interval for the number of statistically independent samples. The width of this interval is W_n .

EFFECTIVE NUMBER OF INDEPENDENT SAMPLES FOR THE ENVELOPES FROM THE SIMULATED GAUSSIAN TIME HISTORIES

These procedures for estimating the effective number of independent samples were applied to a series of simulated Gaussian time histories with varying bandwidths. From this study, the ratio of the effective number of samples to the actual number of samples (N_e/N_a) is shown to be related to the equivalent statistical bandwidth of the sampled envelope time history. The results from this study can then be applied to actual test data, or, if the test data is sufficiently long, these procedures for directly estimating this ratio could be applied to these test data.

The methods used to produce these simulated time histories are described in Appendix A. With filtered white noise produced using a random number generator, five different time histories were generated that had bandwidths varying from the extreme narrowband to wideband. In each case, the spectral shape was a rectangular block centered at 0.51 hertz; the bandwidths for each time history were 0.078, 0.176, 0.293, 0.488 and 0.684 hertz. These bandwidths correspond to runs 2202, 2205, 2203, 2208 and 2206. The variance of each time history was set to one volt. The duration of these time histories was 10,080 seconds at a sample rate of five hertz. The envelope time histories for these cases were obtained using the Rabiner and Schafer¹⁶ method to produce the Hilbert transform. In Appendix A, 20 second long segments of each time history and its envelope are reproduced. Histograms of each time history were obtained and the Chi-Square tests were run to test the input time history against the Gaussian distribution, and the envelope was tested for the Rayleigh distribution. In all cases, the hypothetical distribution was accepted at a 5% level of significance. Examples of the histograms are given. Also shown are plots of their autospectra as well as a plot of each envelope's autocorrelation function.

The procedures for finding the effective sample rates were applied to each of these five envelope time histories. Each time history was divided into 60 intervals ($M=60$) that were 168 seconds long. Six different sample rates were used: 5.0, 2.5, 1.0, 0.5, 0.25 and 0.125 samples per second. Each segment contained 841, 420, 168, 84, 42 or 21 samples (N_a samples) at these respective sample rates. The mean and standard deviation of the largest sample in each segment was estimated; the mean value estimates are given in Figure 3. The solid curve is the expected mean numerically calculated from Equation (15). These results show the affect of sampling. As the sample rate is reduced, the estimated mean is almost constant (same peak values found) until the curve of the expected mean is approached. Here, the estimated mean reduces in value (the affect of undersampling), and next the estimated mean follows the expected mean as the sample rate is reduced further. Since the mean is estimated from 60 samples, a 90% confidence interval is placed around these mean estimates. For these cases, this interval has a width, W_m , from 0.140 to 0.236; a nominal value for each of these cases is 0.200. The width of this interval is shown on

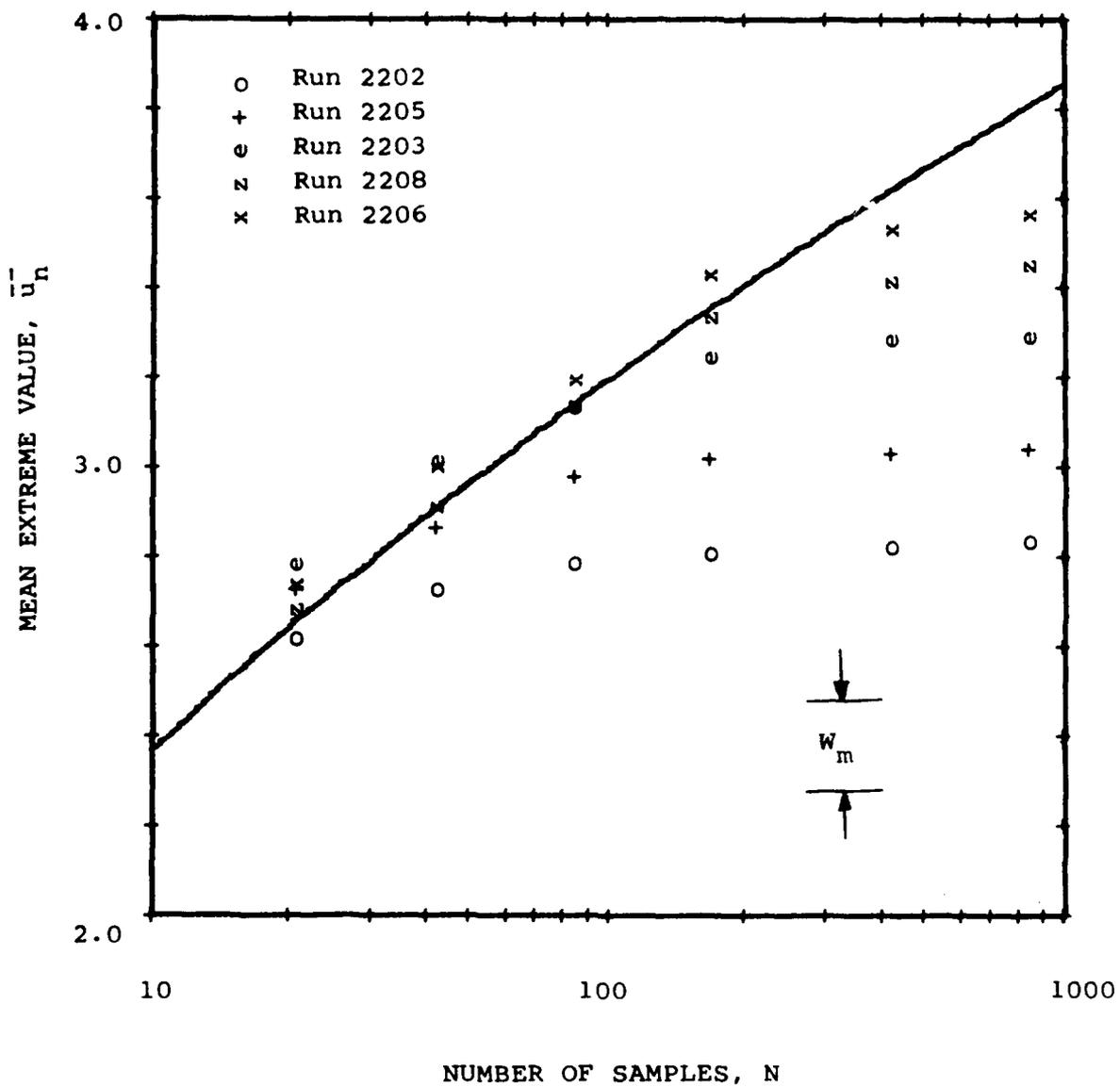


Figure 3 - Mean Extreme Values for Simulated Gaussian Time Histories

Figure 3; it describes the tolerance of these estimates. A summary of results for the five hertz sample rate is given in Table 2.

After equating these mean value estimates with the look-up table obtained from Equation (15), the ratio of the effective number of independent samples to the actual number of samples (N_e/N_a) was found. As a function of the estimated equivalent statistical bandwidth of the sampled envelope, plots of these ratios are given in Figure 4. Regardless of the bandwidth of the original time history, this N_e/N_a ratio is shown to be related only to this equivalent statistical bandwidth of the sampled envelope.

The equivalent statistical bandwidth is described in Appendix B; for our application, this bandwidth measure is related to the temporal extent of the sampled autocorrelation function. The equivalent statistical bandwidth, B_{st} in radians per second is defined as:

$$B_{st} = \frac{\left[\int_0^{\omega_0} G(\omega) d\omega \right]^2}{\int_0^{\omega_0} G(\omega)^2 d\omega} \quad (17)$$

where $\omega_0 = \pi/T$ (one half the sample rate in radians per second) and $G(\omega)$ is the autospectrum of the random process. From Equation (B.10) in Appendix B, the equivalent statistical bandwidth is related to the sampled autocorrelation function of the random process by

$$\frac{\omega_0}{B_{st}} = \sum_{k=-\infty}^{\infty} C(kT)^2 \quad (18)$$

Where $C(kT) = R(kT)/R(0)$ and $R(kT)$ is the autocorrelation function given by

$$R(kT) = E[x(nT) x(nT + kT)] \quad (19)$$

where $x(nT)$ is the random process sampled at equal time intervals, T . The

TABLE 2 - SUMMARY OF RESULTS FROM SIMULATED TIME HISTORY AT FIVE HZ SAMPLE RATE

Run #	Mean	W_m/Mean	N_e	W_n/N_e	ω_0/B_{st}	N_e/N_a	Chi-Square Test	
							N/N_e	Test Stat.
2202	2.835	0.082	34	0.68	65.52	0.040	1.00	4.33
2205	3.042	0.056	62	0.52	35.13	0.074	1.13	12.33
2203	3.291	0.059	135	0.64	20.80	0.161	1.04	13.67
2208	3.450	0.038	229	0.45	11.76	0.272	0.98	7.00
2206	3.565	0.061	342	0.79	8.40	0.407	0.96	7.33

TABLE 3 - AVERAGE OVERPREDICTION OF ENVELOPE MAXIMA

Run No.	Normalized Bandwidth	Average Overprediction in Percent
2202	0.153	1.9
2205	0.345	1.9
2203	0.575	2.7
2208	0.957	4.3
2206	1.341	6.6

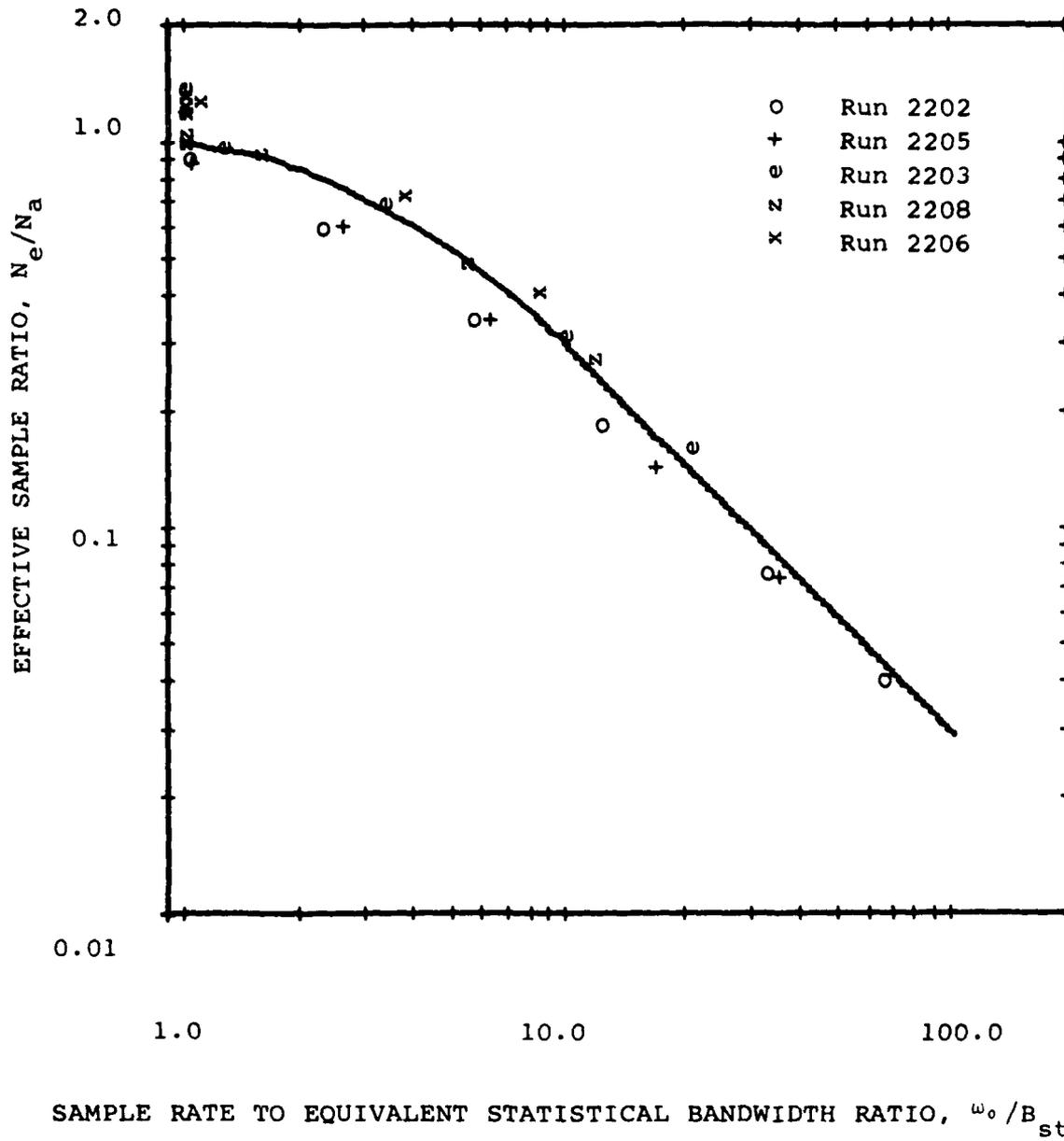


Figure 4 - Effective Sample Ratio for Simulated Time Histories

sample period, T , is the reciprocal of the sampling rate. The equivalent statistical bandwidth was estimated by direct measurement of the autocorrelation function over a finite number of lags such that the estimated autocorrelation function is short (decays to zero) compared to the available number of lags.

The equivalent statistical bandwidth is a constant for any random process as the sample period is changed. As the sample period, T , is increased (reduced sample rate), the autospectrum of the sampled random process is aliased until a limit is approached such that $B_{st}/\omega_0 \ll 1$; this implies that $C(kT)$ is one for $k=0$ and zero elsewhere. In this case, uncorrelated successive samples are used in Equation (19). For Gaussian random processes and for the Rayleigh distributed envelope, this also implies that statistically independent samples are obtained at this sample rate. Next, as the sample period is reduced (increased sample rate), the autospectrum of the sampled random process remains unchanged, so the equivalent statistical bandwidth is constant. The sum of the squared autocorrelation samples (left hand side of Equation (18)) is then directly proportional to ω_0 (one half the sample rate in radians per second).

A similar pair of asymptotic limits are expected for the N_e/N_a ratios given in Figure 4. When $B_{st}/\omega_0 = 1$, this ratio should be one since statistically independent samples are obtained for the Rayleigh distributed envelope. For high sample rates, this ratio should be inversely proportional to the effective statistical bandwidth, since the number of independent samples becomes constant.

The distribution of the largest envelope values from each of the 60 segments was tested against the hypothetical distribution given by Equation (13) to check the values estimated for N_e . Using the equal number of expected occurrences method, this hypothesis was tested using the Chi-Square goodness-of-fit test on histograms of these extreme values. Ten class intervals were used, so this test was conducted with seven degrees of freedom. As shown in Table 2, the Chi-Square test statistic was always less than or equal to 14.07, so, for all five cases, this hypothesis was accepted at the 5% level of significance. This test was performed using the N/N_e value given in Table 2 for the number of statistically independent samples. For runs 2202, 2203 and 2206, the Chi-Square test statistics plotted against N/N_e are given in Figure 5 where N is allowed to vary. Run 2203 fails this test over a wide range of N ; however,

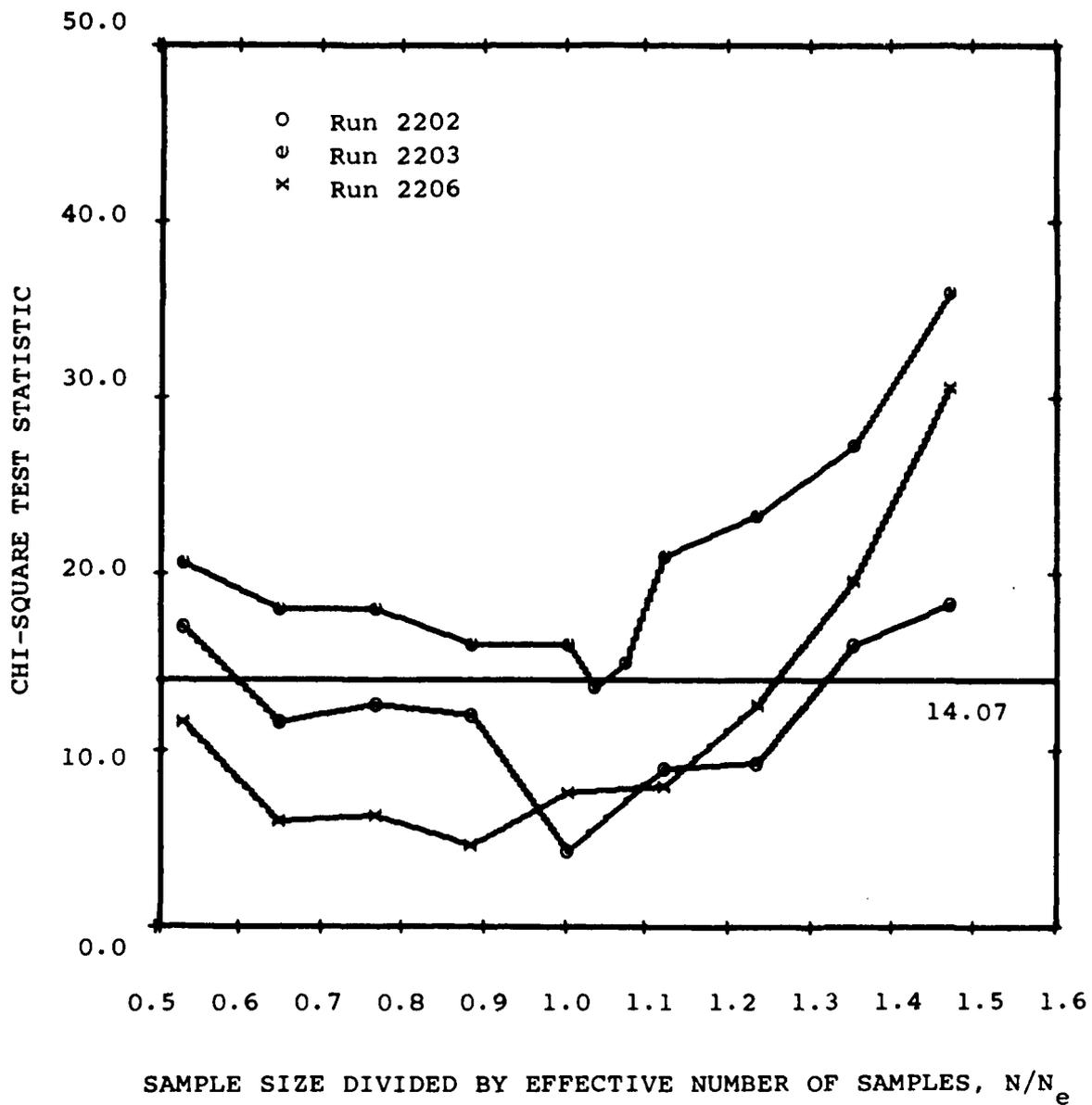


Figure 5 - Chi-Square Test Statistic for Three of the Simulated Gaussian Time Histories

histograms with the hypothetical distribution will fail the Chi-Square test five percent of the time. For these tests with variable N, the envelope time histories were sampled at five hertz.

