

UNCLASSIFIED

AD 4 5 3 9 8 7

DEFENSE DOCUMENTATION CENTER

FOR

SCIENTIFIC AND TECHNICAL INFORMATION

CAMERON STATION ALEXANDRIA, VIRGINIA



UNCLASSIFIED

NOTICE: When government or other drawings, specifications or other data are used for any purpose other than in connection with a definitely related government procurement operation, the U. S. Government thereby incurs no responsibility, nor any obligation whatsoever; and the fact that the Government may have formulated, furnished, or in any way supplied the said drawings, specifications, or other data is not to be regarded by implication or otherwise as in any manner licensing the holder or any other person or corporation, or conveying any rights or permission to manufacture, use or sell any patented invention that may in any way be related thereto.

CATALOGED BY DDC

AS F.U. NO. 453987

453987

NEL REPORT 1243

4 5 3 9 8 7

**PROBABILITY DENSITY ANALYSIS
OF OCEAN AMBIENT AND SHIP NOISE**

M. A. Calderon • Research and Development Report 1248 • 5 November 1964

U. S. NAVY ELECTRONICS LABORATORY, SAN DIEGO, CALIFORNIA 92152 • A BUREAU OF SHIPS LABORATORY

DDC
JAN 11 1965

QUALIFIED REQUESTERS MAY OBTAIN COPIES
OF THIS REPORT FROM DDC.

THE PROBLEM

Investigate the probability density distribution of the amplitude of ocean ambient noise and ship noise; determine any differences in the distributions which might lead to the identification of ship noise masked by a high background noise level. Also, determine, by standard statistical methods, whether the distributions are gaussian or non-gaussian.

RESULTS

1. Ambient ocean noise was found to have a gaussian distribution of amplitudes (in the sense that the moments of the distribution satisfied specific tests) only when the ambient noise was relatively clean, i. e., the noise did not contain high-level ship noise, biological noise, ice noise or any of the other extraneous noises discussed in the text.

2. The group of ship-noise samples recorded at close range contained a large number of samples that had a non-gaussian distribution. However the other types of extraneous noises were found to cause the same kind of deviation from a gaussian distribution, so that it was not possible by these tests to distinguish between a sample with ship noise and a sample with the other types of extraneous noises (such as biological and ice noise).

RECOMMENDATIONS

1. Use the method of moments described here if better accuracy than that given by overlays is desired to estimate the moments and to determine whether a sample is gaussian or non-gaussian.
2. In the probability density analysis of a noise sample, use a range of amplitudes covering at least ± 4 standard deviations; otherwise large errors in the estimates of the moments will frequently result.
3. In future applications of the PDA, have the output of the PDA in a digital form rather than a continuous curve, so that the data will be available in a form more suitable for the calculation of the moments of the distribution.

ADMINISTRATIVE INFORMATION

Work was performed under SR 004 03 01, Task 8119 (NEL L2-4) by members of the Listening Division. The report covers work from January 1962 to June 1963 and was approved for publication 5 November 1964.

The author wishes to express appreciation to the members of the Listening Division who contributed their time to perform much of the data processing, W. P. de la Houssaye who wrote the computer program; and Elaine Kyle who prepared the data for the computer. Thanks are also extended to Fred Dickson, who prepared the illustrations, and to G. M. Wenz, who made many helpful suggestions during the work phase and during the writing of the manuscript.

CONTENTS

INTRODUCTION...	page 5
TEST PROGRAM...	7
Instrumentation...	7
Research Techniques...	8
Data Reduction Techniques...	14
RESULTS...	30
CONCLUSIONS...	37
PD of Ambient Ocean Noise...	37
PD of Ship Noise...	37
Comparison of Test Methods...	37
RECOMMENDATIONS...	38
REFERENCES...	39
APPENDIX A: DESCRIPTION OF THE PDA AND ITS OPERATION...	41-42
APPENDIX B: DETERMINATION OF NUMBER OF DATA POINTS OF EACH SAMPLE...	43-44

TABLES

1	Noise Samples Selected for Analysis, by Location...	page 11
2	Number of Curves Showing Significant Values of Skewness and Kurtosis...	31
3	Locations of Curves Showing Significant Values of Skewness and Kurtosis...	32
4	Locations of Curves Showing Skewness and Kurtosis at 1 Per Cent Probability Level...	33
5	List of Curves for Which Chi-square Was Computed...	35
6	Curves Chosen by Normal Curve Overlay as Being Very Closely Gaussian...	36

ILLUSTRATIONS

- 1 Curve of probability density function of a gaussian random variable... *page 6*
- 2 Block diagram of Probability Density Analyzer system... 7
- 3-4 Selected probability density curves, compared with a normal curve... *12-13*
- 5-6 Examples of PD curves obtained by use of overlay method... *15-17*
- 7 Curve obtained with Probability Density Analyzer... *23*
- 8 Normalized cumulative sums of tabulated values... *25*
- 9 Curves with positive and negative skewness, computed with Edgeworth's series... *26*
- 10 Curves with positive and negative kurtosis... *27*
- 11-12 Cumulative probability of noise samples shown in figures 3 and 4... *28-29*
- 13 Experimental PD curve of random noise showing skewness, calculated by Edgeworth's series... *34*
- A1 Theoretical and experimental PD curves of square wave and sine wave... *41*
- B1 Number of times random noise goes into an interval $\mu \pm 0$ for PD of 0.4, vs. cutoff frequency of low-pass filter... *45*

INTRODUCTION

The study reported here was undertaken to investigate the probability density distribution of the amplitudes of ocean ambient noise and ship noise with respect to various bandwidths in several frequency ranges. The question to be answered was whether ambient noise, without any ship noise or biological noises, can be considered gaussian, and whether the presence of ship noise significantly changes the probability of density distributions. A secondary objective was to investigate methods of data reduction of the probability density curves obtained with the B & K Probability Density Analyzer, using standard statistical tests.

The probability density function, as treated throughout this report, may be defined as follows.

$$p(x) = \lim_{\substack{\Delta x \rightarrow 0 \\ N \rightarrow \infty}} \frac{n_t / \Delta x_t}{N} \quad (1)$$

where x is a random variable, with its range of values divided into a large number of continuous intervals Δx . Measure its instantaneous value a great number of times N . Let n_t be the number of measured values of x in the t th interval (Δx_t).

The above equation can be rewritten as

$$p(x) = \lim_{\substack{\Delta x \rightarrow 0 \\ T \rightarrow \infty}} \frac{\Delta t_t \cdot \Delta x_t}{T} \quad (2)$$

where Δt_t is the amount of time the signal spends in the interval Δx_t and T is the total time of the sample. Equation 2 indicates more clearly how the B & K PDA measures the probability density function. A more detailed explanation can be found in reference 1. (See list of references at end of report.)

The function

$$p(x) = \frac{1}{\sigma \sqrt{2\pi}} \exp \left[-(x-\bar{x})^2 / 2\sigma^2 \right]$$

where \bar{x} is the mean and σ is the standard deviation, is illustrated in figure 1.

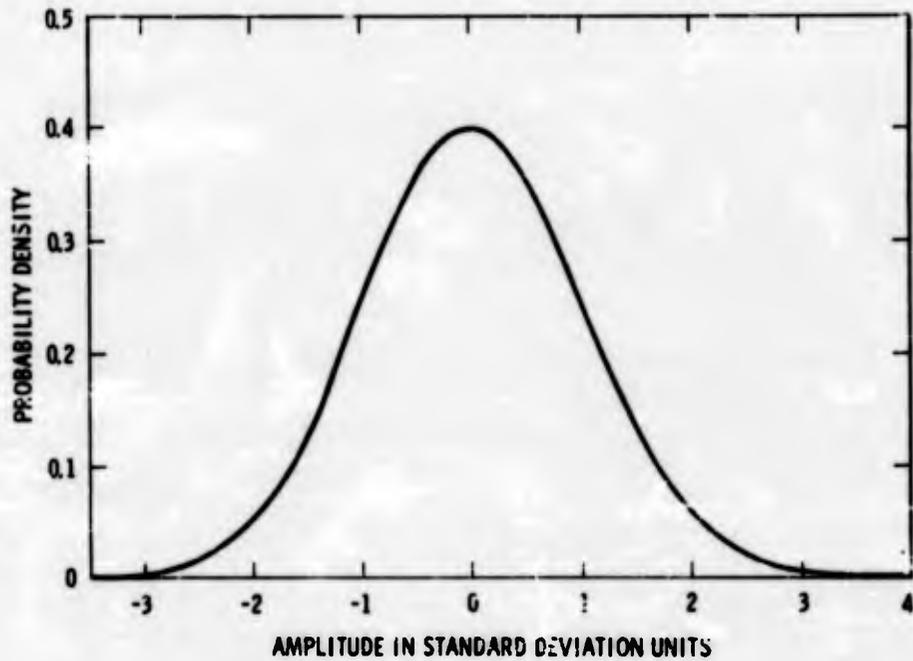


Figure 1. Curve of the probability density function of a gaussian random variable (normalized to a unit area under the curve).

TEST PROGRAM

Instrumentation

The equipment used for the investigation is described below and illustrated in figure 2.

An Ampex Model 350 was used as the record and playback recorder. This model has a good low-frequency response to below 20 c/s.

The filters following the recorder were an Allison Laboratories Model 2-A (used mainly as a low-pass filter) and a B & K Band Pass Filter Set, Type 1611.

A McIntosh amplifier, Model MC30, was used to raise the signal level to 1 volt rms or greater.