The time history is always bounded by its Rayleigh distributed envelope; however, as seen in Figures A.1a to A.1e of Appendix A, the peaks of the time history do not always occur at the peaks of its envelope. For this to occur, the time history maxima must coincide with an envelope maxima and a phase angle of zero or 180 degrees (See Equation (3).) Since the envelope and phase angle are uncorrelated, such a mutual occurrence is unlikely. This characteristic can lead to overestimation of the extreme value reached by the time history. The size of this overestimation was evaluated using the five simulated time histories. For this study, each run was again divided into 60 segments. For each segment, the largest envelope value and the largest absolute time history value were found. The difference between these two values was divided by the largest envelope value, and this ratio was then averaged over the 60 segments to form a measure of the average overprediction. Expressed in percent, these averages are given in Table 3 where the normalized bandwidth is the bandwidth divided by the center frequency. For narrowbanded cases, this average overprediction is small; however, it increases with the bandwidth of the time history. Even in the most widebanded case, the predicted extremes in the envelope can be used in design applications for motion or wave induced force.

APPLICATION TO EXPERIMENTAL DATA

These procedures were applied to two sets of experimental data obtained from wave height measurements. The first data set is a 5940 second segment from the two off-shore platforms operated by the Naval Coastal Systems Laboratory (NCSL) at Panama City, Florida. The second data set consists of three 1750 second segments from hurricane Camille that occurred in 1969. For each time history, Appendix C contains plots of short data segments, histograms, autospectra and envelope autocorrelation function plots. The histograms were tested using the Chi-Square goodness-of-fit test.

The NCSL wave height data were taken at both Stages (off-shore platforms);

Stage 1 is located 11 miles (17.7 km) off-shore in 100 feet (30.5 m) of water and Stage 2 is located 2 miles (3.2 km) off-shore in 60 feet (18.3 m) of water. Wave height was measured using surface piercing wave gages. In this report, this data set is labeled run 2410; channel 1 is data from Stage 2 and channel 2 is from Stage 1. These data represent a mild sea condition; the standard deviation was 0.848 feet at Stage 1 and 0.526 feet at Stage 2. The major advantage of these data is that the sea state was reasonably stationary over the 5940 second run length; tidal mean value shift was removed by subtracting a best fit straight line from these data.

The wave height measurements taken of hurricane Camille are documented by Earle²⁰. The envelope procedures were applied to three 1752 second segments that were taken at 1015, 1130 and 1500 hours; these segments were assigned run numbers 2721, 2724 and 2731. The Camille data is assumed to be stationary over these run lengths. These data represent a rough sea condition; the standard deviation was 5.972 feet for run 2721, 5.838 feet for run 2724 and 9.966 feet for run 2731.

These wave data were analyzed using the same procedures that were applied to the simulated time histories. Each time history was divided into 60 segments and the largest envelope value from each segment was obtained. For the NCSL data, the segments were 99 seconds long and the following sample rates were used: 5.0, 2.5, 1.0, 0.5, 0.25 and 0.125 samples per second. The number of samples, N_a , was 495, 247, 99, 49, 24 and 12. The Camille data was already short, so 29.2 second segments were used at the following three sample rates: 2.5, 1.25 and 0.833 samples per second. For this data set, N_a was 73, 36 and 24. The mean extreme values for these data are given in Figure 6; the Camille data were not undersampled to the point where independent samples were obtained; however, both sets of results show the same trends as those seen in Figure 3. For the NCSL data, the 90% confidence interval ranged from 0.188 to 0.248, so 0.220 is a nominal width for this interval. The confidence interval for the Camille data varied from 0.244 to 0.304; a nominal width for this interval is 0.275. A summary of results for the highest sample rates is given in Table 4.

After equating mean extreme value estimates, the ratio of the effective number of independent samples to the actual number of samples (N_e/N_a) is given

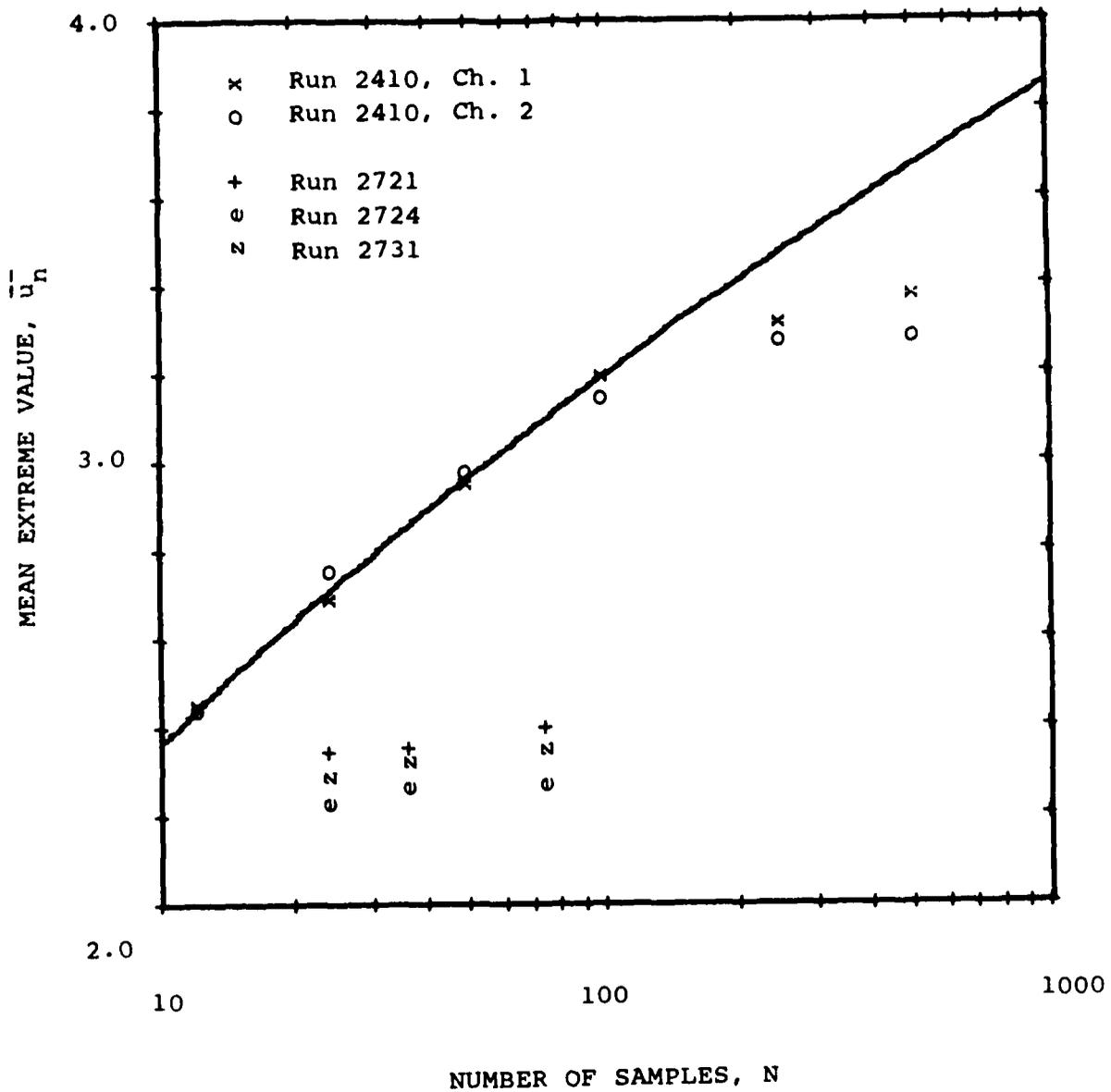


Figure 6 - Mean Extreme Values for Wave Height Data

TABLE 4 - SUMMARY OF RESULTS FROM MEASURED WAVE DATA AT HIGHEST SAMPLE RATE

Run #	Mean	W_m/Mean	N_e	W_n/N_e	ω_0/B_{st}	N_e/N_a	Chi-Square Test	
							N/N_e	Test Stat.
2410 Ch 1	3.376	0.062	179	0.71	11.21	0.362	1.00	7.67
2410 Ch 2	3.280	0.076	130	0.83	19.10	0.263	0.81	16.33
2721	2.400	0.126	11	0.64	22.86	0.151	0.91	4.33
2724	2.273	0.135	8	0.62	32.50	0.110	0.75	17.00
2731	2.353	0.129	10	0.70	27.84	0.137	0.90	8.67

TABLE 5 - HISTOGRAMS OF ENVELOPE MAXIMA FOR SELECTED MEASURED WAVE DATA

Run 2410 Ch 2 $N_e = 130$			Run 2724 $N_e = 8$			Run 2731 $N_e = 10$		
Interval		# of Occur	Interval		# of Occur	Interval		# of Occur
from	to		from	to		from	to	
0.00	2.84	12	0.00	1.66	10	0.00	1.78	10
2.84	2.97	8	1.66	1.85	10	1.78	1.95	9
2.97	3.06	3	1.85	1.98	5	1.95	2.09	4
3.06	3.15	5	1.98	2.11	6	2.09	2.21	7
3.15	3.24	3	2.11	2.23	3	2.21	2.33	4
3.24	3.33	8	2.23	2.36	1	2.33	2.45	4
3.33	3.44	0	2.36	2.50	6	2.45	2.59	4
3.44	3.57	5	2.50	2.68	4	2.59	2.76	4
3.57	3.77	2	2.68	2.94	3	2.76	3.02	4
3.77	∞	14	2.94	∞	12	3.02	∞	10
Test Stat: 30.00			Test Stat: 19.33			Test Stat: 11.00		

TABLE 6 - AVERAGE OVERPREDICTION OF ENVELOPE MAXIMA FOR MEASURED WAVE DATA

Run #	Average Overprediction in Percent
2410 Ch 1	9.8
2410 Ch 2	8.3
2721	7.4
2724	8.3
2731	5.2

in Figure 7 as a function of the estimated equivalent statistical bandwidth. The solid curve is the curve fitted to the simulated time history results given in Figure 4. These results from the wave data closely match the results from the simulated time histories. The N_e/N_a ratio is again shown to be related to the equivalent statistical bandwidth of the sampled envelope.

The distribution of the largest envelope values from each of the 60 segments was tested against the hypothetical distribution given by Equation (13). The Chi-Square goodness-of-fit test results are given in Table 4; the equal number of expected occurrences method was used with a total of 10 intervals. For runs 2410 (channel 2) and 2724, the Chi-Square test statistic is greater than 14.07, so the test failed at the 5% level of significance. The Chi-Square test statistic plotted against N/N_e is given in Figure 8. In those cases where the test failed, larger numbers of occurrences were seen in the first and last intervals; the extreme envelope values appear to be scattered over a wider range than expected by the hypothetical distribution. Histograms for these runs at $N = N_e$ are listed in Table 5; the histogram for run 2731 is included to give an example of a case that passed the Chi-Square test even though it shows wide scatter.

The average overprediction of the envelope maxima compared to the time history maxima is given in Table 6 for each run. The overprediction is several percent larger than the values seen for the simulated time histories.

When applying these procedures to experimental data, the N_e/N_a ratio for a particular run is the principal result that is then used to estimate extreme value statistics using the methods of order statistics. Since N_a (actual) samples were obtained over a known time period, the effective number of statistically independent samples, N_e , over this time period is N_a multiplied by this ratio. This time period is known, so this ratio can be converted to the effective sample rate that gives statistically independent samples (N_e divided by this time period). Extrapolation to longer exposure times is accomplished by varying this time interval and calculating a new value for N_e . The expected or mean extreme value for this time period, for example, is calculated using Equations (13) and (15) where N is equal to N_e . The N_e/N_a ratio is estimated or measured from the available experimental data, so estimates of this mean extreme value are obtained.

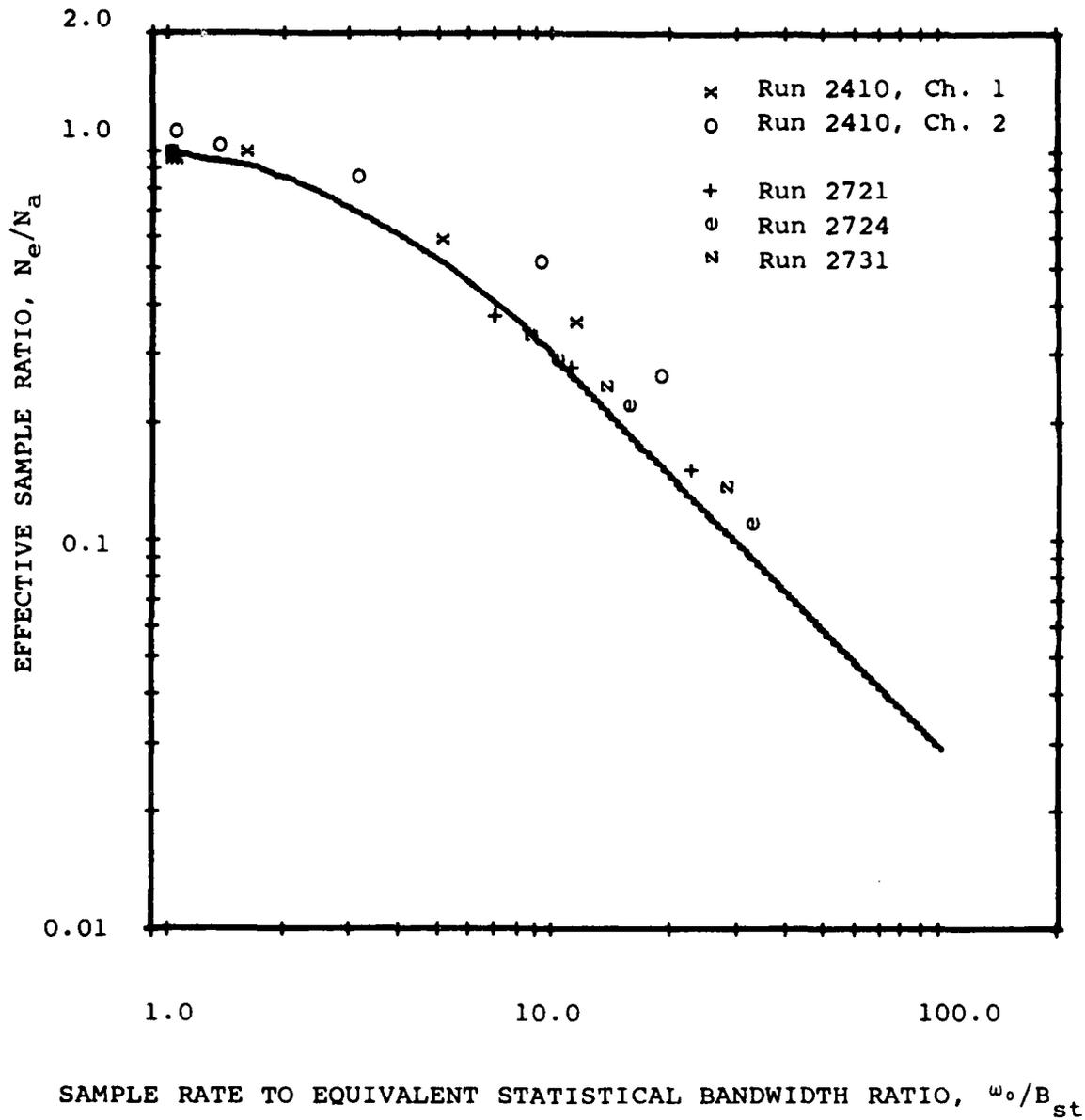


Figure 7 - Effective Sample Ratio for Wave Height Data

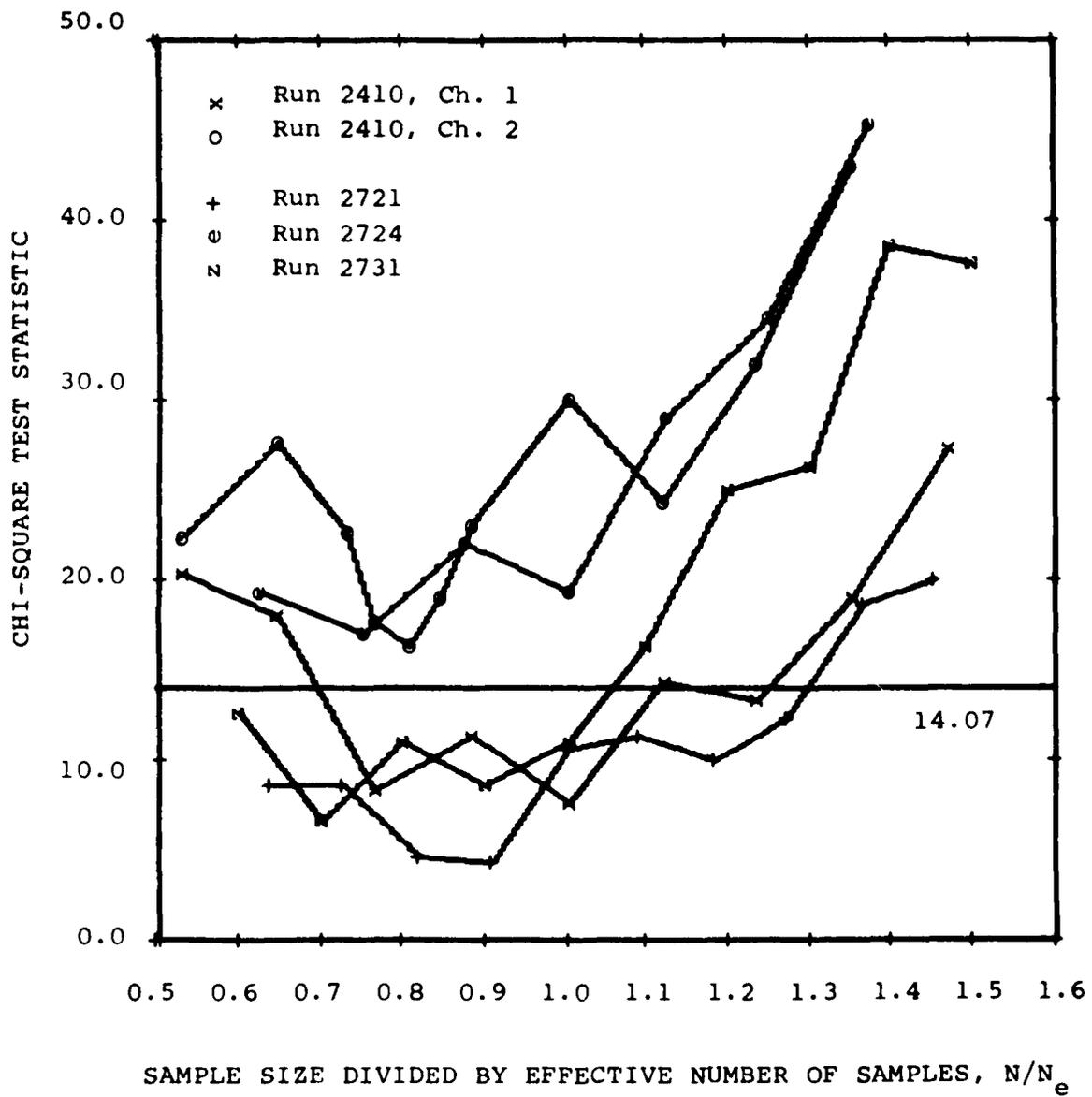


Figure 8 - Chi-Square Test Statistic for Wave Height Data

CONCLUSIONS

In the design of ships and other marine structures, the ability to accurately predict extremes in wave induced motions or forces is limited by the presence of correlated samples in the measured data. This report presents a procedure for measuring the influence of these correlated samples in this measured data; corrections are then applied to these extreme value predictions. The concept of an analytic envelope for arbitrary bandwidth Gaussian random processes is employed in determining this correction; however, extreme value predictions are still made within the framework of standard order statistics. This correction is shown to be related to the equivalent statistical bandwidth of the random process's envelope time history. This relationship is demonstrated for simulated data and for experimentally obtained wave data.

The analytic envelope definition provides a consistent, reasonably easy to apply method for specifying peak occurrences of arbitrary bandwidth Gaussian random processes. Since the envelope is Rayleigh distributed, this approach retains many of the features found in the peak-to-trough approach given by Longuet-Higgins¹. Digital filtering is required to produce the envelope time history, so the computer time required to process experimental data is a factor in selecting this approach.

In a given time history, the envelope maxima are generally larger than the time history maxima. The worst cases seen were an average overprediction of 6.6% for the simulated time histories and 9.8% for the measured wave data. Since the envelope is an upper bound for a given time history, the extreme value predictions will be conservative; these predictions can be used for design applications for motion or wave induced force. When extrapolating to longer exposure times, an important question is the probability law obeyed by the extreme values; the Chi-Square test performed on the largest envelope values from each segment will demonstrate if the assumed distribution is adequate.

APPENDIX A
SIMULATED GAUSSIAN TIME HISTORIES AND THEIR ENVELOPES

The methods used to produce the simulated time histories are described in this appendix. Also presented here are 20 second long plots of each time history and its envelope. Histograms of each time history were obtained and Chi-Square tests were run to test the input time history against the Gaussian distribution and the envelope was tested for the Rayleigh distribution. Examples of the histograms are given. Also shown are plots of their autospectra as well as a plot of each envelope's autocorrelation function.

The method for determining the effective sample rate used in the order statistics estimates requires this series of simulated Gaussian time histories. These time histories were produced using algorithms implemented on a digital computer. Successive calls to the random number generator produced bandlimited white noise; this time history was then filtered using a non-recursive digital filter to achieve the desired spectral shape. The filter was designed using procedures similar to those described by Helms¹⁸. Since the random number generator produces uniformly distributed, statistically independent samples and thirty of these samples were averaged to produce one sample of the bandlimited white noise, the distribution of this noise and the filtered noise is approximately Gaussian.

Five different time histories were produced that had bandwidths varying from the extreme narrowband to wideband. In each case, the spectral shape was a rectangular block centered at 0.51 hertz. The bandwidths were 0.078, 0.176, 0.293, 0.488 and 0.684 hertz. These bandwidths correspond to runs 2202, 2205, 2203, 2208 and 2206. The variance of each time history was set to one volt. The duration of these time histories was 10,080 seconds at a sample rate of five hertz. The envelope time histories were then obtained using the Rabiner and Schafer¹⁶ method to produce the Hilbert transform. Figures A.1a through A.1e show a short segment of each time history; the envelope is plotted on each side of the axis. The segments shown in Figure A.1 present the most extreme event that occurred in each time history. For the narrowbanded cases, the envelope tends to follow the usual concept of a line drawn from one time history maximum

to the next; however, this is not evident in the wideband cases. The estimated or measured autospectra of these time histories are given in Figure A.2. The envelope autospectra are a maximum at or near zero hertz; they tend to show the triangular shape that would be expected from convolving the time history spectrum with itself. The frequency range over which they are non-zero appears to be equal to the range over which the time history spectra are non-zero. The estimated autocorrelation functions of each envelope time history are given in Figure A.3. As expected, the narrowbanded spectra have autocorrelations that go more slowly to zero. These autocorrelation functions are also triangular in shape. A summary of the basic statistics of each time history is given in Table A.1. In Table A.1, the normalized bandwidth is the bandwidth divided by the center frequency (0.51 hertz). The standard deviation is estimated from the Gaussian time history. The normalized maximum is the most extreme level (absolute value) that occurred in that time history divided by the standard deviation. The normalized envelope RMS is the square root of the variance plus the mean squared for the envelope time history divided by the standard deviation; as seen from Equation (10), this ratio should be 1.414. In each case, the estimated RMS is very close to this expected value. The normalized envelope maximum is the most extreme level of the envelope time history divided by the standard deviation. The normalized peak-to-trough maximum is the most extreme peak-to-trough of the Gaussian time history divided by the standard deviation. The peak-to-trough values are divided by two to make them comparable to the envelope results.

Data point histograms were obtained from each time history and its envelope. Using these histograms, the Chi-Square goodness-of-fit test was performed over a range of at least three times the standard deviation of the time history to determine the fit to the Gaussian distribution for the time history and to the Rayleigh distribution for the envelope. The results of these tests are given in Table A.2; in each case, the hypothesis was accepted at the 5% level of significance. The number in parenthesis under the Chi-Square statistic is the threshold for acceptance or rejection of the goodness-of-fit test. The equal interval Chi-Square goodness-of-fit procedures given by Bendat and Piersol¹⁷ were used. Since this test requires statistically independent samples, the two time histories were sampled at the sample period given in

Table A.2. The estimated autocorrelation of each time history has decayed well into the random noise level at this time interval.

For runs 2202, 2203 and 2206, data point and peak-to-trough histograms are presented in Figures A.4, A.5 and A.6; these histograms supplement the Chi-Square test results by giving a visual fit to the hypothetical distributions. The data point histogram for the time history is plotted with the Gaussian distribution, and the data point histogram for the envelope as well as the peak-to-trough histogram is plotted with the Rayleigh distribution. To make the scales comparable to the envelope data point histogram, the peak-to-trough values were divided by two prior to computing their histogram. Every sample was used to create these histograms; these are not the same histograms that were used in the Chi-Square test. The Chi-Square test was not performed on the peak-to-trough histograms since statistically independent samples are required for this test. The two data point histograms consistently match their hypothetical distribution. As the bandwidth is increased, however, the number of occurrences in the main body of the peak-to-trough histogram is seen to increase, while the number of occurrences in the tail begin to decrease; this behaviour is predicted by Tayfun⁵. In these figures, the number of occurrences in equal width class intervals are plotted at each class midpoint.

TABLE A.1 - BASIC STATISTICS OF THE SIMULATED TIME HISTORIES AND THEIR ENVELOPES (OVER ENTIRE RUN)

Run #	Time History			Envelope		Peak-to-Trough (divided by 2)
	Band-width	Sigma, σ volts	Maximum	RMS	Maximum	Maximum
2202	0.153	1.015	5.070	1.409	5.080	5.034
2205	0.345	0.980	4.199	1.414	4.339	4.007
2203	0.575	0.973	4.373	1.411	4.378	4.087
2208	0.957	1.016	3.920	1.412	4.083	3.740
2206	1.341	0.994	5.052	1.412	5.773	3.692

TABLE A.2 - CHI-SQUARE GOODNESS-OF-FIT TEST FOR SIMULATED TIME HISTORIES AND THEIR ENVELOPES

Run #	Sample Period sec	Gaussian Distribution (Time History)		Rayleigh Distribution (Envelope)	
		Degrees of Freedom	Chi-Square Test Stat.	Degrees of Freedom	Chi-Square Test Stat.
2202	16.0	18	21.47 (28.87)*	8	8.83 (15.51)
2205	13.0	17	20.73 (27.59)	10	13.61 (18.31)
2203	10.0	18	15.18 (28.87)	8	7.24 (15.51)
2208	6.0	19	18.27 (30.14)	8	11.49 (15.51)
2206	4.0	29	27.55 (42.56)	9	5.80 (16.92)

* Rejection threshold for Chi-Square Test

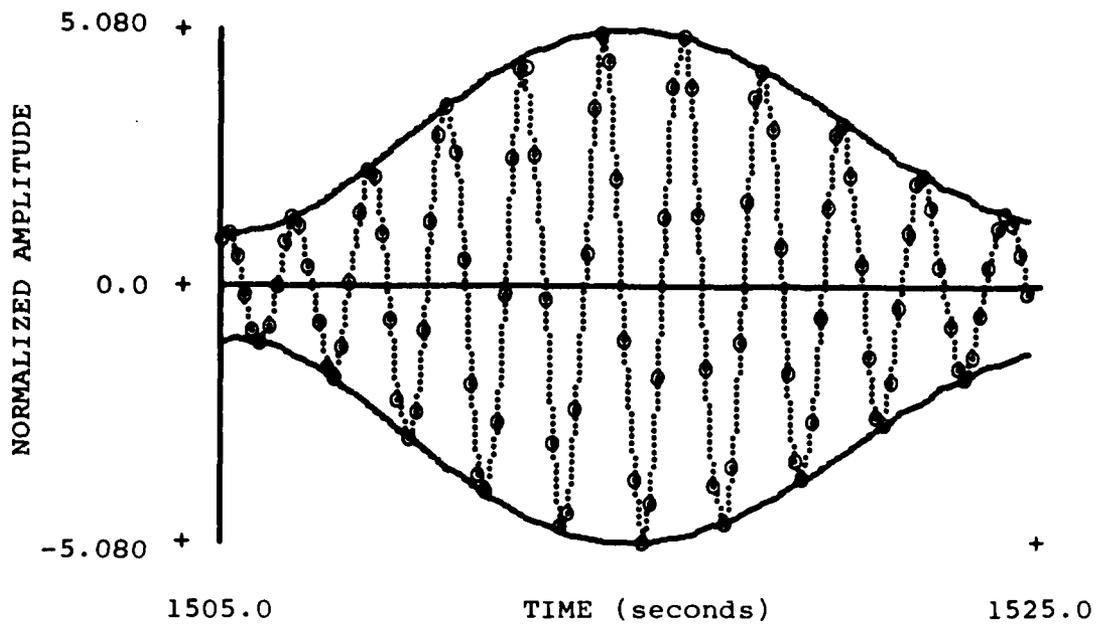


Figure A.1.a - Twenty Second Segment from Run 2202

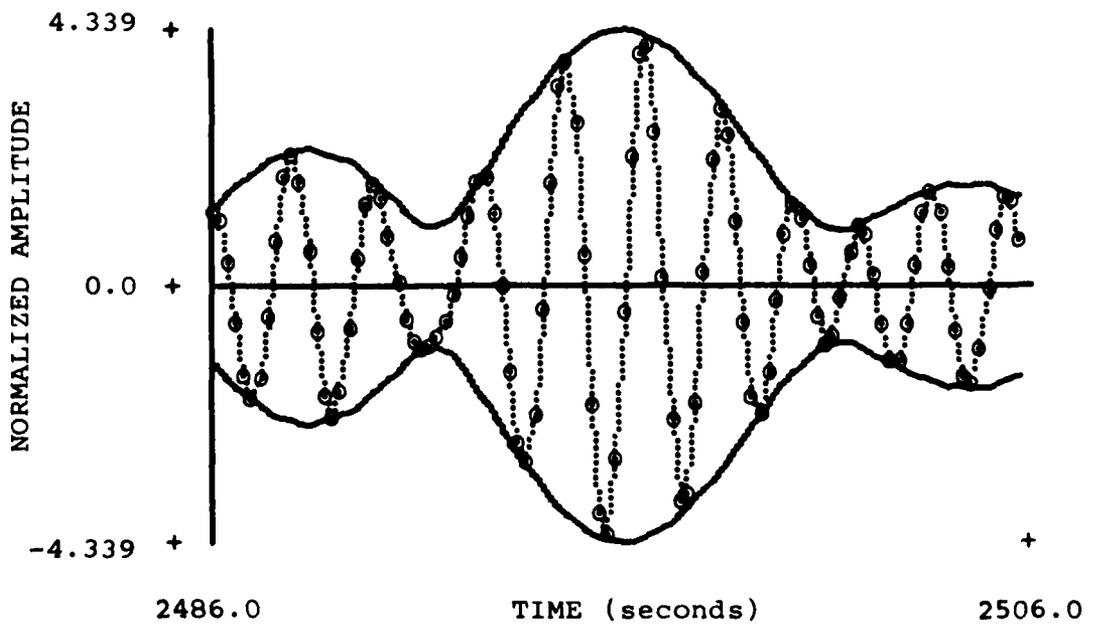


Figure A.1.b - Twenty Second Segment from Run 2205

Figure A.1 - Time History Segments of Simulated Data and their Envelopes

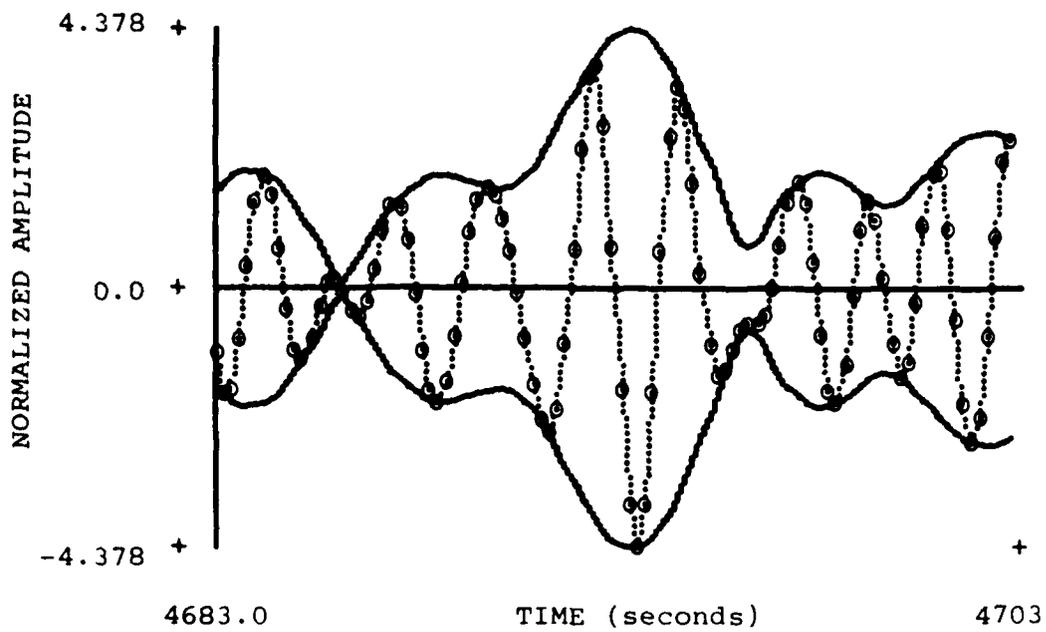


Figure A.1.c - Twenty Second Segment from Run 2203

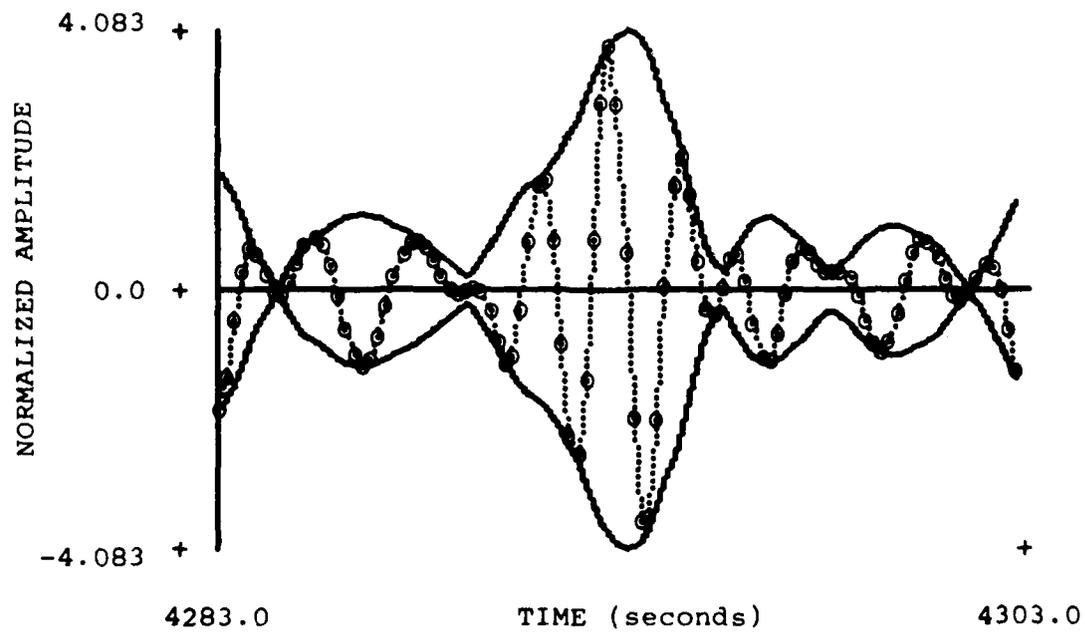


Figure A.1.d - Twenty Second Segment from Run 2208

Figure A.1 - Time History Segments of Simulated Data and their Envelopes (Cont.)

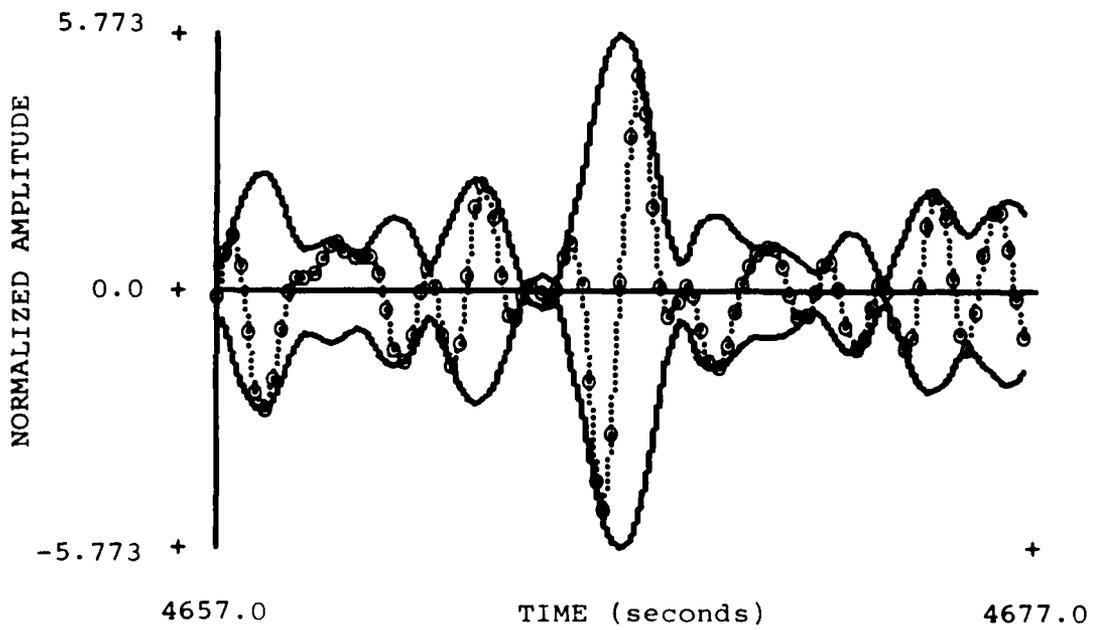


Figure A.1.e - Twenty Second Segment from Run 2206

Figure A.1 - Time History Segments of Simulated Data and their Envelopes (Cont.)