The B & K Probability Density Analyzer, Model 160 (to be referred to as the PDA) was the main piece of equipment and has been primarily designed to obtain the probability density curves of disturbances that are essentially random in character. A brief description of the PDA and its use in this investigation is given in Appendix A. A complete and detailed description of the PDA can be obtained from the instruction manual.¹

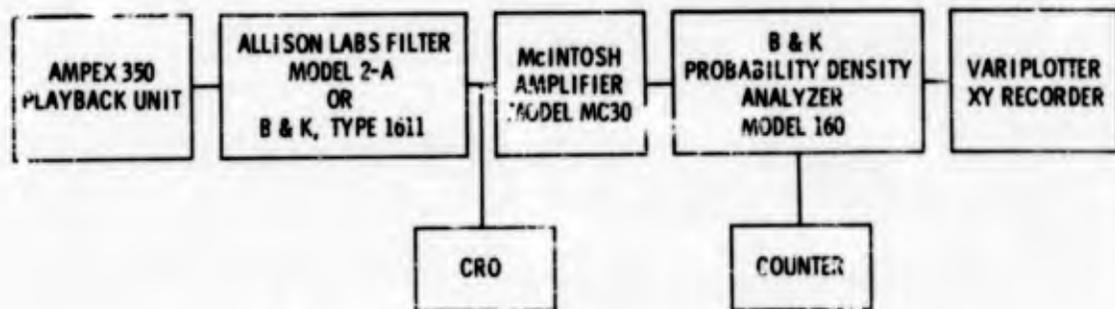


Figure 2. Block diagram of Probability Density Analysis system.

An XY recorder by Electronic Associates, Inc., was used to record the analog X and Y outputs of the PDA.

A cathode ray oscilloscope monitored the signal output of the filter.

The counter used responded to frequencies of at least 10 Mc/s for use with the PDA. The counter can be used in place of an XY recorder and, in fact, is essential if measurements are to be made at low probability densities.

Research Techniques

Data which had been recorded for previous ambient-noise studies were available for this study. These samples had been recorded on 10 $\frac{1}{2}$ -inch reels of $\frac{1}{4}$ -inch tape, at 3 $\frac{3}{4}$ inches per second, and were from three locations. Two groups had been made in shallow water -- one, about 2 miles from the western side of an island off the coast of Southern California, and the other in the Bering Straits. These consisted of ambient-noise samples recorded at regular intervals throughout the day, so that one reel covered data for one day. The third location represented was in deep water in the North Pacific between Hawaii and Alaska; most of these samples were of longer duration than the other two groups, but covered only a few days.

Samples of ship noise were desired, so that their probability density curves might be compared with those of "clean" ambient noise. Recordings were made of ships entering San Diego Harbor, with the sampling made at approximately the closest point of approach. These included Navy surface ships, submarines (surfaced), and commercial ships.

Several factors were considered in choosing the data samples to be used in this study.

1. "Clean" ambient noise was used to determine whether the distributions of the amplitudes were gaussian.

or near-gaussian according to certain tests which will be discussed later. Ambient noise was judged to be "clean" when it was free from ship noise, biological noises, or any man-made sounds when the sample was monitored. A band-pass filter and oscilloscope were used to determine whether 60-c/s hum or any other single frequency components were present in the noise sample.

2. All noise samples should be stationary for their entire length. When the sample is ambient ocean noise, this condition will not in general be true. For a noise sample to be stationary it is necessary for the sample parameters, the means and the variances, to remain unchanged as measured from samples taken at different times. It is possible that no significant difference in the sample parameters will be found if the time between samples is short enough. In a previous study² it was concluded that ocean noise is a slowly varying, not a stationary, process. This conclusion was based on a comparison of samples that were 3 or more minutes apart. However, no significant difference was found among the values of some other samples which were only 3 minutes or less apart. Thus it appears reasonable to assume that ocean noise is stationary during a short interval of time (less than 3 minutes).

3. The PDA requires a noise sample of about 30 minutes duration for a complete automatic analysis of the amplitudes from -3.00 to +3.00 standard deviations.

The need for a long noise sample that is stationary can be satisfied by recording a short noise sample on magnetic tape and then making a loop of the tape. A loop length was selected according to the following requirements.

a. The loop should be short enough so that the noise could be considered stationary and so that the entire loop could be analyzed for each amplitude interval. The PDA (in the particular position used) requires 30 seconds to sweep a range of amplitudes equal to the window width, which is 0.1 times the rms value of the input signal. A sample length of 7 seconds met all the above requirements

and this gives a loop size of 52.5 inches, which was conveniently handled.

b. The recorded noise on the loop should be continuous, i.e., there should be no blank intervals on the loop, since a blank interval would change the average rms value of the recorded noise.

A typical analysis procedure was as follows. A portion of data was selected for analysis from the recorded data available. The noise was re-recorded on a loop. The loop was played back at $7\frac{1}{2}$ ips and the analysis proceeded as indicated by the diagram in figure 2. The filter was set to the desired bandwidth, and the noise was amplified to 1 volt rms or greater. The PDA was carefully calibrated and adjusted just before each analysis. Its input level of noise was adjusted to 1 volt rms by its potentiometer, thus normalizing its output.

Probability density of the amplitudes was recorded on the Y scale of the XY recorder and the amplitude around which the probability density was measured was on the X scale. Some factors were selected to give a deflection of 4 inches on the Y scale for a probability density range of 0 to 0.4, and a deflection of 1 inch per standard deviation of amplitude on the X scale. The automatic sweep time of the PDA was set at $X = -3.00$ standard deviations, and would automatically sweep through to $X = +3.00$ standard deviations, based on a 1-volt rms input. Total running time was about 30 minutes. This procedure was repeated for each bandwidth on every loop analyzed.

Table 1 lists the number of samples analyzed from each location, the total number of probability density curves obtained from the samples, and the filter used to analyze these curves. When the Allison Laboratories filter was used, the system cutoff frequency at the low end was about 20 c/s and the upper cutoff frequency was determined by the filter which was set at 2500, 1500, 1200, 600, 300, or 150 c/s. The B & K filter was used in both the octave

and third-octave positions for center band frequencies of 100, 200, 400, 800, and 1600 c/s.

TABLE 1. NOISE SAMPLES SELECTED FOR ANALYSES. BY LOCATION.
(FOR THE BANDWIDTHS USED. SEE ABOVE)

LOCATION	NUMBER OF NOISE SAMPLES	NUMBER OF f_d CURVES OBTAINED	FILTER USED FOR ANALYSIS OF DATA
SOUTHERN CALIFORNIA	9	29	8 SAMPLES WITH ALLISON LABS FILTER; 1 SAMPLE WITH ALLISON LABS AND B & K
BERING STRAITS	9	24	ALLISON LABS
NORTH PACIFIC	8	65	B & K
SAN DIEGO (SHIP SE IN HARBOR)	9	36	ALLISON LABS

Actual probability density curves of ambient noise are shown in figures 3 and 4. The large fluctuations in some of the traces are caused by substantial variations in the level of the noise sample. Since some of the curves appeared to be closely gaussian, the methods used to measure the parameters of the distribution included overlays, calculated moments, and cumulative probability graphs. Tests of significance and the chi-square "goodness of fit" tests were used to determine what values of skewness and kurtosis were improbable at a 5 or 1 per cent probability level.

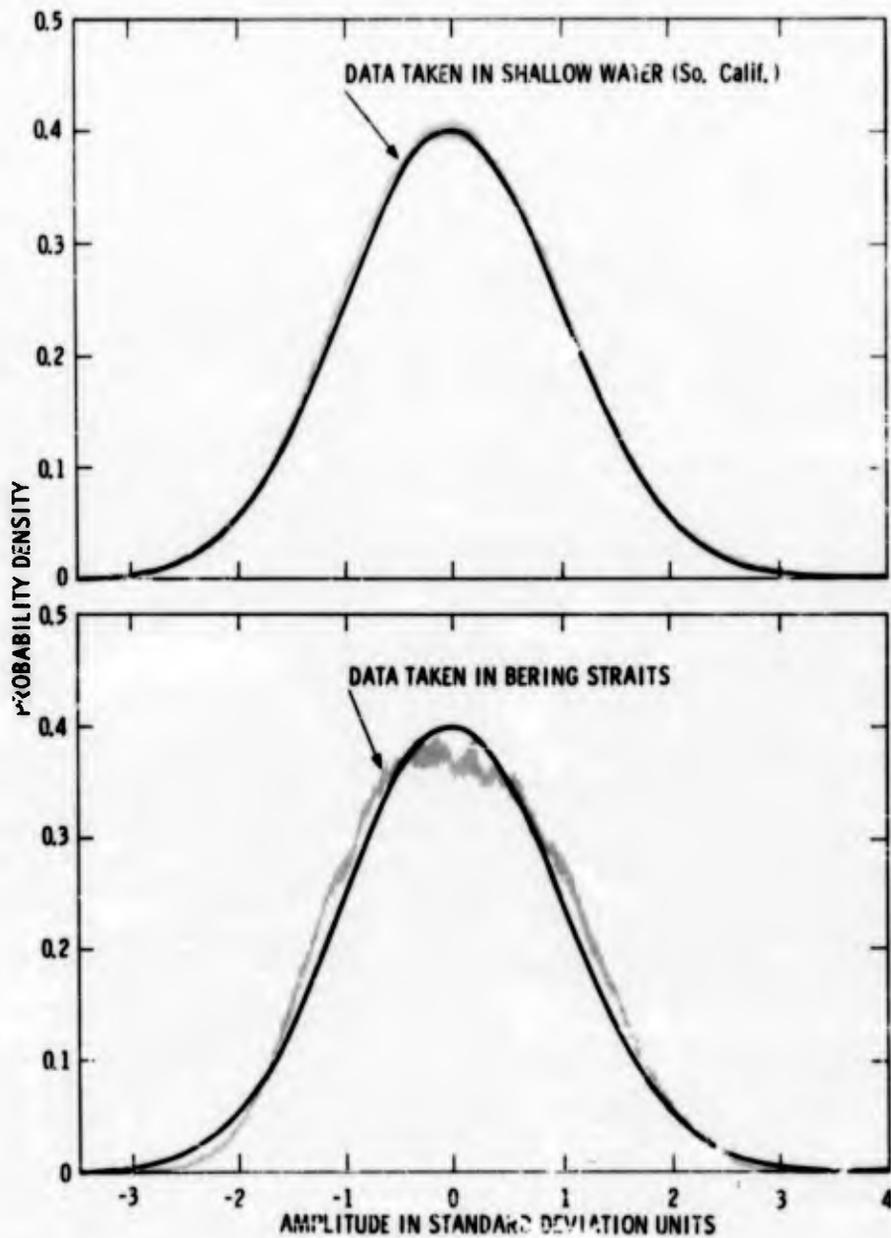


Figure 3. Example of some PD curves taken in shallow water, compared with a normal curve.

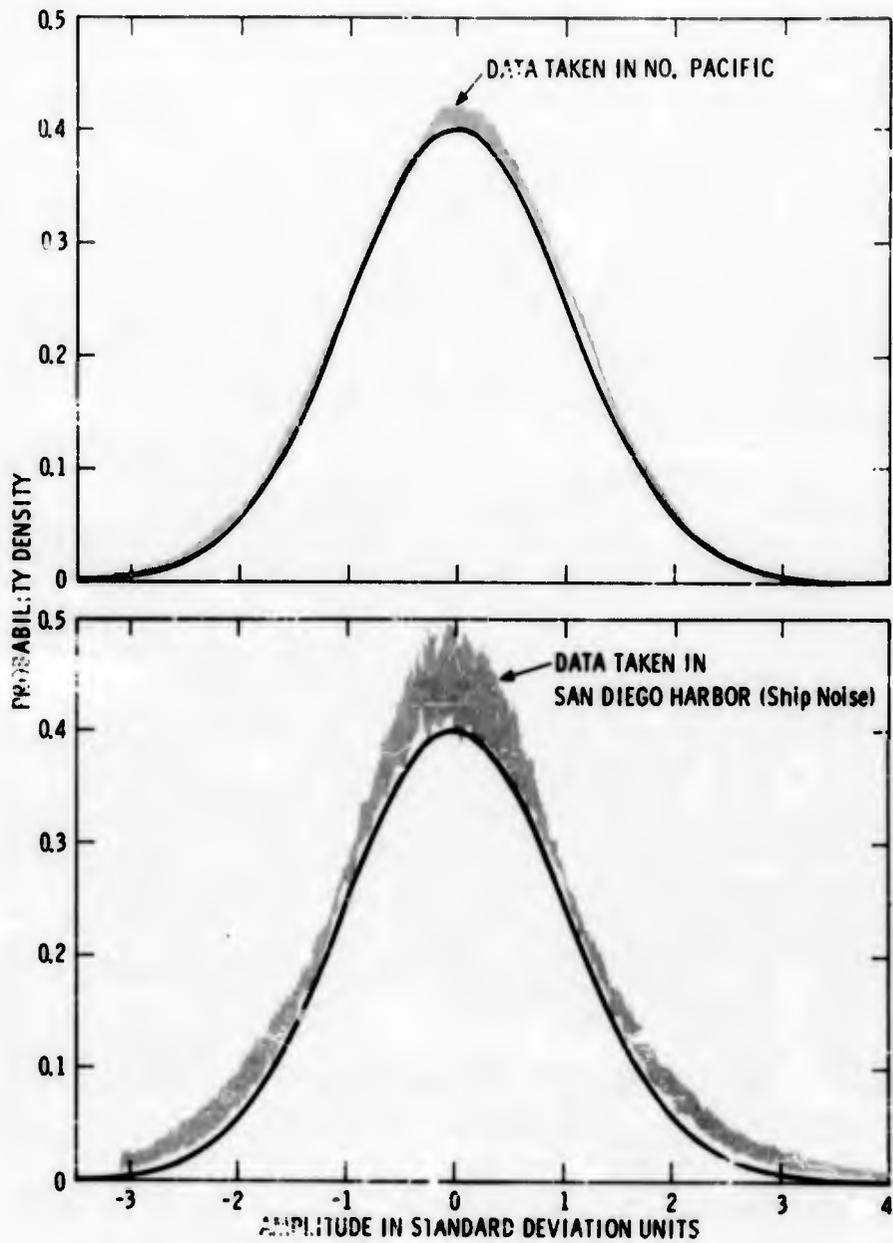


Figure 4. Examples of some PD curves taken in both shallow and deep water, compared with a normal curve.

Data Reduction Techniques

OVERLAY METHOD

Since it was expected that the probability density curves obtained with the PDA would have a gaussian or nearly gaussian distribution, an overlay with a gaussian curve was used. The curve had parameters of a mean equal to zero and a standard deviation equal to one. Figure 5 illustrates the use of this method with two curves, one judged to be gaussian and the other non-gaussian. Some probability density curves obtained with the PDA were judged to be very nearly gaussian.

One disadvantage of the overlay method is that decisions about how well a particular curve compares with the overlay are purely subjective. Skewness and kurtosis can be detected, but the magnitudes of these moments cannot be estimated with accuracy. An extension of the overlay method which will allow estimates of skewness and kurtosis is described here.

The extension is an overlay with several curves instead of just one. Each curve has a different set of values for skewness and kurtosis. The curves are positioned over the actual probability density curve and the parameters are estimated by interpolation between the two closest curves. The curves of the overlay can be computed with the use of Edgeworth's series approximation for nearly gaussian distributions.³ The first four terms of this series are

$$f(x) = h(x) - \frac{\theta_1}{3!} h^3(x) + \frac{\theta_2}{4!} h^4(x) + \frac{10\theta_1^2}{6!} h^6(x) \quad (3)$$

where $h(x)$ is the normalized gaussian distribution, $h^n(x)$ is the n th derivative of $h(x)$, θ_1 is the standardized skewness, and θ_2 is the standardized kurtosis.

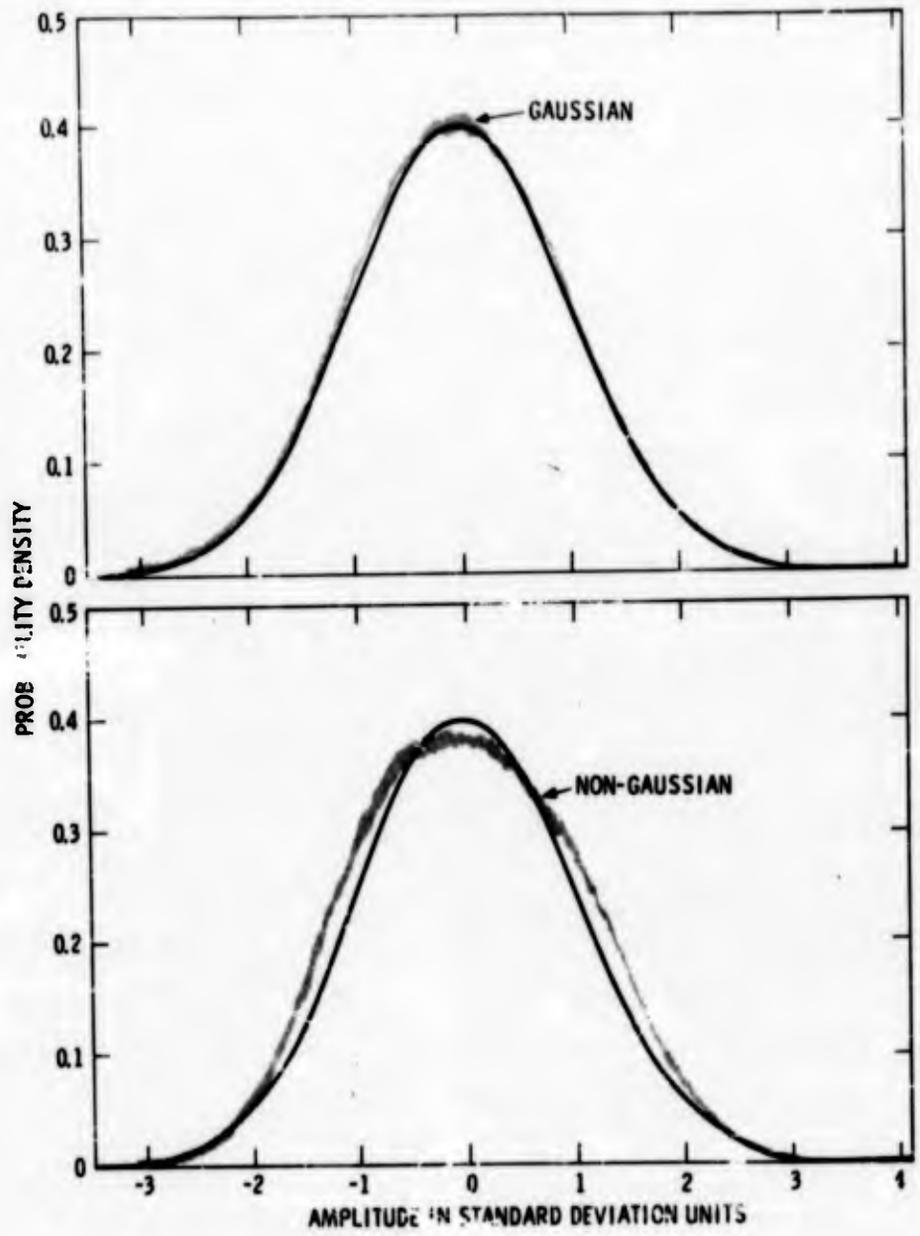


Figure 5. Examples of two PD curves which were determined to be gaussian or non-gaussian, using a normal curve as an overlay.