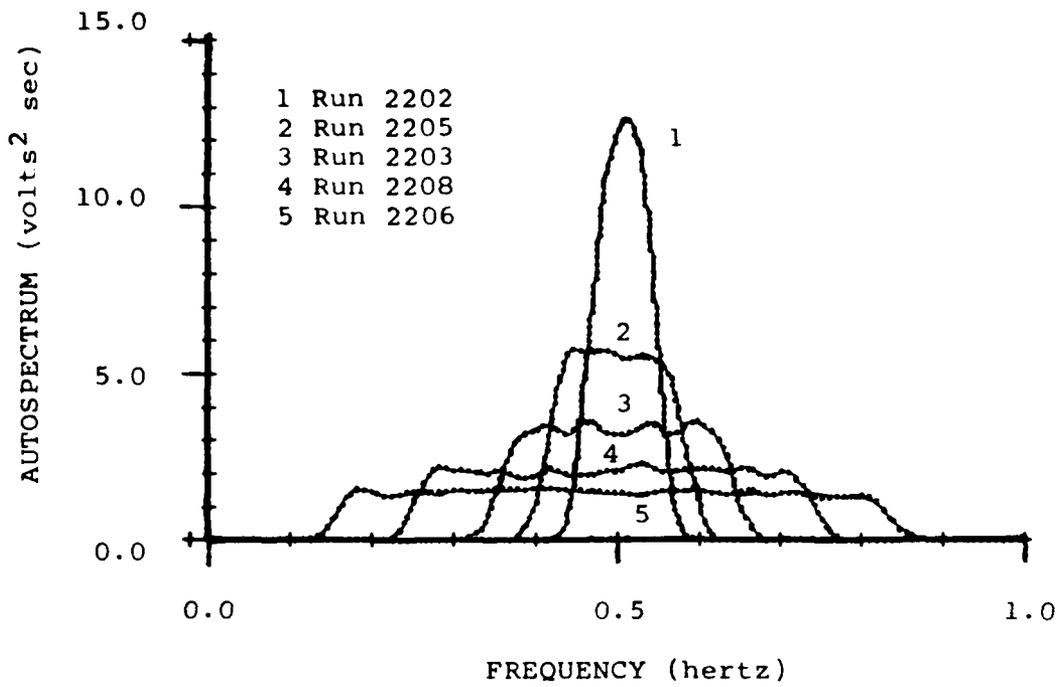


Figure A.2.a - Autospectra of Simulated Gaussian Time Histories

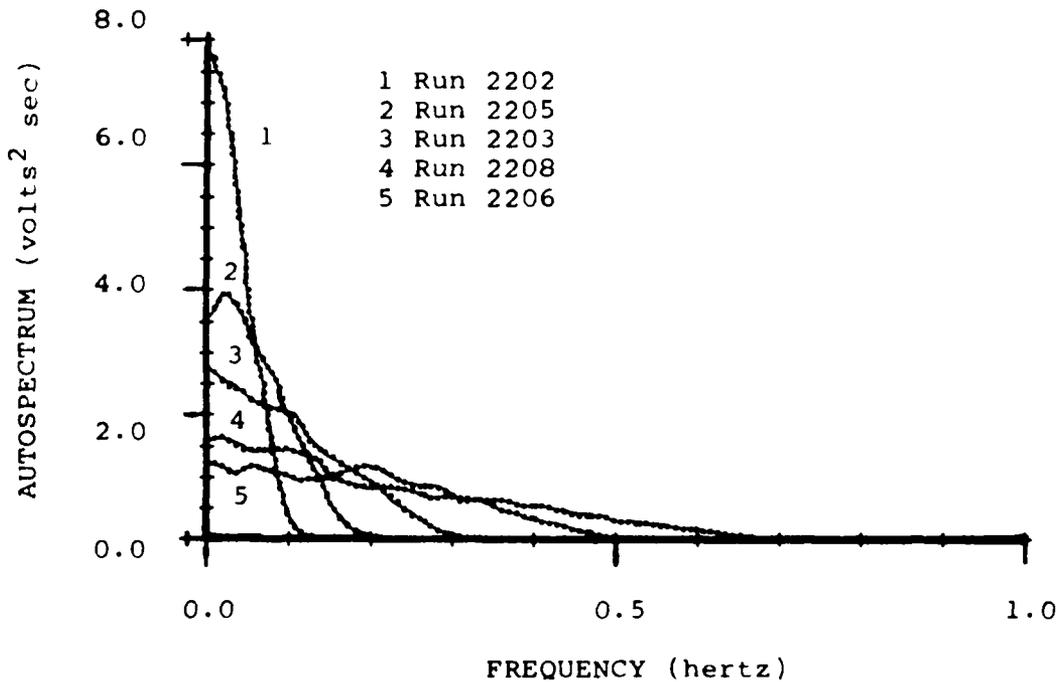


Figure A.2.b - Autospectra of Envelope Time Histories

Figure A.2 - Autospectra for Simulated Time Histories and their Envelopes

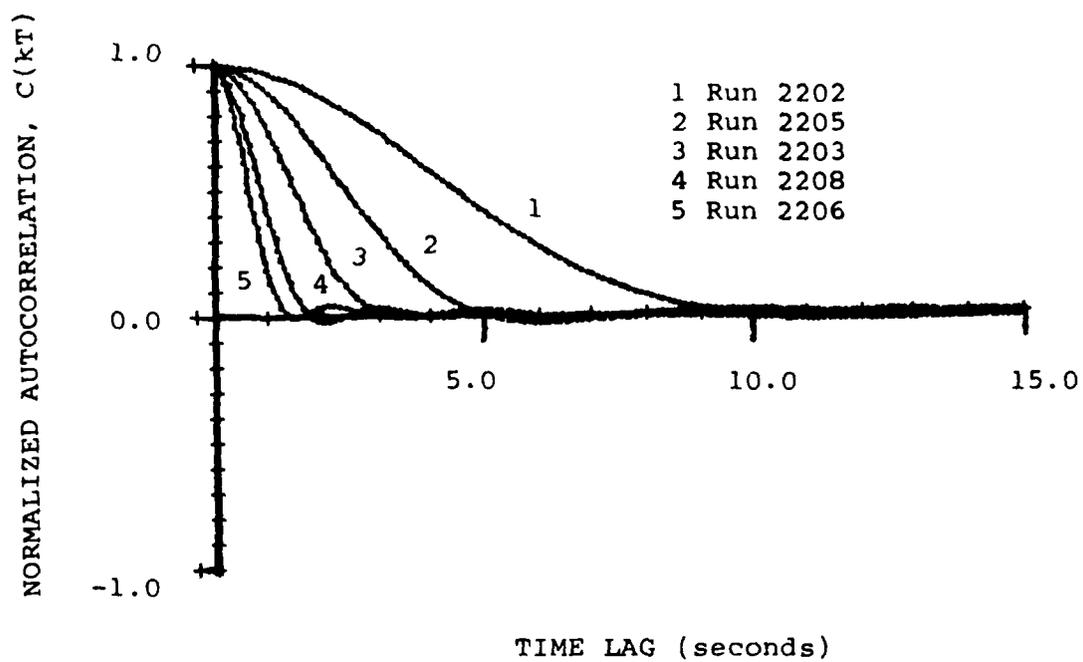


Figure A.3 - Autocorrelation Functions for Envelopes of Simulated Time Histories

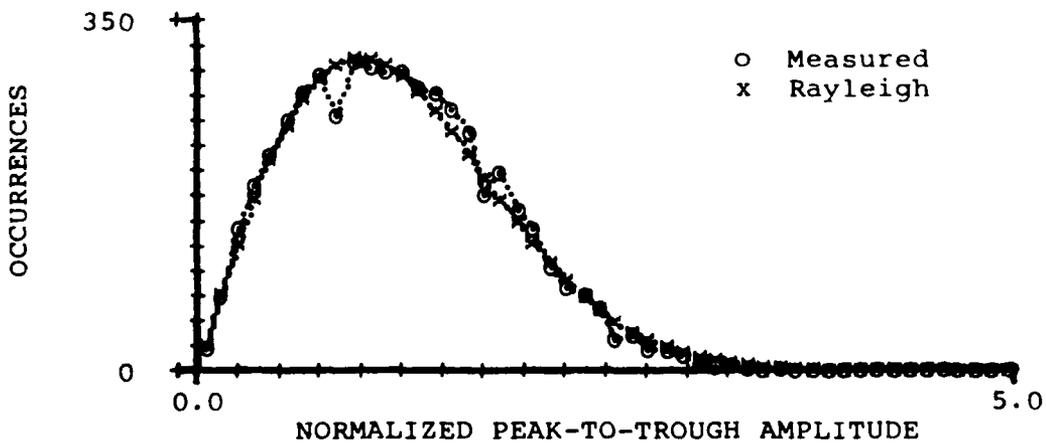
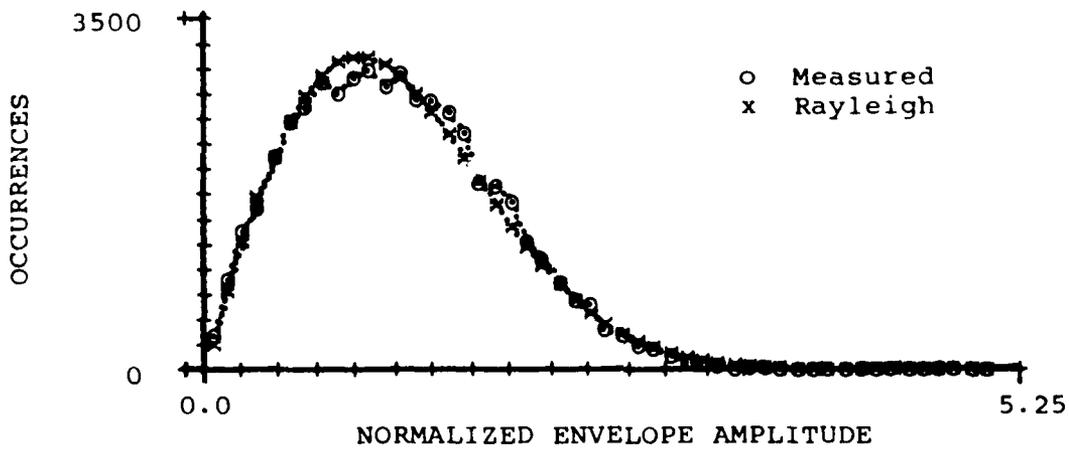
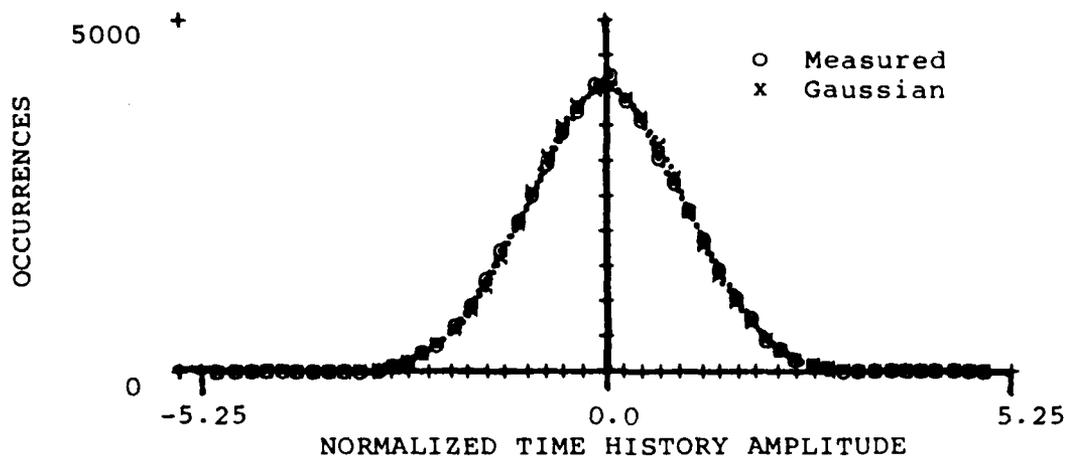


Figure A.4 - Histograms for Run 2202

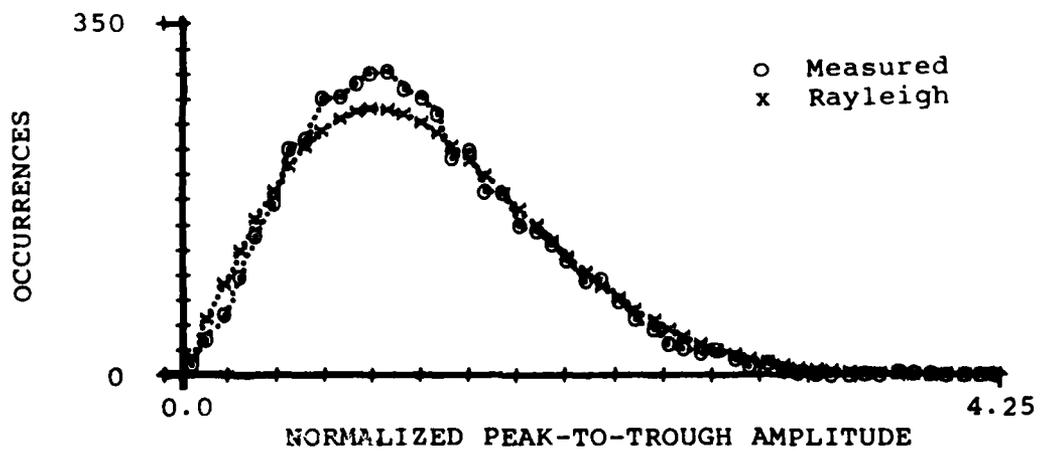
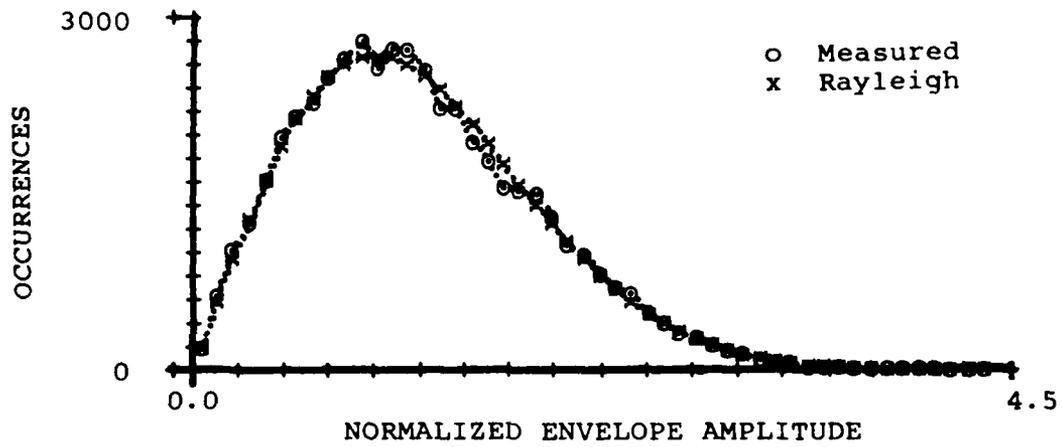
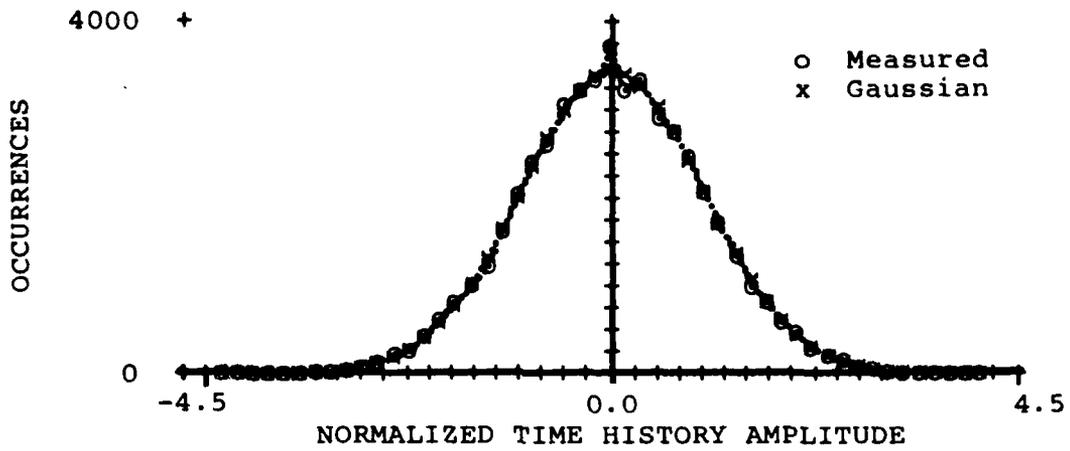


Figure A.5 - Histograms for Run 2203

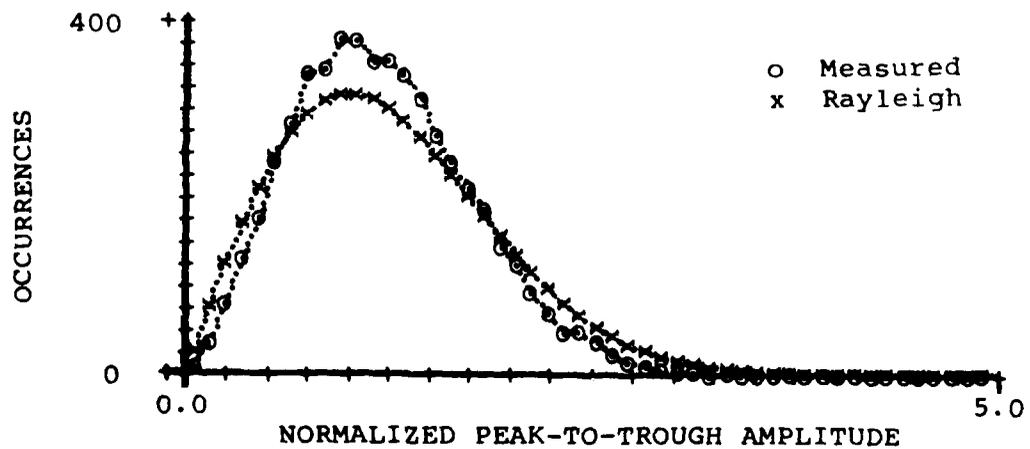
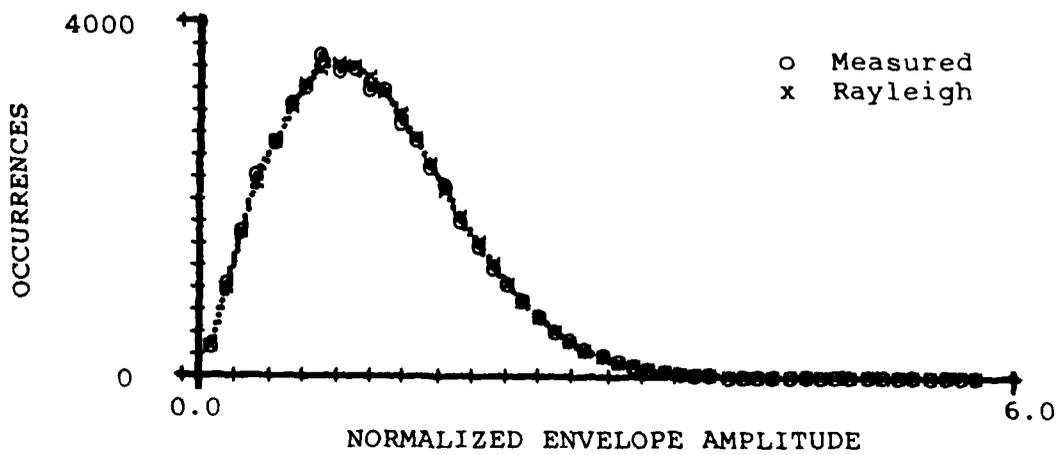
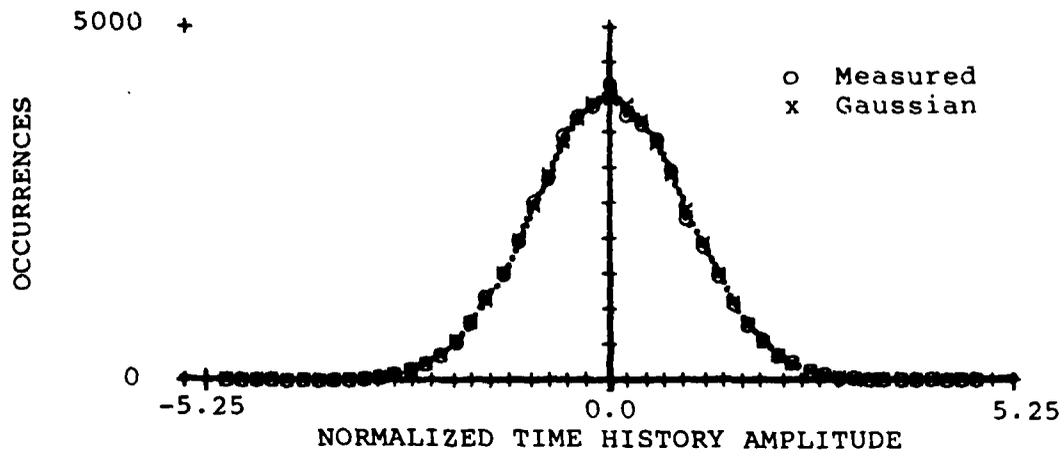


Figure A.6 - Histograms for Run 2206

APPENDIX B
RELATIONSHIP BETWEEN EQUIVALENT STATISTICAL BANDWIDTH AND
THE AUTOCORRELATION FUNCTION

The definition of the equivalent statistical bandwidth of a filter is given by Bendat and Piersol¹⁷; their definition is now extended to the equivalent statistical bandwidth of a random process with any distribution. From Bendat and Piersol, this "is the bandwidth of a hypothetical rectangular filter which would pass a signal with the same mean square value statistical error as the actual filter when the input is white noise". The mean square value statistical error relationships derived by Bendat and Piersol assume a Gaussian distributed random process; however, this bandwidth definition has meaning for random processes other than Gaussian. The equivalent statistical bandwidth, B_{st} , in radians per second of a random process is defined as:

$$B_{st} = \frac{\left[\int_0^{\omega_0} G(\omega) d\omega \right]^2}{\int_0^{\omega_0} G(\omega)^2 d\omega} \quad (B.1)$$

where $\omega_0 = \pi/T$ (one half the sample rate in radians per second) and $G(\omega)$ is the autospectrum of the random process. Since this random process is sampled, it is bandlimited ($G(\omega) = 0$); it has nonzero values only in the frequency interval $-\omega_0 \leq \omega \leq \omega_0$. The relationship between the equivalent statistical bandwidth and the autocorrelation function of the random process is now derived; this derivation is independent of the distribution of the random process. It is based on sampled random processes; however, this derivation can be extended to continuous processes.

Balakrishnan¹⁹ proved the following relationships between autocorrelation and the autospectrum for sampled, stationary random processes. The autocorrelation function, $R(kT)$, is given by

$$R(kT) = E[x(nT) x(nT + kT)] \quad (B.2)$$

where $x(nT)$ is the sampled time history. The Fourier transform pair are

$$R(kT) = \frac{1}{2} \int_{-\omega_0}^{\omega_0} G(\omega) e^{ikT\omega} d\omega \quad (B.3)$$

and

$$G(\omega) = \frac{1}{\omega_0} \sum_{k=-\infty}^{\infty} R(kT) e^{-ikT\omega} \quad (B.4)$$

The top integral in Equation (B.1) is the variance of the random process, σ^2 , the zeroth autocorrelation lag:

$$\sigma^2 = R(0) = \frac{1}{2} \int_{-\omega_0}^{\omega_0} G(\omega) d\omega = \int_0^{\omega_0} G(\omega) d\omega \quad (B.5)$$

The lower integral in Equation (B.1) is found by substituting Equation (B.4) for $G(\omega)$:

$$\int_0^{\omega_0} G(\omega)^2 d\omega = \frac{1}{2\omega_0^2} \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} R(kT) R(lT) \int_{-\omega_0}^{\omega_0} e^{-i\omega T(k+l)} d\omega \quad (B.6)$$

Since

$$\int_{-\omega_0}^{\omega_0} e^{-i\omega T(k+l)} d\omega = \begin{cases} 2\omega_0, & k+l = 0 \\ 0, & \text{otherwise} \end{cases} \quad (B.7)$$

then

$$\int_0^{\omega_0} G(\omega)^2 d\omega = \frac{1}{\omega_0} \sum_{k=-\infty}^{\infty} R(kT)^2 \quad (B.8)$$

Substituting Equations (B.5) and (B.8) into Equation (B.1) gives the relationship between the equivalent statistical bandwidth and the autocorrelation

function of the random process:

$$B_{st} = \omega_0 \frac{R(0)^2}{\sum_{k=-\infty}^{\infty} R(kT)^2} \quad (B.9)$$

or

$$\frac{\omega_0}{B_{st}} = \sum_{k=-\infty}^{\infty} C(kT)^2 \quad (B.10)$$

Where $C(kT)$ is the normalized autocorrelation function,

$$C(kT) = \frac{R(kT)}{R(0)} \quad (B.11)$$

Equation (B.9) gives the relationship between the equivalent statistical bandwidth and the autocorrelation function. At a fixed sample rate, the equivalent statistical bandwidth is inversely proportional to the sum of the squared autocorrelation samples. In general, the statistical bandwidth is a measure of the extent (in time) of the autocorrelation function; it summarizes this feature of the autocorrelation.

APPENDIX C
MEASURED WAVE HEIGHT DATA

The experimental data were obtained from two sets of wave height measurements. The first data set is a 5940 second segment from the two off-shore platforms operated by the Naval Coastal Systems Laboratory (NCSL) at Panama City, Florida. The second data set consists of three 1752 second segments from hurricane Camille that occurred in 1969. This appendix contains plots of short data segments, histograms, autospectra and envelope autocorrelation function plots for each data run. Chi-Square goodness-of-fit tests were performed on the histograms.

The NCSL wave height data were taken at both Stages (off-shore platforms); Stage 1 is located 11 miles (17.7 km) off-shore in 100 feet (30.5 m) of water and Stage 2 is located 2 miles (3.2 km) off-shore in 60 feet (18.3 m) of water. Wave height was measured using surface piercing wave gages. In this report, this data set is labeled run 2410; channel 1 is data from Stage 2 and channel 2 is from Stage 1. These two wave height records are reasonably stationary over the 5940 second run length; tidal mean value shift was removed by subtracting a best fit straight line from these data. The hurricane Camille wave height data are documented by Earle²⁰. The three 1752 second long data segments selected for analysis were taken at 1015, 1130 and 1500 hours; these segments were assigned run numbers 2721, 2724 and 2731. The Camille data is assumed to be stationary over these run lengths.

Both sets of wave data were processed to obtain the envelope time history, autospectra, envelope autocorrelation functions, data point histograms and peak-to-trough histograms, as well as basic statistics. The envelope time histories were then obtained using the Rabiner and Schafer¹⁶ method to produce the Hilbert transform. A summary of the basic statistics of each time history is given in Table C.1. The standard deviation is estimated from the wave time history. The normalized maxima is the most extreme level (absolute value) that occurred in that time history divided by the standard deviation. The normalized envelope RMS is the square root of the variance plus the mean squared for the envelope time history divided by the standard deviation; as seen from Equation

(10), this ratio should be 1.414. In each case, the estimated RMS is very close to this expected value. The normalized envelope maxima is the most extreme level of the envelope time history divided by the standard deviation. The normalized peak-to-trough maxima is the most extreme peak-to-trough of the wave time history divided by the standard deviation.

Figures C.1 through C.12 present time history segments, autospectra, histograms and autocorrelation function estimates from these time histories. These figures are organized by wave data set and run number in the following manner: For run 2410, 20 second segments of each time history are given in Figure C.1 for channel 1 and Figure C.4 for channel 2; for the Camille wave data, 60 second segments are given in Figures C.8 for run 2721, Figure C.11 for run 2724 and Figure C.14 for run 2731. Each segment brackets the most extreme event that occurred during that run; the envelope is plotted on each side of the axis. For the corresponding runs, wave and envelope autospectrum estimates are plotted in Figures C.2, C.5, C.9, C.12 and C.15. Similarly, the data point histograms and peak-to-trough histograms are plotted in Figures C.3, C.6, C.10, C.13 and C.16. Autocorrelation estimates for the envelope time history are given in Figure C.7 for channels 1 and 2 of run 2410; the autocorrelation estimates for runs 2721, 2724 and 2731 are given in Figure C.17.

The envelope autospectra given in Figures C.2, C.5, C.9, C.12 and C.15 show the general features of the envelope autospectra for the simulated data presented in Appendix A. Similarly, the autocorrelation functions given in Figures C.7 and C.17 share the general features of the simulated data autocorrelation functions. The wave autospectrum for run 2410, Channel 1 is bimodal; however, this feature does not seem to affect its envelope auto spectrum or its autocorrelation function estimate.

Data point and peak-to-trough histograms are presented in Figures C.3, C.6, C.10, C.13 and C.16; these histograms supplement the Chi-Square test results presented below by giving a visual fit to the hypothetical distributions. The data point histograms for the time histories (Figures C.3a, C.6a, C.10a, C.13a and C.16a) are plotted with the Gaussian distribution, and the data point histograms for the envelope (Figures C.3b, C.6b, C.10b, C.13b and C.16b) as well as the peak-to-trough histograms (Figures C.3c, C.6c, C.10c, C.13c and C.16c) are plotted with the Rayleigh distribution. Every sample was

used to create these histograms; these are not the same histograms that were used in the Chi-Square tests described below. Except for the peak-to-trough histograms, a good qualitative match to their hypothetical distribution is seen. The peak-to-trough histograms appear to have more occurrences near the peak of the hypothetical distribution, and less occurrences in the tail; this feature was more obvious with the widebanded simulated time histories. In every case, the wave data point histograms have occurrences at a higher level on the positive side of the distribution than they do on the negative side; non-linear wave effect probably account for this feature; however, except for channel 1 of run 2410, the other wave data passed the Chi-Square test for the Gaussian distribution.

An additional set of data point histograms was obtained from each time history and its envelope. Using these histograms, the Chi-Square goodness-of-fit test was performed over a range of at least three times the standard deviation of the time history for channels 1 and 2 of run 2410 and over a range of at least two times the standard deviation for runs 2721, 2724 and 2731. These tests were performed to determine the fit to the Gaussian distribution for the wave data and to the Rayleigh distribution for the envelope. The Chi-Square test was not performed on the peak-to-trough histograms since statistically independent samples are required for this test. The results of these tests are given in Table C.2; except for run 2410 channel 1, the hypothesis was accepted at the 5% level of significance. The number in parenthesis under the Chi-Square statistic is the threshold for acceptance or rejection of the goodness-of-fit test. The equal interval Chi-Square goodness-of-fit procedures given by Bendat and Piersol¹⁷ were used. Since this test requires statistically independent samples, the two time histories were sampled at the sample period given in Table C.2. The estimated autocorrelation of each time history has decayed well into the random noise level at this time interval.

TABLE C.1 - BASIC STATISTICS OF THE MEASURED WAVE DATA AND THEIR ENVELOPES (OVER ENTIRE RUN)

Run #	Time History		Envelope		Peak-to-Trough (divided by 2)
	Sigma, σ feet	Maximum	RMS	Maximum	Maximum
2410 Ch 1	0.526	4.643	1.411	4.738	3.379
2410 Ch 2	0.848	4.476	1.417	4.509	3.445
2721	5.972	3.674	1.418	3.878	2.877
2724	5.838	4.722	1.419	4.764	4.374
2731	9.966	4.827	1.418	4.849	3.692

TABLE C.2 - CHI-SQUARE GOODNESS-OF-FIT TEST FOR MEASURED WAVE DATA AND THEIR ENVELOPES

Run #	Sample Period sec	Gaussian Distribution (Time History)		Rayleigh Distribution (Envelope)	
		Degrees of Freedom	Chi-Square Test Stat.	Degrees of Freedom	Chi-Square Test Stat.
2410 Ch 1	15.0	17	30.34 (27.59)*	9	15.74 (16.92)
2410 Ch2	15.0	15	11.89 (25.00)	7	3.06 (14.07)
2721	30.0	6	4.32 (12.59)	4	7.58 (9.49)
2724	30.0	7	8.53 (14.07)	5	5.40 (11.07)
2731	30.0	7	8.12 (14.07)	4	1.14 (9.49)