Estimates of the skewness and kurtosis can be found with the above method; but it does not give any indication of whether these estimates are significantly different from the expected values, if the sample is taken from a gaussian distribution. Using the previous overlay, a method can be developed so that a sample can be accepted or rejected at any desired level of probability. Basically the method is to have two of the curves on the overlay plotted so that they will represent the maximum deviations allowed in the particular parameter of a sample with (N) points. The method will be developed for kurtosis, but a similar method can be used for skewness.

The variance of kurtosis is given by⁴

$$\text{var}(g_2) = 24/N \quad (4)$$

for large N . This holds for a sample taken from a normal parent population. The standard deviation of kurtosis is $(24/N)^{\frac{1}{2}}$; if the kurtosis is distributed normally, then from the ratio of a particular value of kurtosis (g_2') and the standard deviation we can obtain the probability of getting a value of kurtosis as large or larger than g_2' . The ratio is

$$\frac{g_2'}{(24/N)^{\frac{1}{2}}} = R \quad (5)$$

The probability of getting a value of kurtosis as large as or larger than g_2' is given by the amount of area under a normal curve outside the $-R$ and $+R$ standard deviations. A value of $R = 1.96$ corresponds to a probability level of

5 per cent, or 1/20th the total area. A ratio as large as 1.96 may be considered sufficiently improbable and hence g_2' can be assumed to result from a non-gaussian distribution. The sample would therefore be rejected as coming from a gaussian distribution. The value of g_2' therefore depends on N , $g_2' = 1.96(24/N)^{1/2}$. Edgeworth's series would then be used to compute two curves, one with $-g_2'$ (for negative kurtosis) and one with $+g_2'$ (for positive kurtosis). These curves would represent the limits, at a 5 per cent probability level, within which a sample of N points would be considered as coming from a gaussian distribution.

Figure 6 shows two curves as they would appear in the overlay. These two curves are the limits for a sample

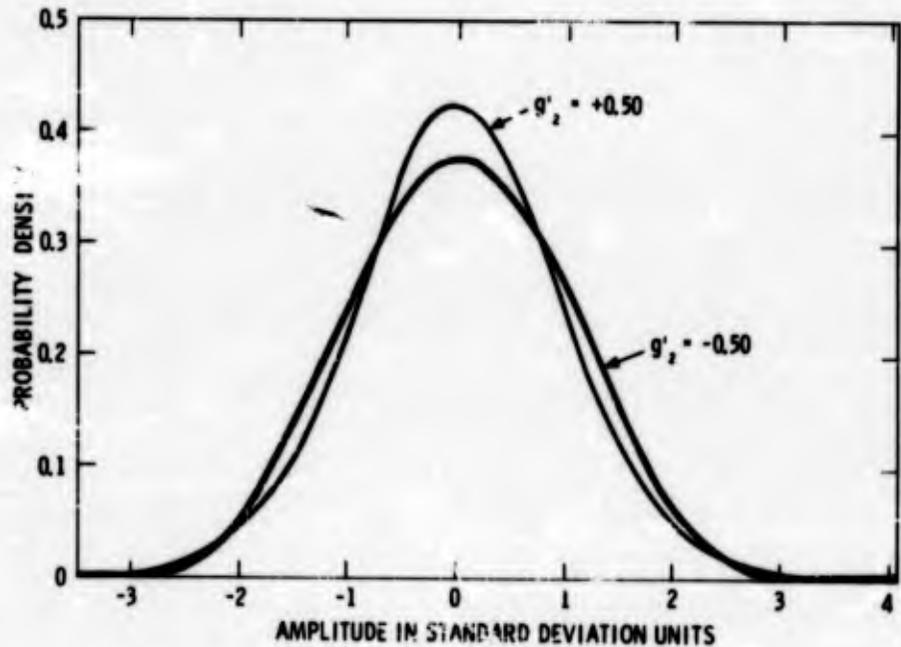


Figure 6. Overlay indicating g_2' of +0.50 and of -0.50. A curve having a value of kurtosis as large or larger than these values will be non-gaussian at a 5 per cent level for a sample of 370 points or, equivalently, a bandwidth of about 55 c/s.

of bandwidth about 55 c/s, with N given by the equation $N = 6.7 f_0$, where f_0 is the bandwidth. The equation is obtained from Appendix B, using a time constant $T = 2.3$ seconds.

The overlay method was not used extensively because of the complexity that comes from considering different values of N and also different combinations of skewness and kurtosis in the same sample. A method using computed moments of the curves is described next; it was felt that this method would yield accurate values of the mean, standard deviation, skewness, and kurtosis.

METHOD OF MOMENTS

The method of moments is basically a general method of forming estimates of the parameters of a distribution by means of a set of measured sample values. The first few moments of the actual distribution are calculated and these are used as estimates of the moments of the parent population. On the basis of these moments a suitable theoretical distribution curve is selected. For any particular distribution curve the moments are functions of the parameters of that curve. The parameters are determined and tests of significance are made on the skewness and kurtosis.

The moments about the origin are defined as⁵

$$m_r' = \sum_t p_t(x) x_t^r \quad (6)$$

where $p_t(x)$ is the probability that a value selected at random from the population will lie in the t th class. The variate x with which we are concerned may be discrete or continuous.

The moment

$$m_1' = \sum_t p_t(x) x_t \quad (7)$$

is defined as the mean value of x , $m_1' = \bar{x}$.

Another more important set of moments is obtained by changing the origin to the arithmetic mean. Equation 8 defines the moments about the mean.

$$m_r = \sum_i f_i(x) (x_i - \bar{x})^r \quad (8)$$

For computing purposes, the relations between the m_r and the m_r' are convenient. Expressing the m_r in terms of the m_r' we have the relations

$$m_1 = m_1' \quad (9a)$$

$$m_2 = m_2' - (m_1')^2 \quad (9b)$$

$$m_3 = m_3' - 3m_2'm_1' + 2(m_1')^3 \quad (9c)$$

$$m_4 = m_4' - 4m_3'm_1' + 6m_2'(m_1')^2 - 3(m_1')^4 \quad (9d)$$

Grouping errors are negligible, so Sheppard's corrections are not applied.

These moments can be expressed in standard units by the use of a standardized variable z , by dividing the variable x by s_x , the standard deviation.

$$z = \frac{(x - \bar{x})}{s_x} \quad (10)$$

The standardized moments are defined by the equations

$$\alpha_r = \frac{m_r}{s_x^r}, \text{ for } r = 1, 2, 3, \text{ and } 4 \quad (11)$$

The first four standardized moments are

$$\alpha_1 = 0 \quad (12a)$$

$$\alpha_2 = 1 \quad (12b)$$

$$\alpha_3 = \frac{n_3}{\sigma^3} \quad (12c)$$

$$\alpha_4 = \frac{n_4}{\sigma^4} \quad (12d)$$

The third moment, α_3 , is a measure of the skewness of the distribution. A positive value indicates a distribution with a longer positive tail than a negative tail.

The fourth standardized moment, α_4 , is a measure of the kurtosis of the distribution. In some cases it is a measure of the "peakedness" of the distribution, though it is now understood that the length and size of the tails are very important in this measurement.

For a normal curve the values of α_3 and α_4 will be 0 and 3, respectively. We redefine the skewness and kurtosis as

$$\vartheta_1 = \alpha_3 \quad (13a)$$

$$\vartheta_2 = \alpha_4 - 3 \quad (13b)$$

so that ϑ_2 is 0 for a normal curve.

It is not very likely that the third and fourth moments of a random sample will be zero. Depending on the distribution and on the actual sample values, the third and fourth moments will have some value different from zero. To determine whether this difference is significant, it is necessary to use the variances of the third and fourth moments.⁴

$$\text{var}(\vartheta_1) = 6N(N-1)(N-2)^{-1}(N-1)^{-1}(N-3)^{-1} \quad (14a)$$

$$\text{var}(\varrho_2) = 24N(N-1)^2(N-3)^{-1}(N-2)^{-1}(N-3)^{-1}(N-5)^{-1} \quad (14b)$$

For large N use,

$$\text{var}(\varrho_1) = 6/N \quad (15a)$$

$$\text{var}(\varrho_2) = 24/N \quad (15b)$$

The hypothesis to be tested is that the data sample is taken from a gaussian distribution. To test the hypothesis compare ϱ_1 to $(6/N)^{\frac{1}{2}}$ and ϱ_2 to $(24/N)^{\frac{1}{2}}$ (see ref. 5), then

$$\begin{aligned} \text{if } \frac{\varrho_1}{(6/N)^{\frac{1}{2}}} > 1.96 & \text{ reject the hypothesis at the 5 per cent level} \\ & > 2.57 & \text{ reject the hypothesis at the 1 per cent level.} \end{aligned}$$

Similarly, for ϱ_2

$$\begin{aligned} \text{if } \frac{\varrho_2}{(24/N)^{\frac{1}{2}}} > 1.96 & \text{ reject the hypothesis at the 5 per cent level} \\ & > 2.57 & \text{ reject the hypothesis at the 1 per cent level.} \end{aligned}$$

CHI-SQUARE "GOODNESS OF FIT" TEST

The χ^2 test will be applied to the hypothesis that a sample of N individuals forms a random sample from a population with a given probability distribution. The parameters of a distribution are known and are not estimated from the sample itself. Later a modification will be given for the situation where the parameters are estimated from the sample.

The quantity*

$$\chi^2_s = \sum_{l=1}^k \frac{(F_l - NP_l)^2}{NP_l} \quad (16)$$

is a measure of the deviation of the sample from the expectation, where F_l is the number of observed frequencies in the l th interval, and NP_l is the number of expected frequencies in the l th interval as predicted by the theoretical distribution. Karl Pearson proved that the above quantity, in the limit, is the ordinary χ^2 distribution which is now tabulated in most statistics books.

The χ^2_s computed with equation 16 is compared with the 5 per cent point for $(k-1)$ degrees of freedom from a χ^2 distribution table. The tabulated value of χ^2 at the 5 per cent probability level with 29 degrees of freedom is 42.6. Now, if χ^2_s , as calculated by equation 16, is greater than 42.6, the hypothesis is rejected by this test; that is, the sample is non-gaussian.

The application of the χ^2 test to the data was as follows (fig. 7). Let $f(x)$ represent a probability density curve obtained from the PDA. Divide the curve into 30 intervals from $x = -3.0$ to $x = +3.0$. Let ξ_l be the mid-point of Δx_l , one of the intervals, and $f(\xi_l)$ the value of the probability density at ξ_l . The area under the curve is then estimated by $\sum A_l'$, where $A_l' = f(\xi_l)\Delta x$. Let $A_l =$

$$\int_{z_{l-1}}^{z_l} \varphi(z) dz, \text{ where } \varphi(z) \text{ is the theoretical probability}$$

* See ref. 5, pp. 197-200.

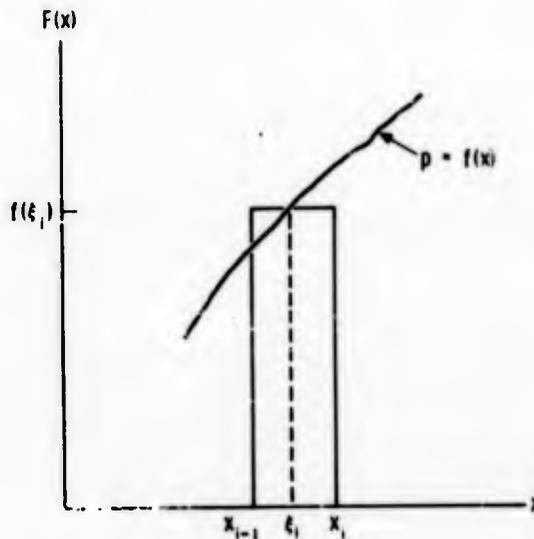


Figure 7. Curve obtained with the PDA ($p = f(x)$) vs. measured probability density at ξ_i ($f(\xi_i)$). The area within the rectangle (A_i') is an estimate of the area under the curve for the interval

$$\Delta x = x_i - x_{i-1}$$

$f(\xi_i)$: MEASURED PD AT ξ_i
 $p = f(x)$: CURVE OBTAINED WITH PDA

density distribution with which the experimental curve is being compared. Also let $n_t = NA_t'$, and $e_t = NA_t$.

$$\chi^2_s = \sum_t \frac{(n_t - e_t)^2}{e_t} \quad (17)$$

$$\chi^2_{\text{fit}} = N \sum_t \frac{(A_t' - A_t)^2}{A_t} \quad (18)$$

N is estimated according to the method described in Appendix B.

The only modification needed when the parameters are estimated from the sample itself is a reduction in the number of degrees of freedom by one for each parameter that is estimated from the sample. For example, if a gaussian distribution is assumed and the mean and standard deviation are estimated from the sample, then the number of degrees of freedom are $(k-1)-2$, where k is the number of groups.

CUMULATIVE PROBABILITY PLOTS

The use of cumulative probability paper was also investigated. On this type of plot a gaussian distribution is represented by a straight line. The data are normalized and plotted, and deviations from a gaussian distribution are seen as departures from a straight line. The data are the $P_i(\xi_i)$ used for the method of moments where ξ_i is the midpoint of $\Delta x_i = x_i - x_{i-1}$. The points plotted are the normalized cumulative sums, i.e., the first point is

$$\frac{P_1(\xi_1)}{\Sigma P_i(\xi_i)}$$

the second point is

$$\frac{P_1(\xi_1) + P_2(\xi_2)}{\Sigma P_i(\xi_i)}, \text{ etc.}$$

The range used is from $x = -3.00$ to $x = +3.00$ so that a gaussian curve resembles figure 8. The curve deviates from a straight line at the ends because of the small amount of area (0.27 per cent) outside three standard deviations. This curve should be used for comparison with the data instead of the straight line.

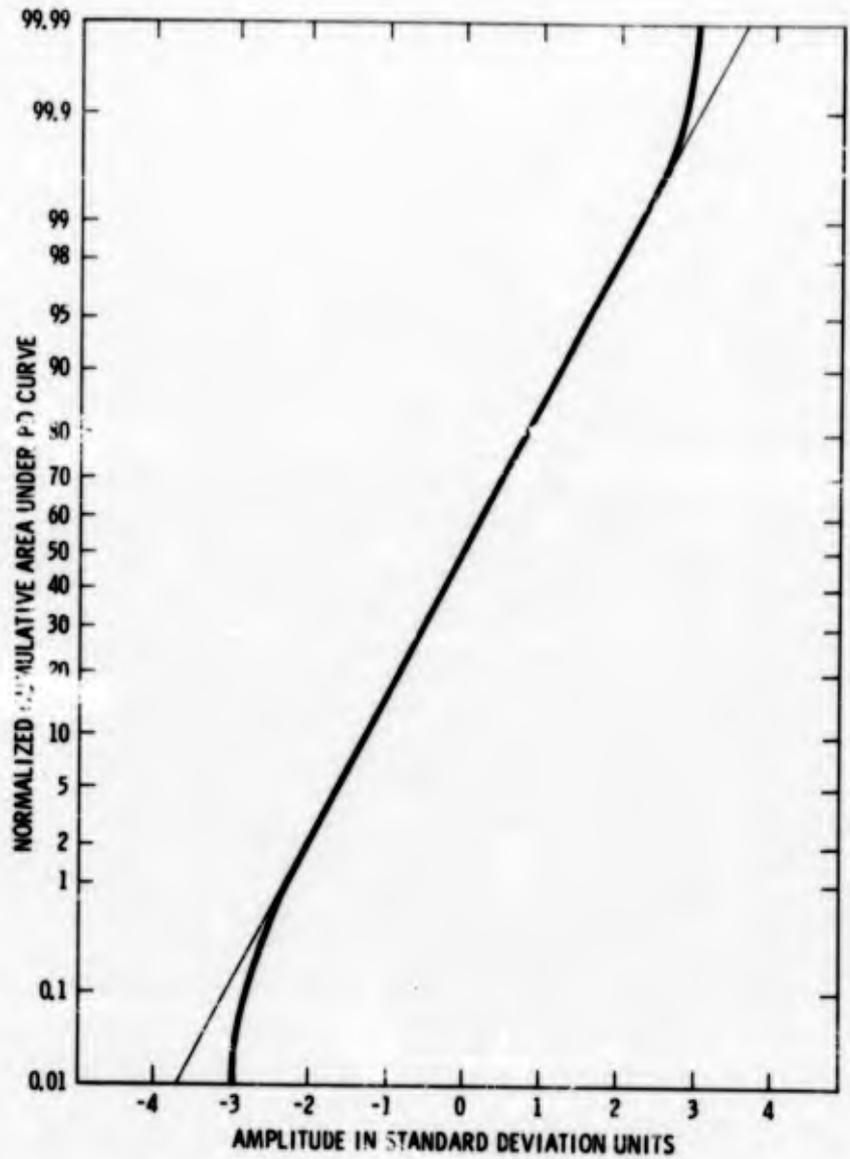


Figure 8. Normalized cumulative sums of tabulated values of the area of a normal curve for values of x between -3.00 and $+3.00$ standard deviations.

The type of deviations that would result for curves with skewness and kurtosis are shown in figures 9 and 10. The distribution curves for certain amounts of skewness and kurtosis are computed using Edgeworth's first four

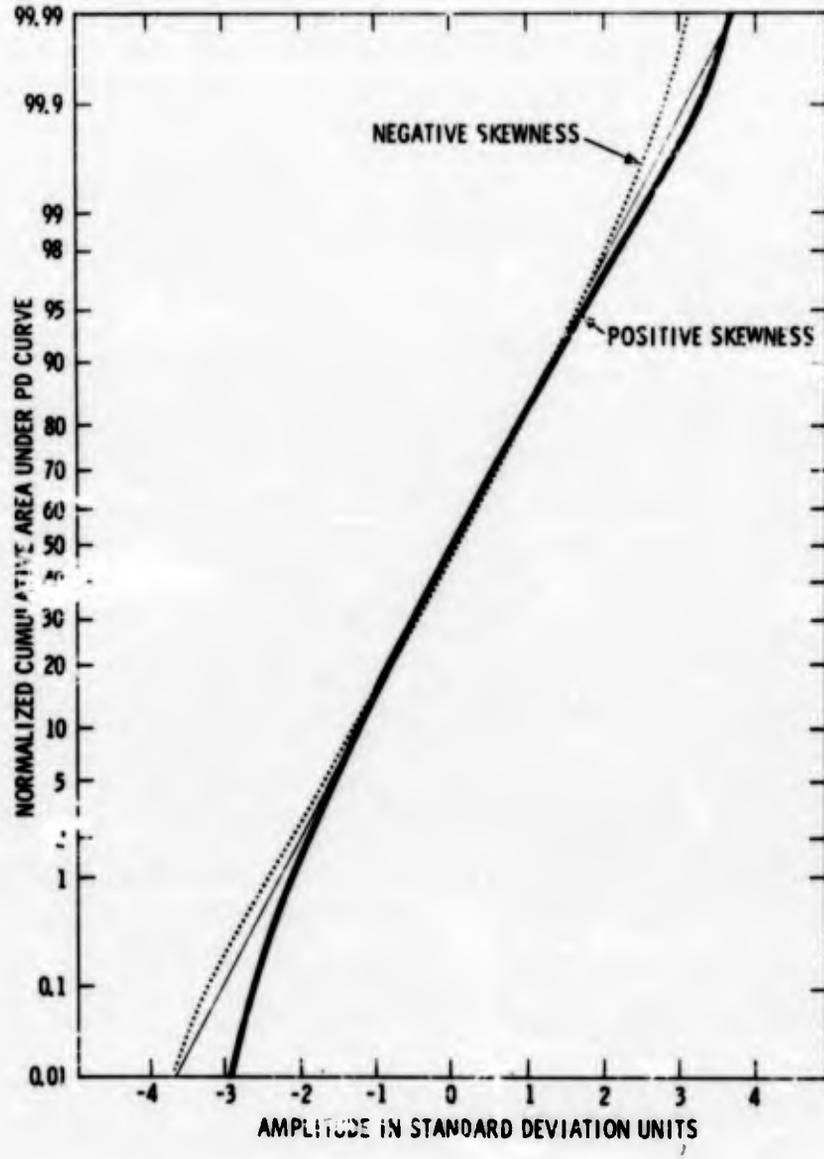


Figure 9. Curves illustrating positive and negative skewness, computed using Edgeworth's series.

terms, equation 4. The $f(x)$ were computed using a mean of zero and a standard deviation of one. For skewness, $\varrho_1 = \pm 0.24$, $\varrho_2 = 0$ were used. For kurtosis, $\varrho_1 = 0$, $\varrho_2 = \pm 0.48$ were used.