* Rejection threshold for Chi-Square Test

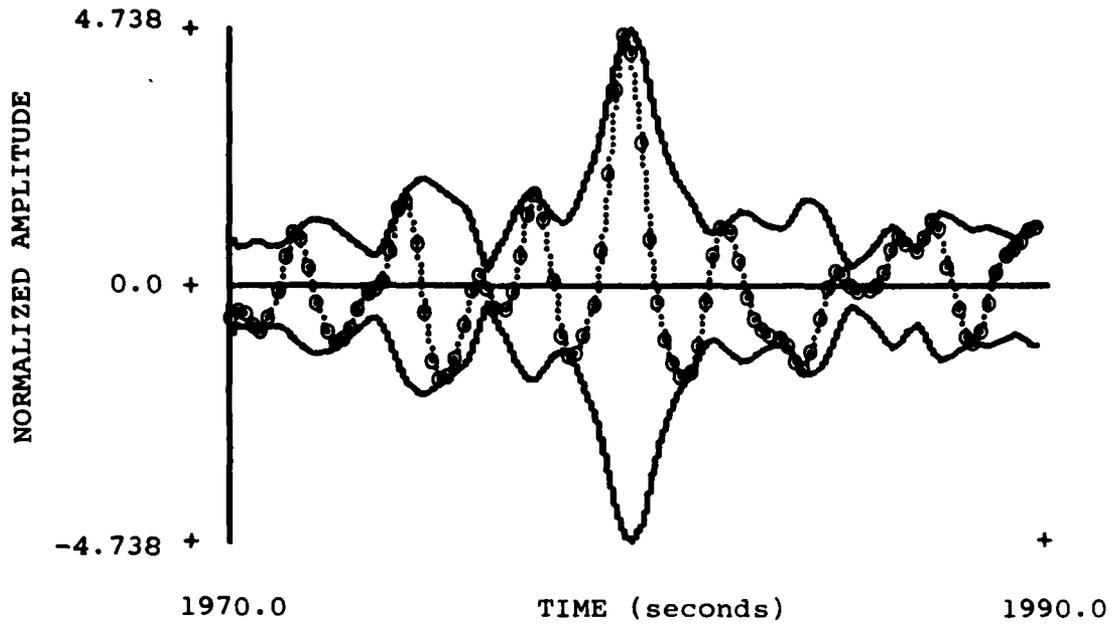


Figure C.1 - Time History Segment from Run 2410, Channel 1

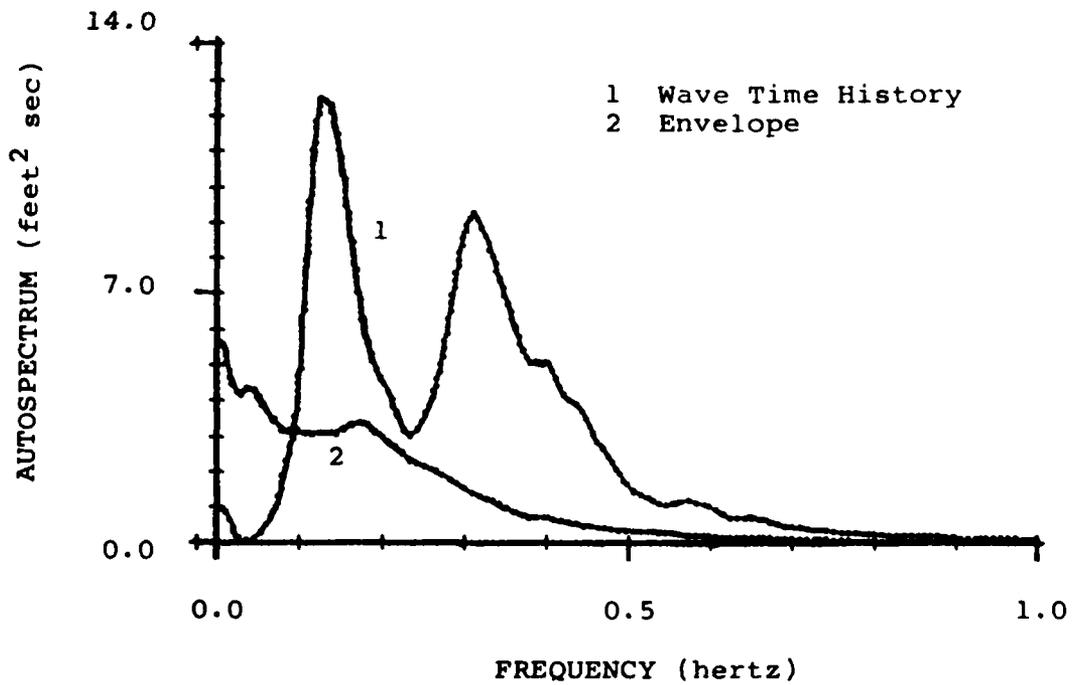


Figure C.2 - Autospectra of Time History and Envelope for Run 2410, Channel 1

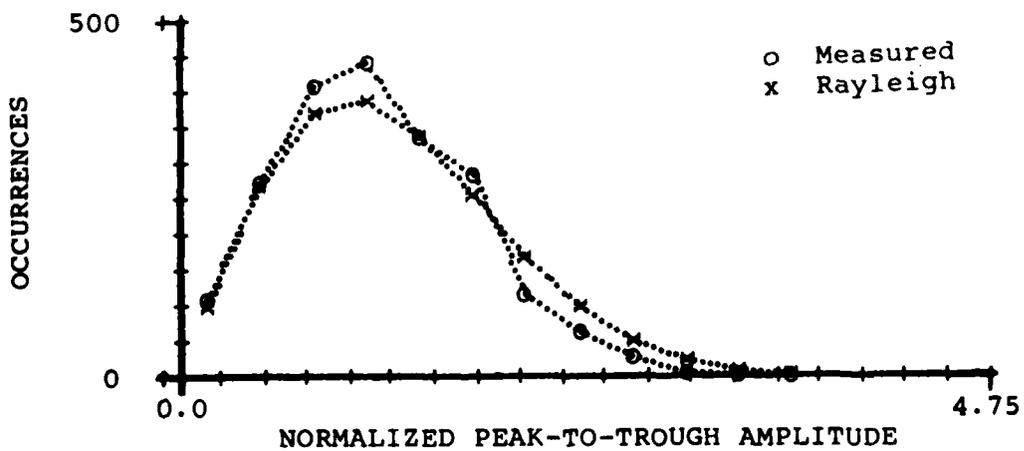
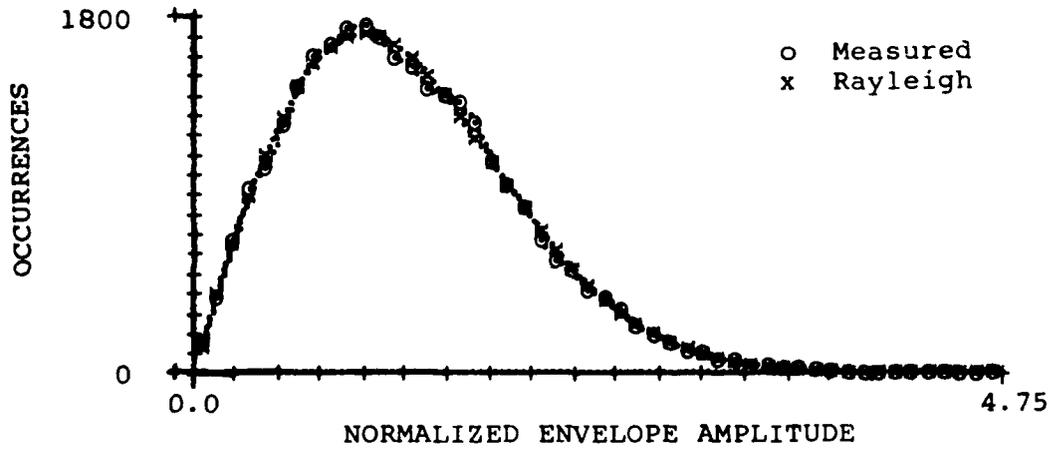
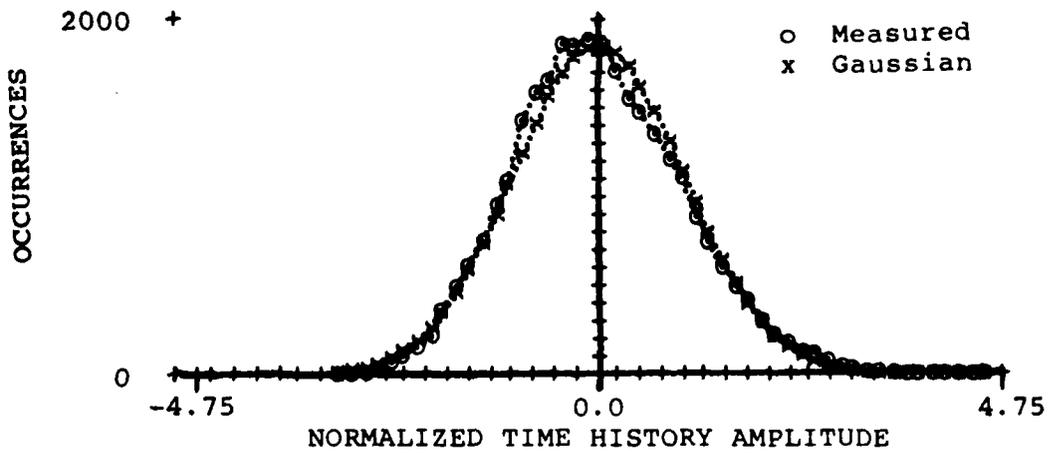


Figure C.3 - Histograms for Run 2410, Channel 1

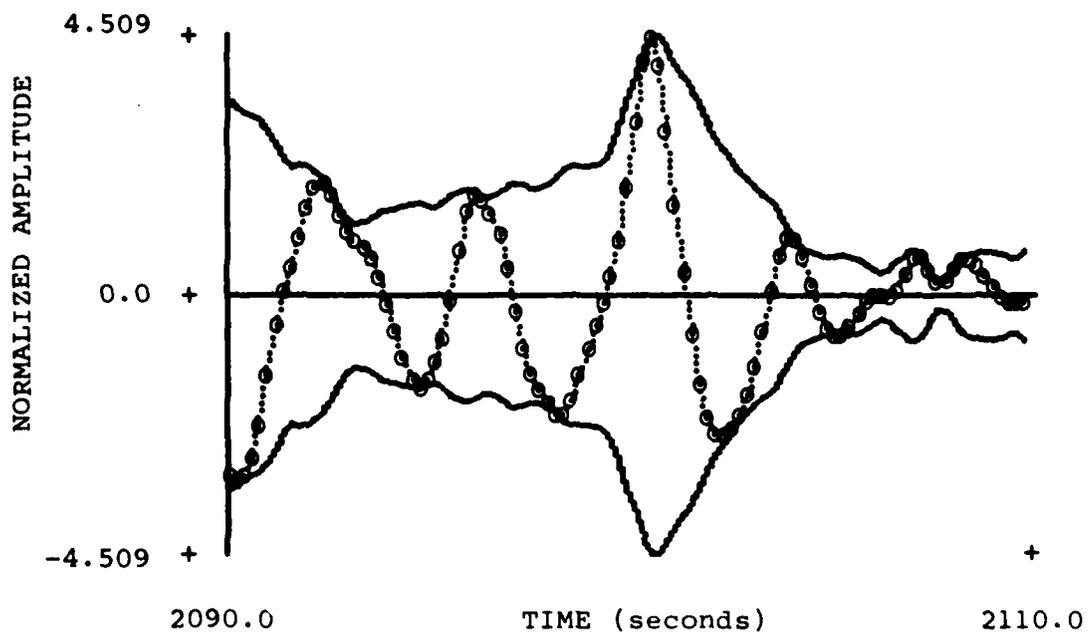


Figure C.4 - Time History Segment from Run 2410, Channel 2

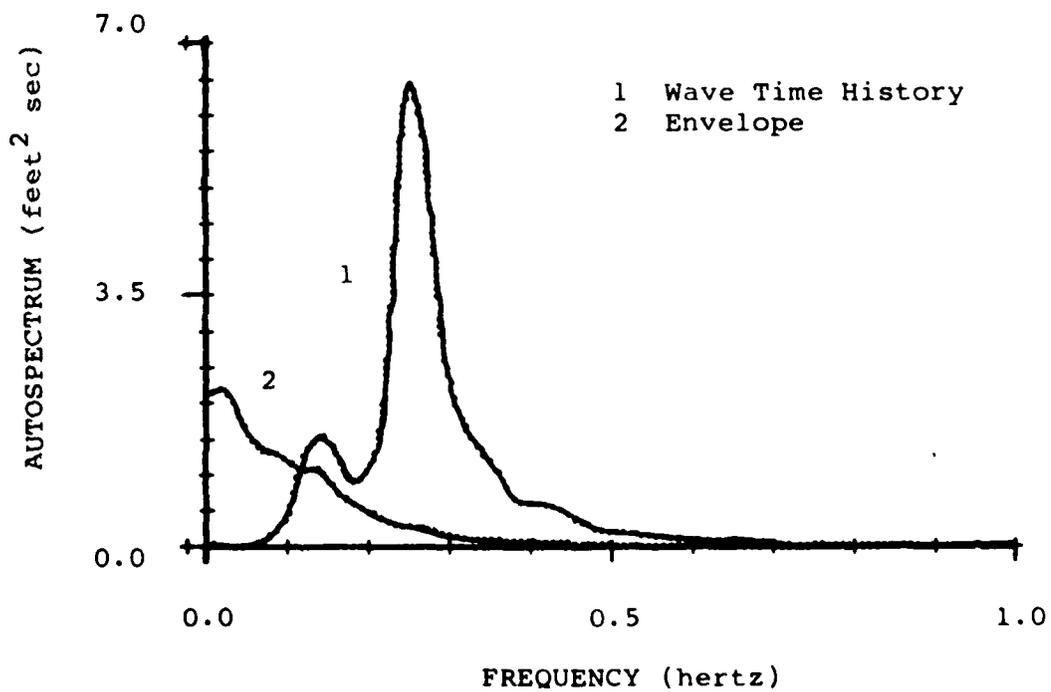


Figure C.5 - Autospectra of Time History and Envelope for Run 2410, Channel 2

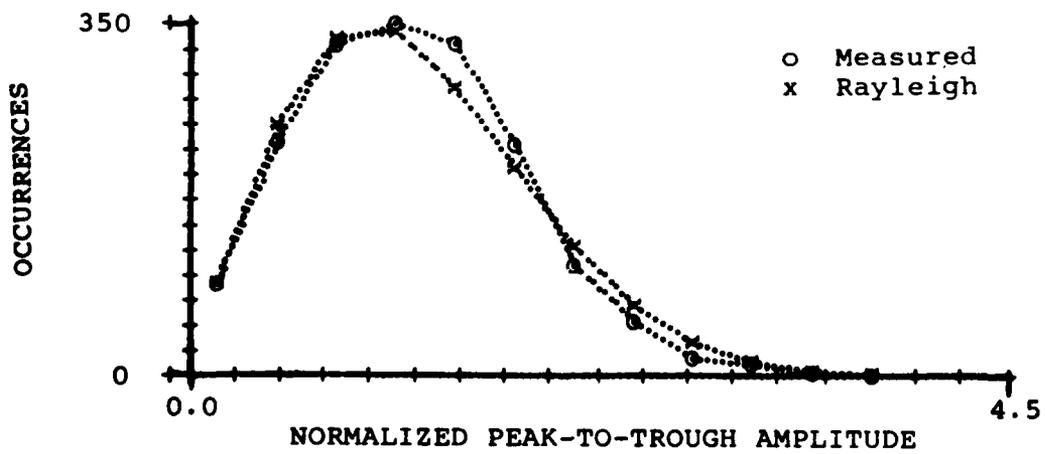
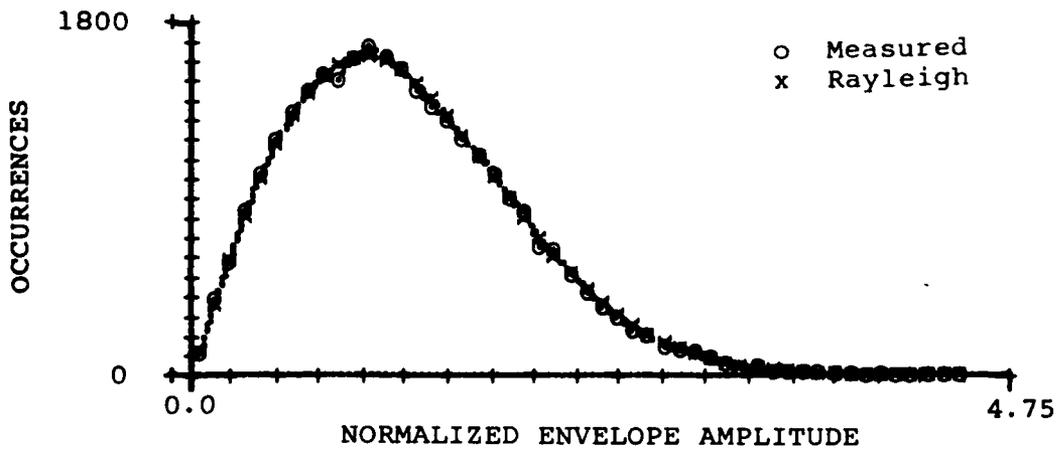
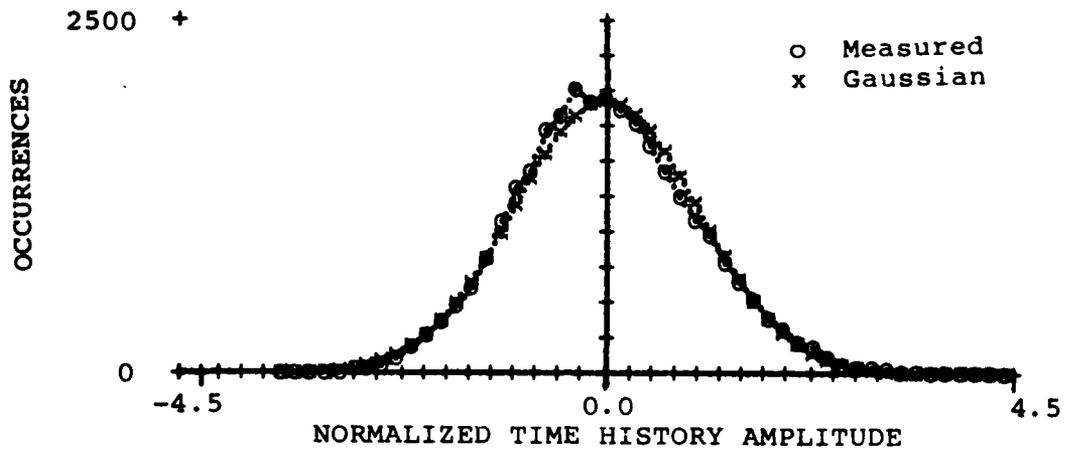


Figure C.6 - Histograms for Run 2410, Channel 2

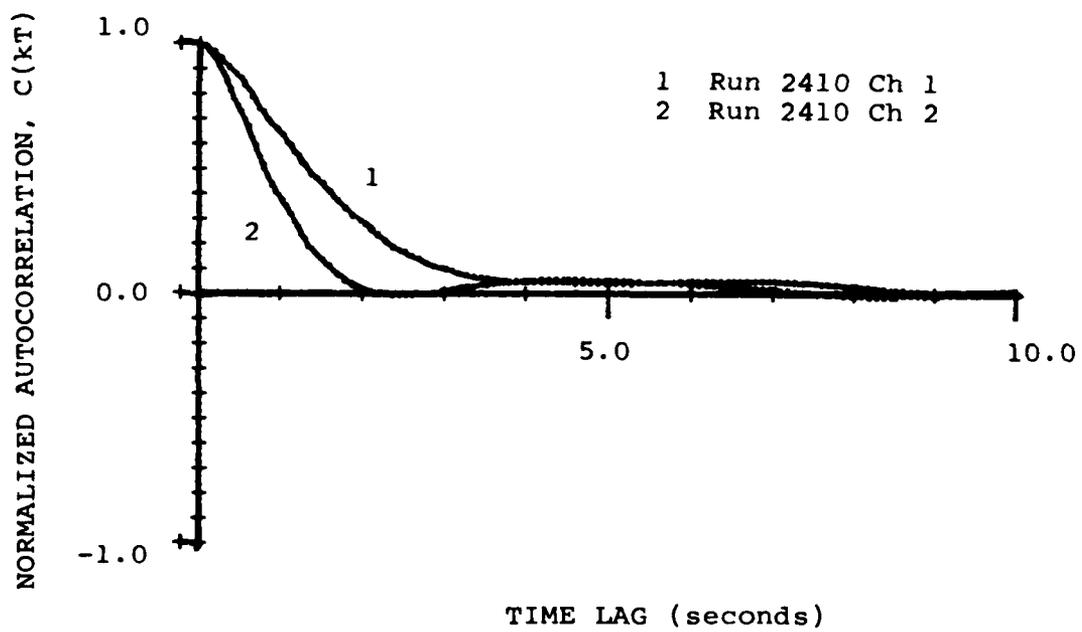


Figure C.7 - Autocorrelation Functions for Envelopes of Run 2410, Channels 1 and 2