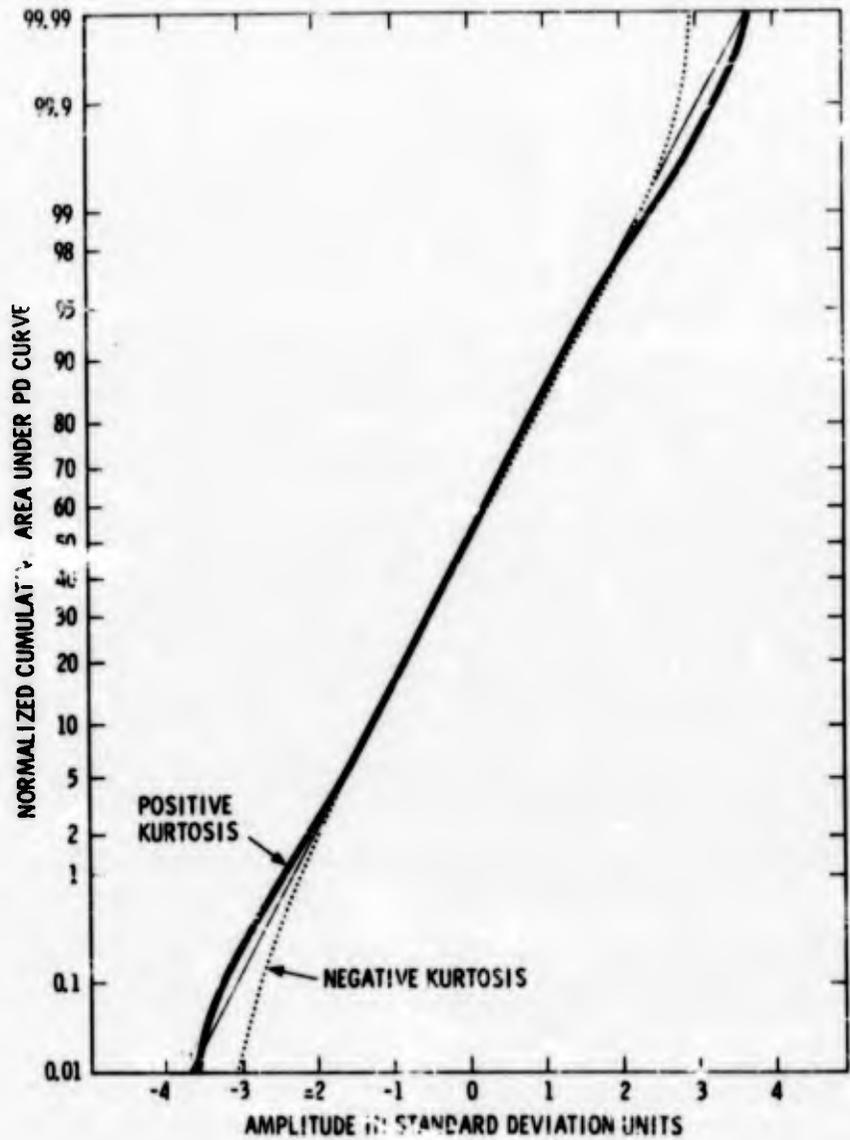


Figure 10. Curves illustrating positive and negative kurtosis, computed using Edgeworth's series.

Several cumulative probability examples are shown in figures 11 and 12.

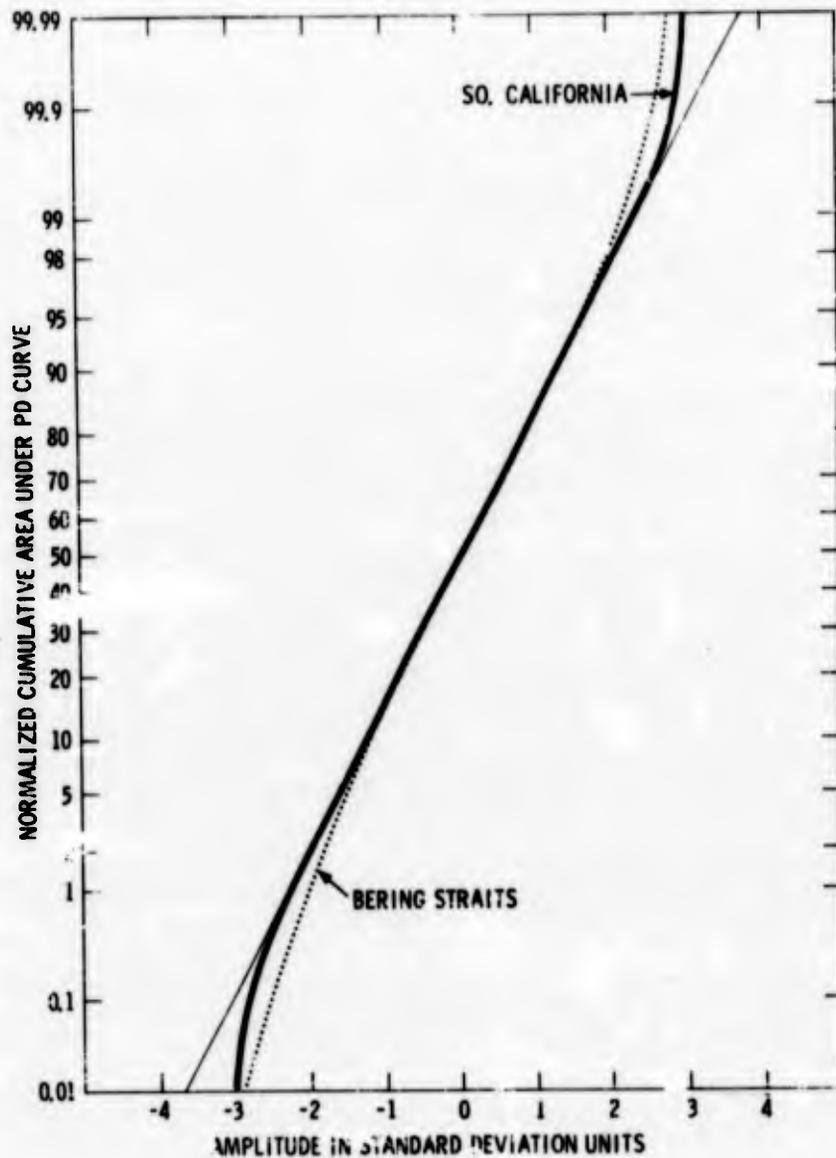


Figure 11. Cumulative probability of noise samples shown in figure 3.

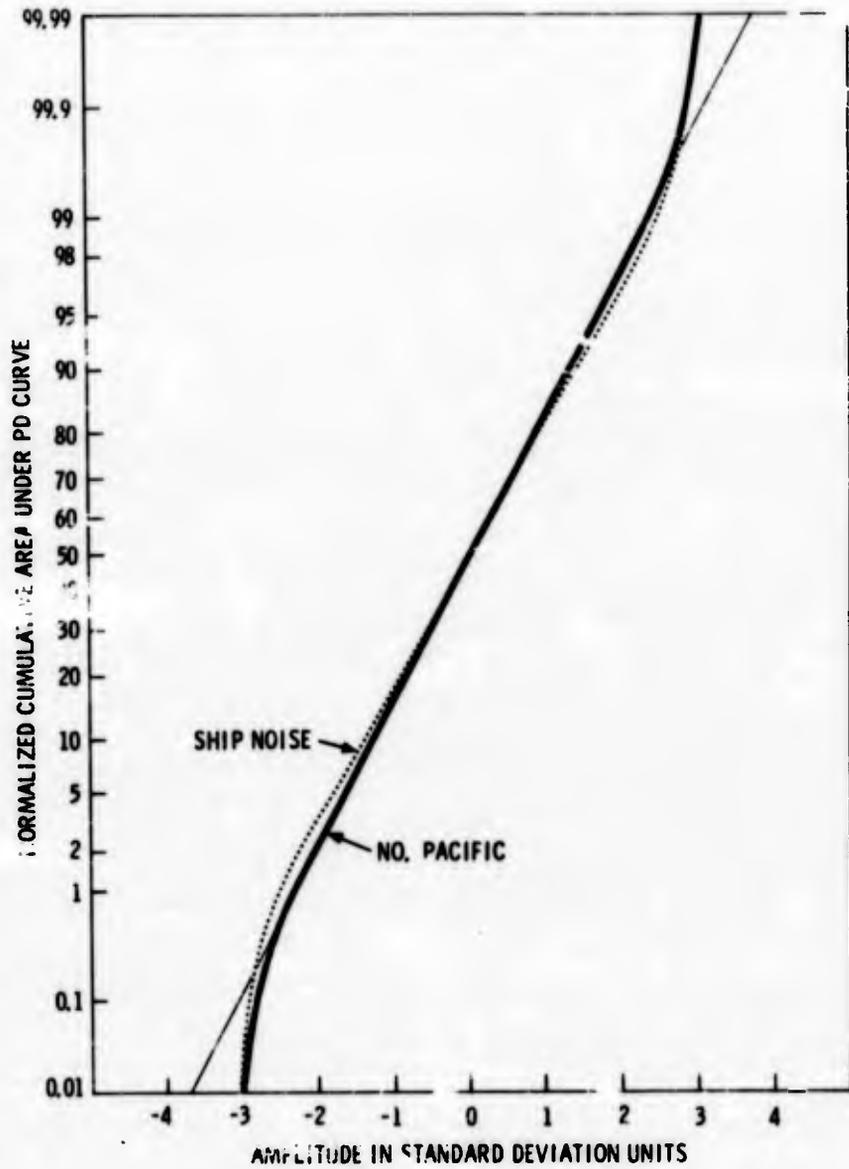


Figure 12. Cumulative probability of noise samples shown in figure 4

RESULTS

An overlay of a normal curve was constructed and used on the probability density curves obtained with the XY recorder. This method was found to have a limited usefulness since it would not yield results with the desired amount of accuracy in all cases. Cases where it could be used are for curves which do not deviate very much from the overlay. As the deviations become greater it becomes more difficult to estimate the parameters accurately. It was decided not to use an overlay with more than one curve (as described in "Data Reduction Techniques") because, even though the curves were normalized to a standard deviation of 1, the error in setting the input to a value of 1 volt rms can be as much as 10 per cent -- although in most cases it was less than 5 per cent. Such error makes it difficult to estimate the skewness and especially the kurtosis, since the overlay will not fit if the standard deviation is other than 1.

With the method of moments it is possible to get estimates of the mean, the standard deviation, the skewness, and the kurtosis. The limited range of amplitudes analyzed (which was thought to be sufficient before the data reduction) causes an error in the calculated moments for those samples which have amplitudes extending beyond three standard deviations. The error is more evident in the higher (3rd and 4th) moments, because of the higher powers of x used in the calculations of these moments. A correction can be applied to the computed skewness and kurtosis. The correction takes into account the "ignored" amplitudes, i. e., it compares the computed moment with the moment of a truncated normal distribution. Actually only the moment of kurtosis was corrected, since skewness was not found very frequently among the noise samples.

The hypothesis to be tested is that the noise samples are taken from a noise with a gaussian distribution of amplitudes. The test used is the same one described on page 21 and it is applied to both the skewness and the kurtosis. The

hypothesis is then rejected by the test if both moments of the sample are significant at the 5 per cent level, and also if either moment is significant at the 1 per cent level.

Table 2 lists the sample for which the 3rd and/or 4th moments were found to be significantly different from the

TABLE 2. NUMBER OF CURVES SHOWING SIGNIFICANT VALUES OF SKEWNESS AND KURTOSIS.

SAMPLE NO.	g ₁ (SKEWNESS)		g ₂ (KURTOSIS)		COMMENTS
	5%	1%	5%	1%	
13*	X		X		BIOLOGICAL NOISE
17			X		CLEAN NOISE
20*			X	X	BIOLOGICAL NOISE
21*			X	X	LOW-FREQ. HUM, CLANKS
23			X		CLEAN NOISE
27*			X	X	SHIP NOISE
28*			X	X	CLEAN NOISE
29*	X		X	X	MOSTLY CLEAN NOISE
42			X		SHIP NOISE
50*			X	X	ICE NOISE
51*			X	X	ICE NOISE
52			X		ICE NOISE
57*	X	X	X	X	ICE NOISE, 60 C/S
60*			X	X	60 C/S
62*			X	X	60 C/S
69			X		
84*			X	X	
100*			X	X	
135*	X	X			ALL THE FOLLOWING ARE SHIP NOISE SAMPLES
138	X				
139*	X	X			
142			X		
143*	X	X	X	X	
147			X		
149			X		
150*			X	X	
151*			X	X	
159			X		
163*			X	X	

*Hypothesis rejected for that sample.

expected values. An asterisk by the sample number indicates that the hypothesis was rejected by that particular sample according to the conditions of the previous paragraph. Comments on the type of noise (clean, biological, ship noise, etc.) in the sample are also given. The number of noise samples that were found to be significant are listed by location in table 3. Note that the shallow-water locations have many more non-gaussian samples than the deep-water North Pacific location.

TABLE 3. LOCATIONS OF CURVES SHOWING SIGNIFICANT VALUES OF SKEWNESS AND KURTOSIS AT 1 AND 5 PER CENT PROBABILITY LEVEL.

LOCATION	g ₁ (SKEWNESS)		g ₂ (KURTOSIS)		NO. OF CURVES	NO. OF SAMPLES
	5%	1%	5%	1%		
SHALLOW, SO. CALIF.	2	0	8	5	29	9
SHALLOW, ALASKA	1	1	7	5	24	9
NORTH PACIFIC	0	0	3	2	65	8
SAN DIEGO (SHIP NOISE IN HARBOR)	4	3	8	4	36	9

If the hypothesis that the noise samples are taken from a gaussian distribution is true, then it is expected that one sample in 20 may have significant parameters at the 5 per cent level. If it is not true, it is expected that more than one sample in 20 may have significant parameters.

Table 3 shows that the values of skewness for the first three locations are well within the expected number, except perhaps for the Southern California location, where two samples out of 29 were found with significant amounts of skewness. However, since this location had considerable

biological and other non-gaussian type noises, a greater number of rejected samples is to be expected. The table also shows that for kurtosis the expected number (one out of 20) was exceeded for both of the shallow-water locations but was not exceeded for the deep-water (North Pacific) location. For the ship-noise samples, the number of expected significant values of skewness and kurtosis are exceeded at all levels. Of interest are the number of significant values of skewness in the ship-noise data, since there were not many of these for the other three groups. For all groups the data indicate that kurtosis is a sensitive indicator for the presence of ship noise, biological noise, ice noise, and in general any type of noise which has a non-gaussian distribution.

Table 4 indicates the bandwidth of the noise samples which had significant values of skewness and kurtosis at the 1 per cent level. For the most part the bandwidths involved are the larger ones.

TABLE 4. NOISE SAMPLES FROM FOUR LOCATIONS, SHOWING SIGNIFICANT MOMENTS OF SKEWNESS AND KURTOSIS AT 1 PER CENT PROBABILITY LEVEL.

SOUTHERN CALIFORNIA			BERING STRAITS		
BANDWIDTH	g_1 (1%)	g_2 (1%)	BANDWIDTH	g_1 (1%)	g_2 (1%)
20-2500	0	5	BROADBAND	0	1
20-1500	0	0	20-1200	1	2
			20-600	0	2
a			b		
NORTH PACIFIC			SAN DIEGO (SHIP NOISE IN HARBOR)		
BANDWIDTH	g_1 (1%)	g_2 (1%)	BANDWIDTH	g_1 (1%)	g_2 (1%)
1 OCTAVE 1600 c.f.	0	2	20-1200	3	3
1/3 OCTAVE 1600 c.f.	0	1	600-1200	0	1
c			d		

An Edgeworth series was fitted to an experimental curve using the parameters calculated by this method, to insure that they were good estimates to the extent that they provided a good fit to the data. Figure 13 is the trace of an experimental curve of random noise which shows considerable skewness. The fit of an Edgeworth series approximation to the experimental curve is seen to be very good.

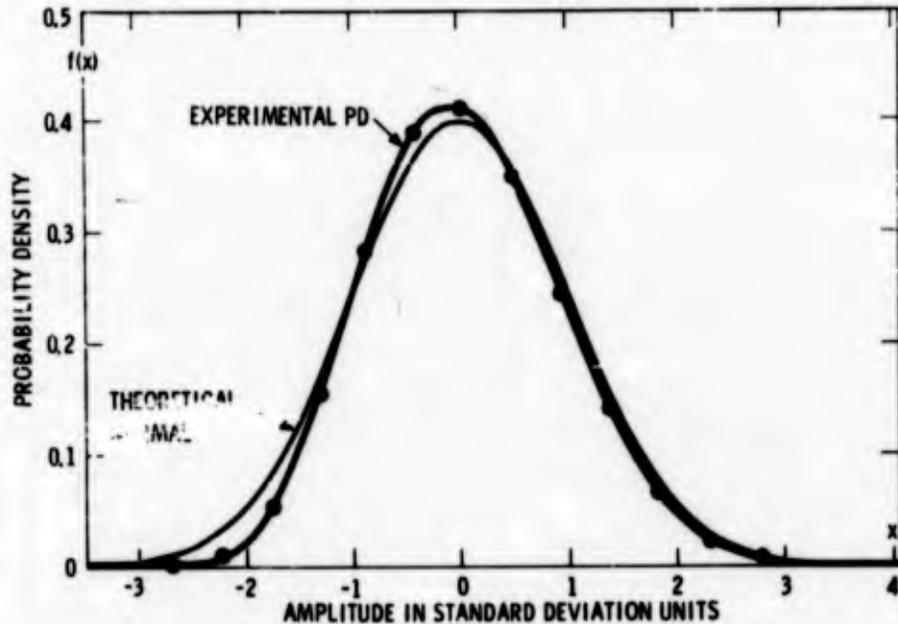


Figure 13. Experimental PD curve of random noise showing considerable skewness. The points shown are calculated using Edgeworth's series with the moments calculated from the curve itself by the method of moments described in the text. A theoretical normal curve is shown for comparison.