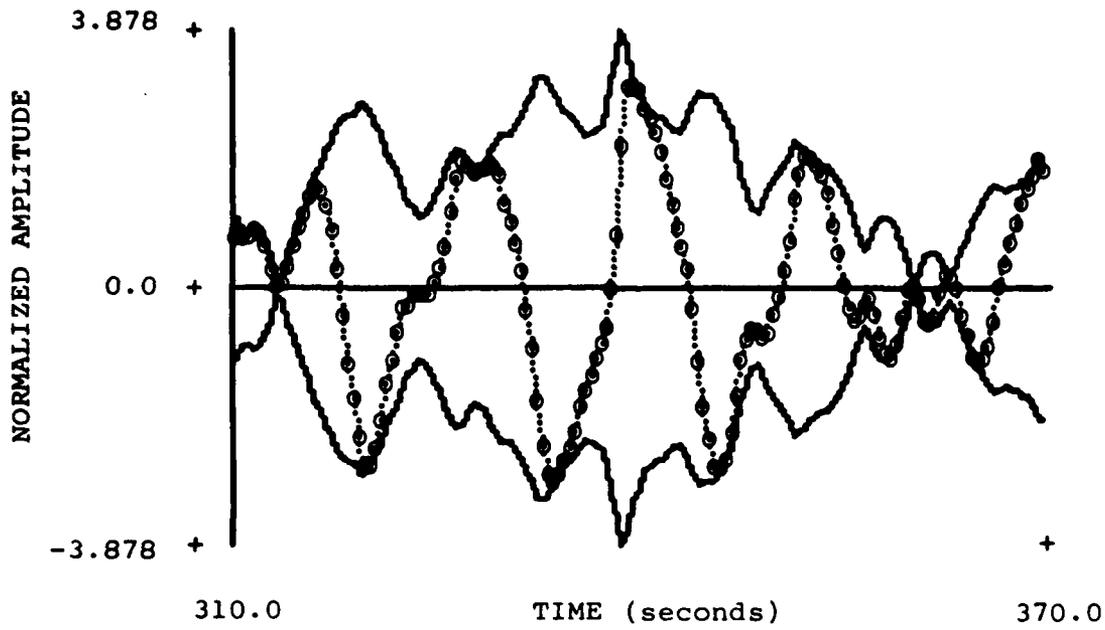


Figure C.8 - Time History Segment from Run 2721

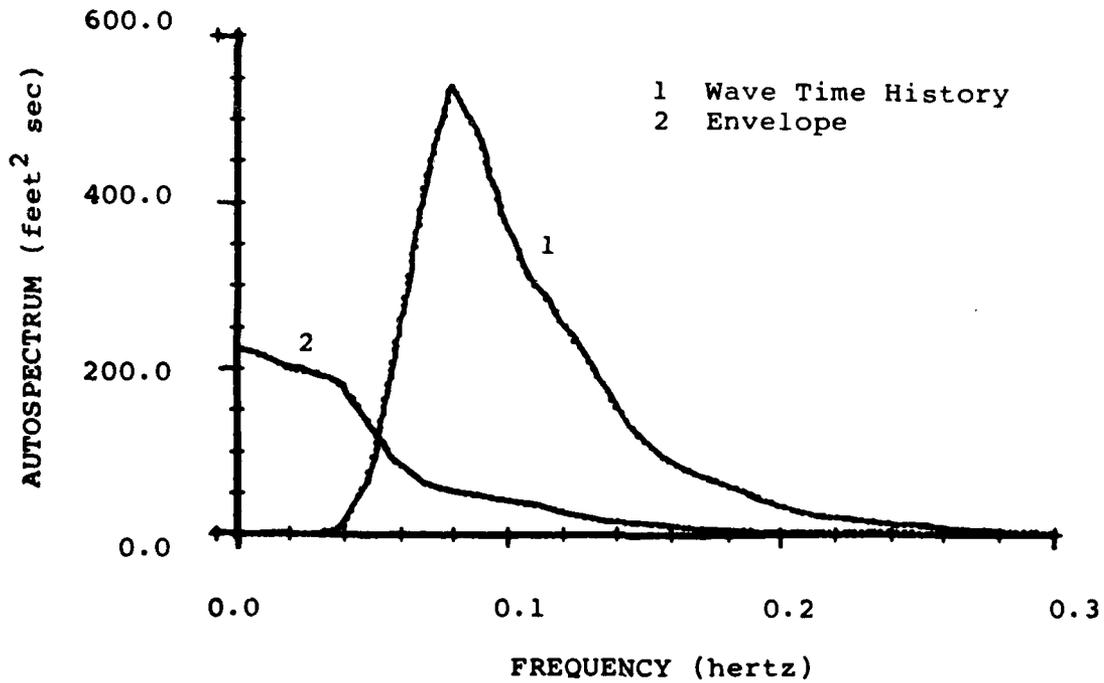


Figure C.9 - Autospectra of Time History and Envelope for Run 2721

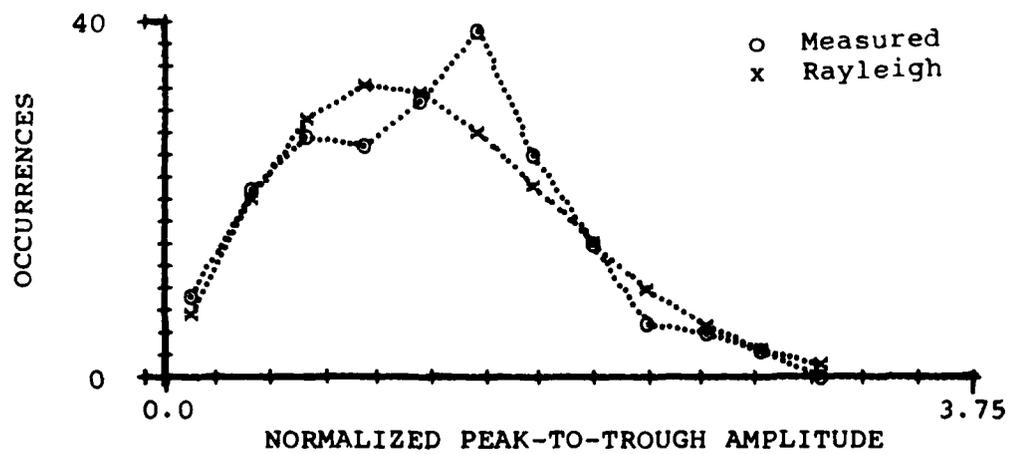
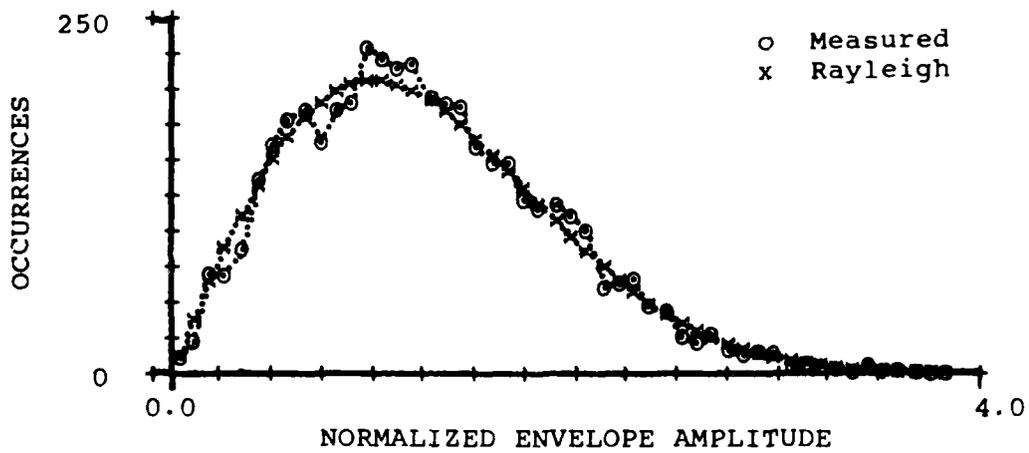
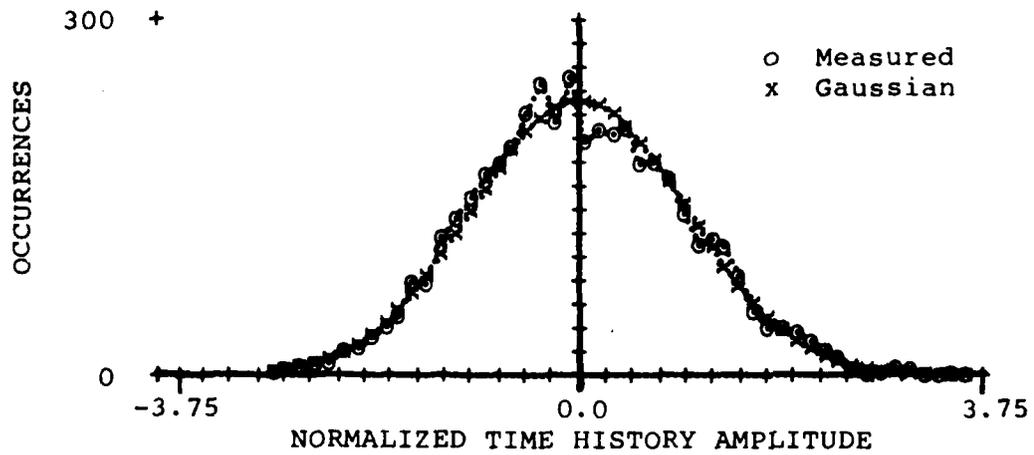


Figure C.10 - Histograms for Run 2721

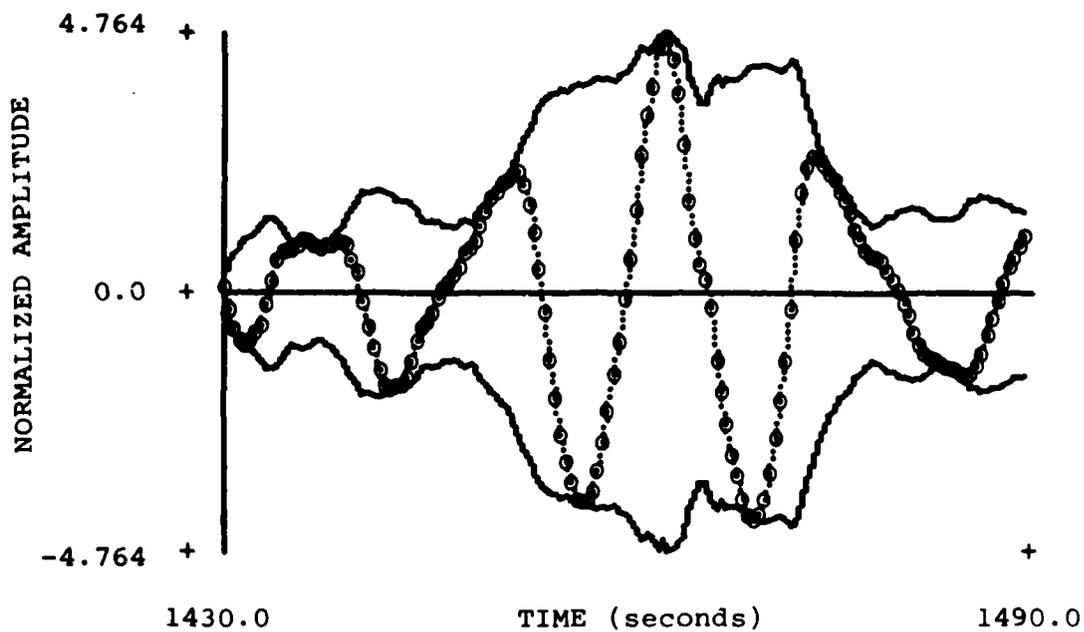


Figure C.11 - Time History Segment from Run 2724

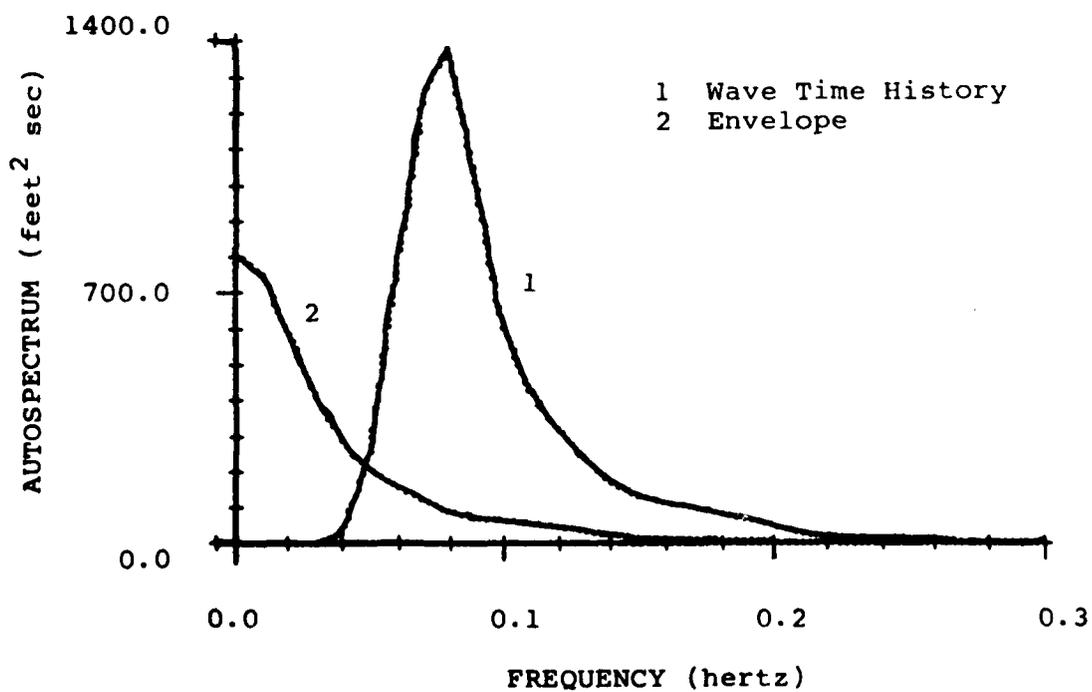


Figure C.12 - Autospectra of Time History and Envelope for Run 2724

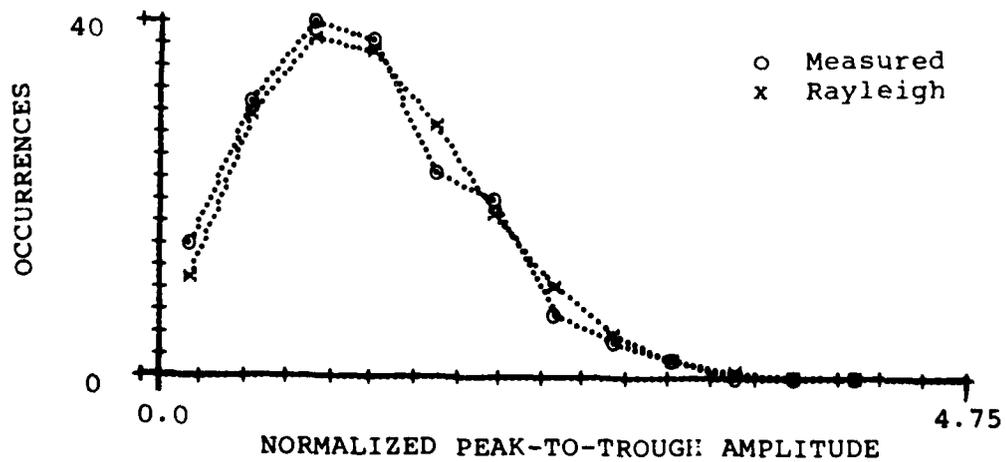
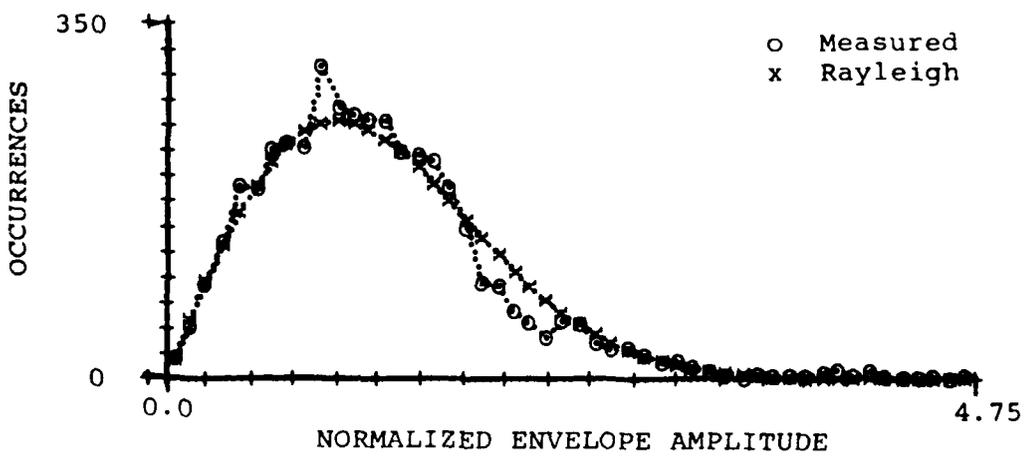
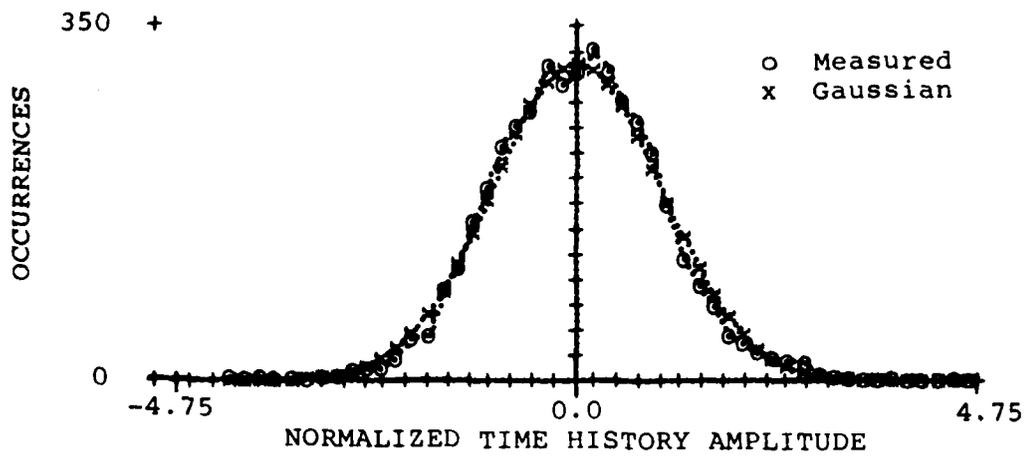


Figure C.13 - Histograms for Run 2724

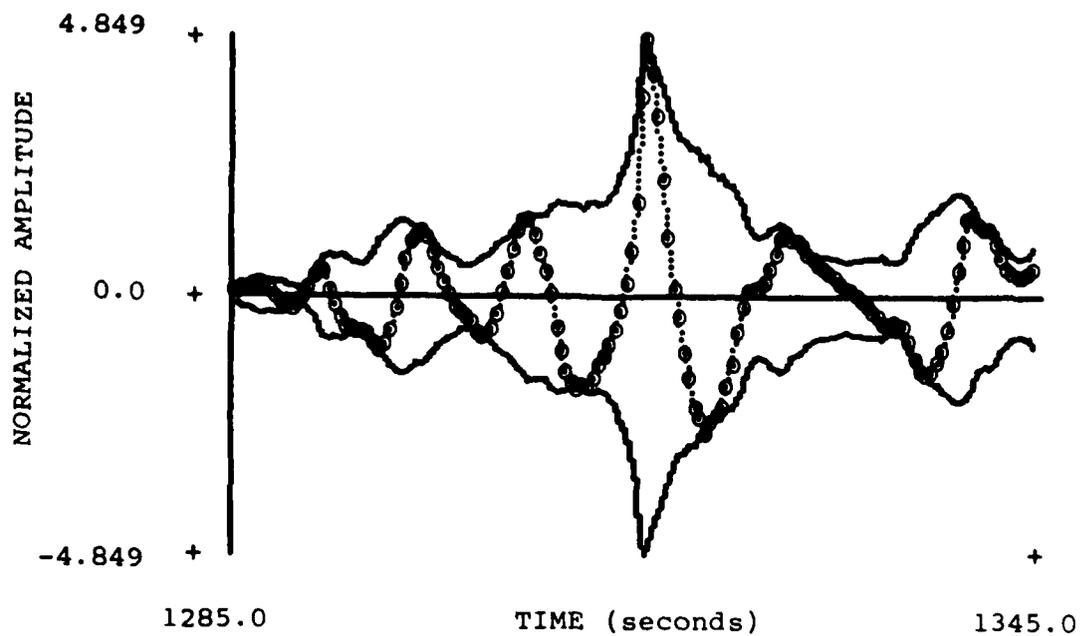


Figure C.14 - Time History Segment from Run 2731

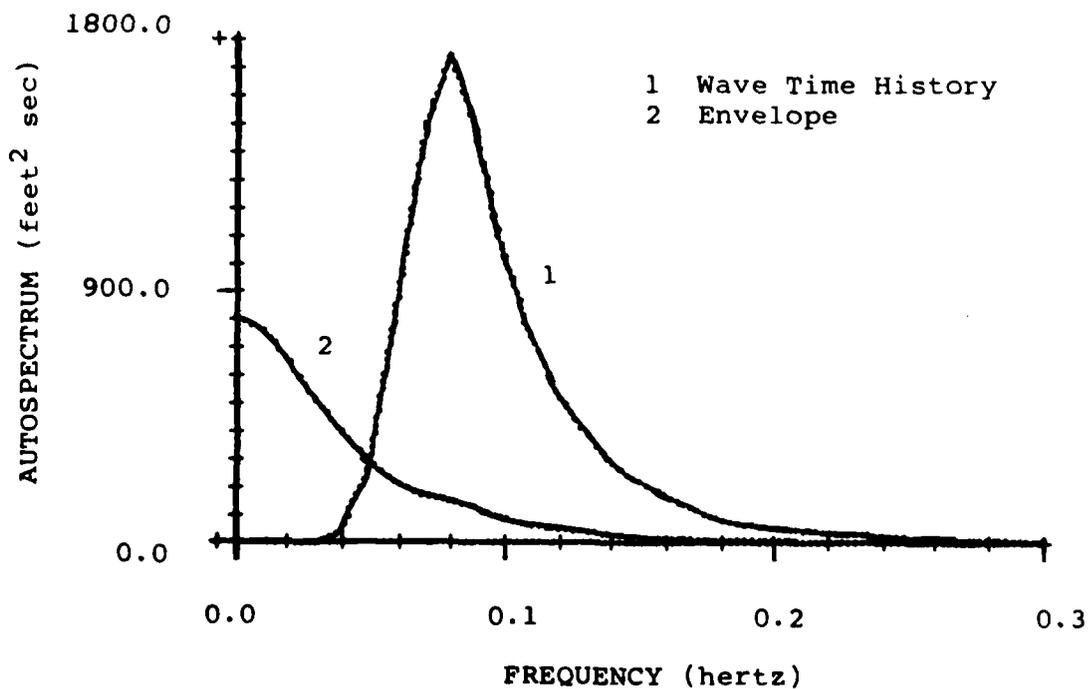


Figure C.15 - Autospectra of Time History and Envelope for Run 2731

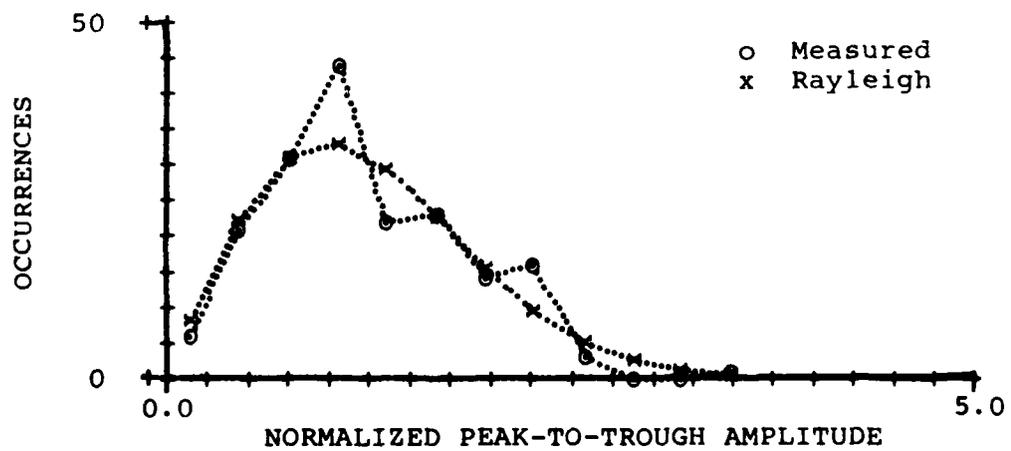
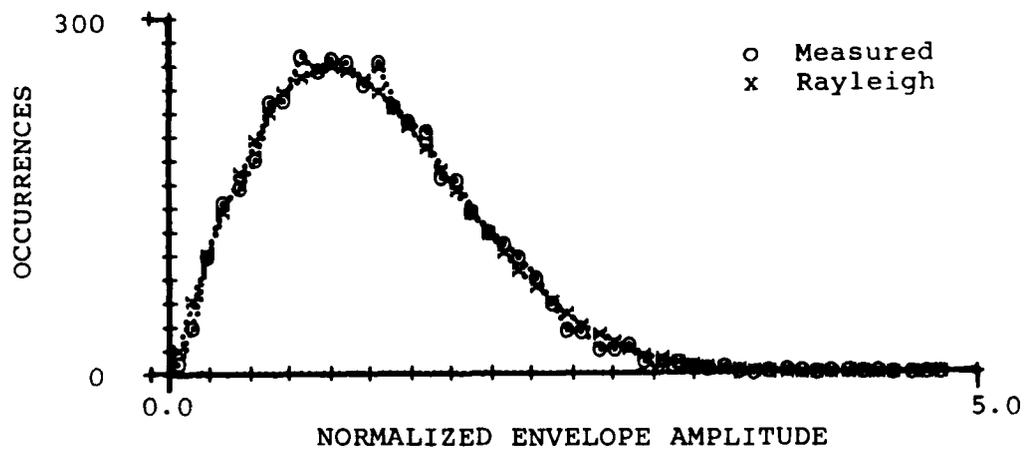
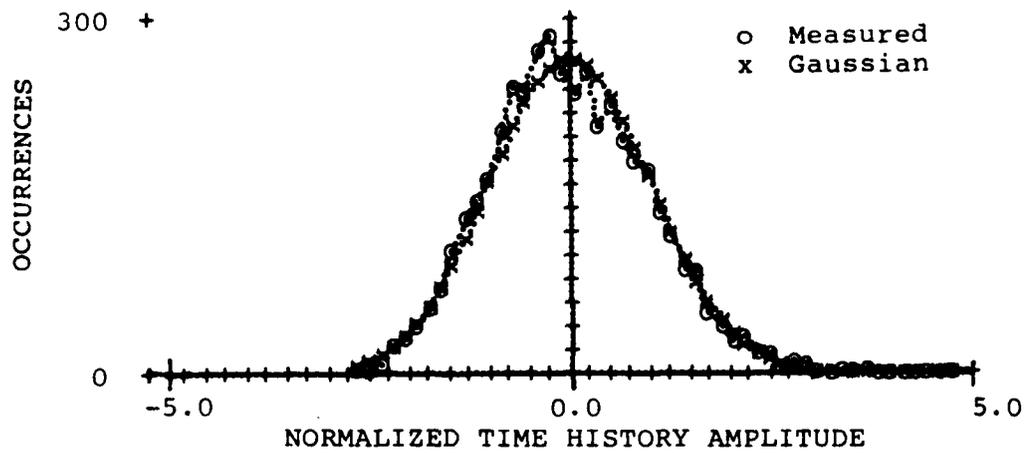


Figure C.16 - Histograms for Run 2731

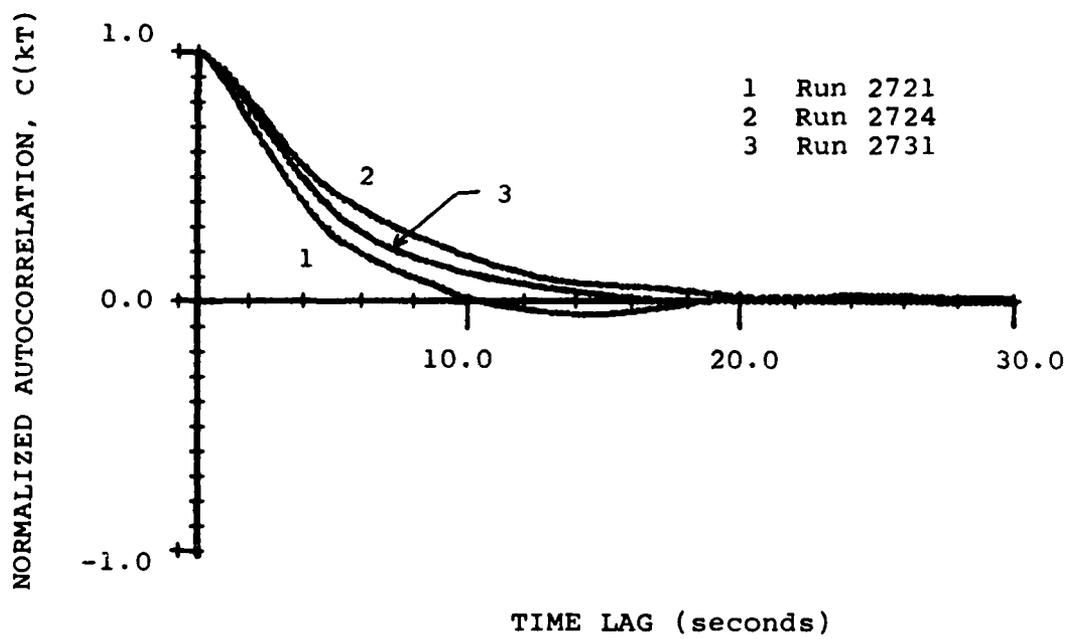


Figure C.17 - Autocorrelation Functions for Envelopes of
Runs 2721, 2724 and 2731

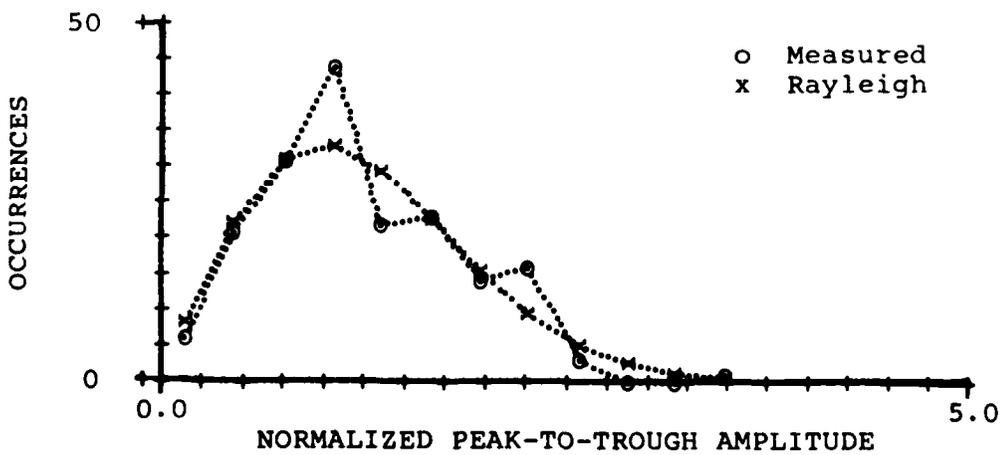
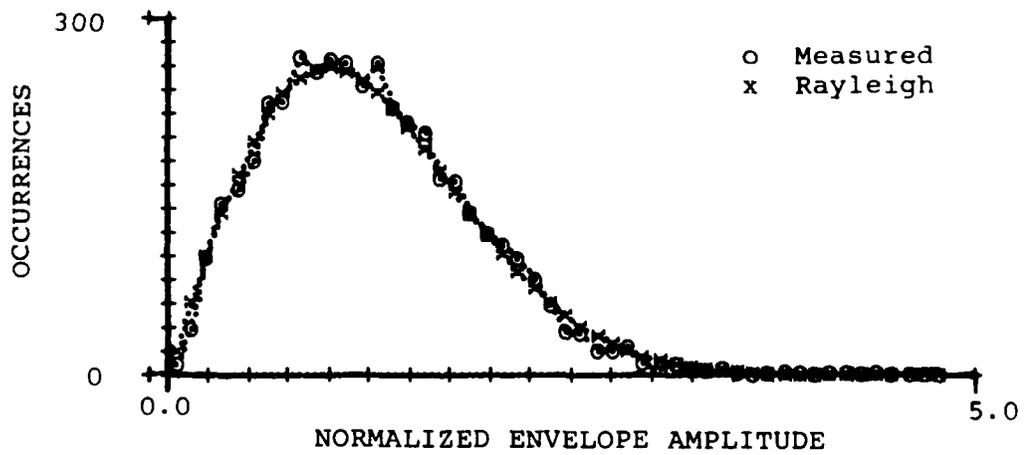
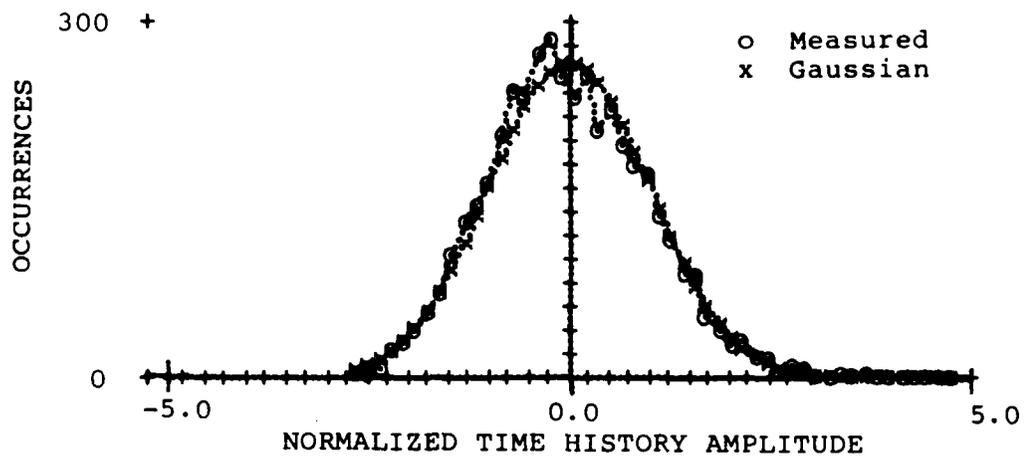


Figure C.16 - Histograms for Run 2731

REFERENCES

1. Longuet-Higgins, M.S., "On the Statistical Distribution of the Heights of Sea Waves," *Journal of Marine Research*, Vol. 11, No. 3, p. 245 (1952).
2. Cartwright, D.E. and Longuet-Higgins, M.S., "The Statistical Distribution of the Maxima of a Random Process," *Proc. Roy. Soc. London, Ser. A*, 237, p. 212 (1956).
3. Ochi, M.K., "On Prediction of Extreme Values," *Journal of Ship Research*, Vol. 17, No. 1, p. 29 (Mar 1973).
4. Rice, S.O., "Mathematical Analysis of Random Noise," *Bell System Technical Journal*, Vol. 23 (1944) and Vol. 24 (1945).
5. Tayfun, M.A., "Distribution of Crest-to-Trough Wave Heights," *J. Wat. Ways Port Coastal Ocean Div. Am. Soc. Civ. Engrs.*, 107(WW3), p. 149 (Aug 1981).
6. Ochi, M.K., "Generalization of Rayleigh Probability Distribution and Its Application," *Journal of Ship Research*, Vol. 22, No. 4, p. 259 (1978).
7. Forristall, G.Z., "On the Statistical Distribution of Wave Heights in a Storm," *Journal of Geophysical Research*, Vol. 83, No. C5, p. 2353 (1978).
8. Epstein, B., "The Distribution of Extreme Values in Sample whose Members are Subject to a Markov Chain Condition," *Annals of Mathematical Statistics*, Vol. 20, p. 590 (1949).
9. Ochi, M.K., "Extreme Values of Waves and Ship Responses Subject to the Markov Chain Condition," *Journal of Ship Research*, Vol. 23, No. 3, p. 188 (Sept 1979).
10. Naess, A., "Extreme Values of a Stochastic Process Whose Peak Values are Subject to the Markov Chain Condition," *Norwegian Maritime Research*, No. 1, Vol. 11, p. 16 (1983).
11. Naess, A., "Extreme Value Estimates Based on the Envelope Concept," *Applied Ocean Research*, Vol. 4, No. 3, p. 181 (1982).
12. Middleton, D., An Introduction to Statistical Communication Theory, McGraw Hill, N.Y., N.Y. (1960).
13. Arens, R., "Complex Processes for Envelopes of Normal Noise," *IRE Transactions on Information Theory*, Vol. IT-3, p. 204 (Sept 1957).
14. Dugundji, J., "Envelopes and Pre-Envelopes of Real Waveforms," *IRE Transactions on Information Theory*, Vol. IT-4, p. 53 (Mar 1958).
15. Rice, S.O., "Envelopes of Narrow-Banded Signals," *Proceedings of the IEEE*, Vol. 70, No. 7, p. 692 (July 1982).

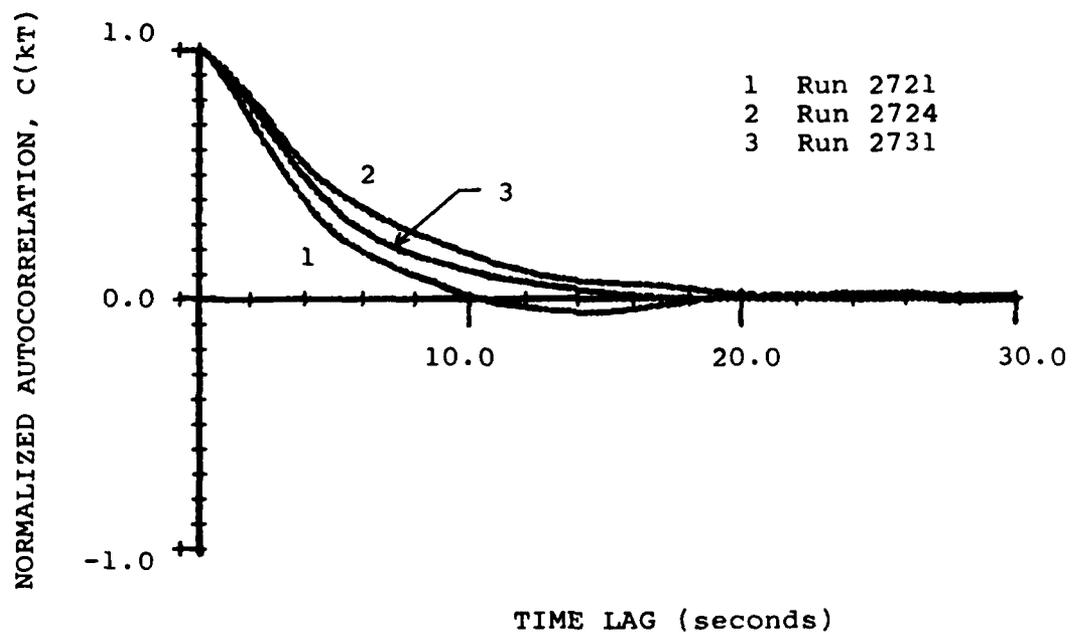


Figure C.17 - Autocorrelation Functions for Envelopes of
Runs 2721, 2724 and 2731

REFERENCES (Cont.)

16. Rabiner, L.R. and Schafer, R.W., "On the Behavior of Minimax FIR Digital Hilbert Transformers," *The Bell System Technical Journal*, Vol. 53, No. 2, p. 363 (Feb 1974).
17. Bendat, J. and A. Piersol, Random Data: Analysis and Measurement Procedures, Wiley-Interscience, N.Y., N.Y. (1971).
18. Helms, H.D., "Nonrecursive Digital Filters: Design Methods for achieving Specifications on Frequency Response," *IEEE Trans. Audio and Electroacoustics*, Vol. AU-16, No. 3, p. 336 (Sept 1969).
19. Balakrishnan, A.V., "A Note on the Sampling Principle for Continuous Signals," *IRE Trans. Information Theory*, vol. IT-3, p. 143 (June 1957).
20. Earle, M.D., "Extreme Wave Conditions During Hurricane Camille," *Journal of Geophysical Research*, Vol. 80, No. 3, p.377 (Jan 1975).

DTNSRDC ISSUES THREE TYPES OF REPORTS

1. DTNSRDC REPORTS, A FORMAL SERIES, CONTAIN INFORMATION OF PERMANENT TECHNICAL VALUE. THEY CARRY A CONSECUTIVE NUMERICAL IDENTIFICATION REGARDLESS OF THEIR CLASSIFICATION OR THE ORIGINATING DEPARTMENT.

2. DEPARTMENTAL REPORTS, A SEMIFORMAL SERIES, CONTAIN INFORMATION OF A PRELIMINARY, TEMPORARY, OR PROPRIETARY NATURE OR OF LIMITED INTEREST OR SIGNIFICANCE. THEY CARRY A DEPARTMENTAL ALPHANUMERICAL IDENTIFICATION.

3. TECHNICAL MEMORANDA, AN INFORMAL SERIES, CONTAIN TECHNICAL DOCUMENTATION OF LIMITED USE AND INTEREST. THEY ARE PRIMARILY WORKING PAPERS INTENDED FOR INTERNAL USE. THEY CARRY AN IDENTIFYING NUMBER WHICH INDICATES THEIR TYPE AND THE NUMERICAL CODE OF THE ORIGINATING DEPARTMENT. ANY DISTRIBUTION OUTSIDE DTNSRDC MUST BE APPROVED BY THE HEAD OF THE ORIGINATING DEPARTMENT ON A CASE-BY-CASE BASIS.