The chi-square test was performed on a few selected curves and the results are shown in table 5. An X marks those noise samples for which χ^2 was equal to or greater than 40.6, which is the tabulated value of χ^2 for 29 degrees of freedom at a 1 per cent probability level. The value of χ^2 is given for the samples for which the hypothesis was not rejected by the test. The figure also shows whether

SAMPLE NO.	χ^2_s	g_1	g_2
13	X	X(5%)	X(5%)
14	31.2		
15	17.5		
16	32.6		
17	X		X(5%)
18	18.2		
30	14.7		
32	2.8		
34	14.4		
37	3.9		
41	2.7		
42	32.4		X(5%)
49	40		
66	3.3		
68	38		
85	5.4		
104	2.2		
112	42		
132	5.5		
134	X		
135	X	X(1%)	
147			X(5%)
150	X		X(5%)

TABLE 5. CURVES FOR WHICH χ^2_s WAS COMPUTED, COMPARED WITH RESULTS BY METHOD OF MOMENTS. THE X'S INDICATE THOSE CURVES WHICH REJECTED THE HYPOTHESIS.

the 3rd and 4th moments were significant at the 5 per cent or 1 per cent level. Generally the results of the chi-square test agree with the results of the method of moments.

Only one hypothesis was tested and this was that the noise samples were taken from a gaussian distribution with the mean equal to zero and the standard deviation equal to one. A better hypothesis is to assume that the distribution is gaussian with a mean $\bar{x} = m$ and standard deviation $s = \sigma_x$, where m and σ_x are estimated from the sample itself.

Table 6 shows parameters for several noise samples which were selected by the overlay normal curve as very closely approximating a gaussian curve. The four computed moments of each sample are given (the mean, the standard deviation, the skewness, and the kurtosis). The ratio of the skewness and kurtosis to the square root of their variances are given. The computed value of χ^2_s is given for comparison. All three methods agree that these samples

TABLE 6. NOISE SAMPLES CHOSEN, BY VARIOUS METHODS, AS BEING VERY CLOSELY GAUSSIAN.

SAMPLE NO.	MEAN	COMPUTED MOMENTS			g_1	g_2	χ^2_c
		STD. DEV.	g_1 (SKEWNESS)	g_2 (KURTOSIS)	$(\text{var } g_1)^{1/2}$	$(\text{var } g_2)^{1/2}$	
30	0.00008	0.97868	-0.00198	-0.067	0.85	1.45	14.7
32	0.00458	0.97796	0.01884	-0.005	0.36	0.05	2.8
37	0.01141	1.00828	0.00793	-0.030	0.13	0.25	3.9
41	0.00690	0.99298	0.00965	0.003	0.22	0.03	2.7
67	0.05984	0.99511	-0.01290	0.014	0.33	0.18	
68	0.06427	0.98920	-0.01077	0.030	0.39	0.52	36
83	0.02767	0.98493	-0.02185	0.009	0.56	0.12	5.4
98	0.04498	0.97240	-0.01331	0.019	0.24	0.16	
104	0.00398	0.99926	-0.01461	0.003	0.30	0.03	2.2
112	0.04519	1.00764	-0.00009	0.006	0.60	0.11	42

are very closely gaussian. The largest value of chi-square occurs for sample no. 112, which has a very good shape; however it does have a mean which is different from zero, thus causing the large χ^2_c . Sample no. 68 also has a large χ^2_c for the same reason.

Cumulative probability graphs of the noise samples can reveal if the curve has skewness or kurtosis and can also reveal other deviations. However this method was not used beyond plotting a few curves. The graphs would perhaps require an overlay to estimate skewness and kurtosis but, as before, the sample distribution needs to be standardized for each sample before the overlay can be applied.

Analysis with the cumulative probability method was not performed, as it was felt that it would not provide much more information than the method of moments and the chi-square test, and the estimates would not be as accurate as those obtained by the method of moments. This method does, however, give a better indication of whether a noise sample is gaussian than does the overlay method on the XY probability density curves.

CONCLUSIONS

PD of Ambient Ocean Noise

The results obtained from the analysis of ambient ocean-noise samples from the three locations indicate that the hypothesis (that is, the assumption that the ocean samples are taken from a gaussian noise distribution) is not rejected when using "clean" ambient noise. The hypothesis is rejected for "contaminated" ambient ocean noise. The contamination may be ship noise, biological noise, noise from ice (in polar regions), etc. Thus it has been shown that ambient ocean noise, under certain conditions, can be assumed to have a gaussian distribution of amplitudes.

PD of Ship Noise

The same hypothesis used for the ambient noise samples was used to test the ship samples. The results indicate that ship-noise samples are not gaussian, since six out of nine samples tested rejected the hypothesis that the samples were taken from a gaussian distribution. It was not found possible to distinguish between the types of contamination of the ambient noise; that is, there were no obvious differences in the probability density distributions for ambient noise contaminated by ship noise and that contaminated by other sources.

Comparison of Test Methods

The method of moments was found to be the most suitable, of the four methods of data reduction, for providing the most accurate and useful estimates of the parameters. However, when using this method the tails of the distribution should not be left out of the calculations, since their contribution at the higher moments is very significant. The

chi-square test, although it does not provide estimates of the parameters of the distribution, does provide a good indication of the "fit" of the sample distribution to the theoretical distribution. The chi-square method can be used to give an independent check on the method of moments.

RECOMMENDATIONS

1. Use the method of moments described here if better accuracy than that given by overlays is desired to estimate the moments and to determine whether a sample is gaussian or non-gaussian.
2. In the probability density analysis of a noise sample, use a range of amplitudes covering at least ± 4 standard deviations; otherwise large errors in the estimates of the moments will frequently result.
3. In future applications of the PDA, have the output of the PDA in a digital form rather than a continuous curve, so that the data will be in a form more suitable for the calculation of the moments of distribution.

REFERENCES

1. Bolt, Beranek and Newman, Inc., Report 895, Probability Density Analyzer, Summary Report, by H. L. Fox, September 1963
2. Bell Telephone Laboratories Contract Nonr-2461(00): Technical Report 10, A Study of the Bivariate Distribution Function of the Ocean Noise, by A. H. Green, p. II-2 - II-3, 13 April 1962
3. Cramér, H., Mathematical Methods of Statistics, p. 229-230, Princeton University Press, 1946
4. Kenney, J. F. and Keeping, E. S., Mathematics of Statistics, 2d ed., v. 2, p. 189-190, Van Nostrand, 1951
5. Kenney, J. F. and Keeping, E. S., Mathematics of Statistics, 3d ed., v. 1, p. 90-103, 116, Van Nostrand, 1954

APPENDIX A:

DESCRIPTION OF THE PDA AND ITS OPERATION

The PDA was designed primarily for the determination of the probability density (PD) of random noise. Approximations of the PD curves of other waveforms, such as sine waves, square waves, and others can also be obtained. The PD of a sine wave and square wave as obtained with the PDA are shown in figure A1. Also shown are the theoretical PD's of the two curves.

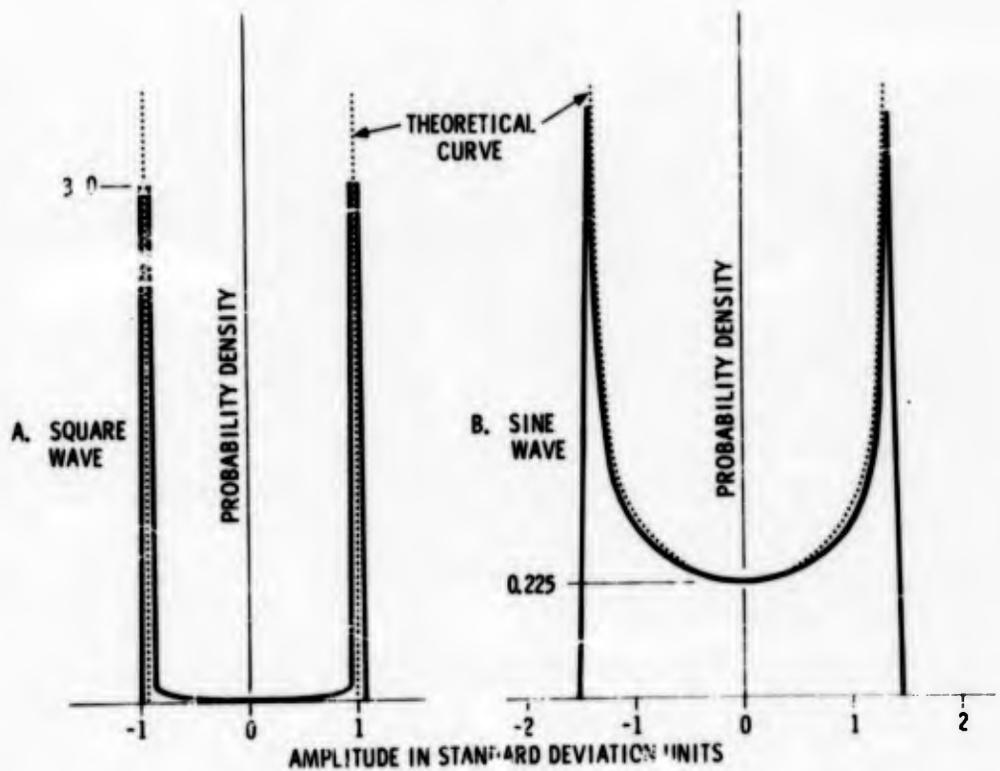


Figure A1. Theoretical and experimental probability density curves of the square wave and the sine wave.

The PDA has one input and several types of outputs. The input level to the PDA is controlled by a precision 10-turn potentiometer and the level can be in the range from 1 volt to 50 volts rms. The frequency range is from dc to 10 kc/s and this range should never be exceeded. The input level is set with the aid of a true-rms voltmeter.

There are two analog outputs. The X (amplitude) output is from -10 to +10 volts and corresponds to +5 to -5 times the rms input level. The Y (probability density) output is from 0 to about 7.5 volts and corresponds to a PD from 0 to 10. However a meter on the instrument gives a reading of PD on either of two scales from 0 to 0.4, or 0 to 1.0. A PD greater than one can be measured using the digital outputs, of which there are two: a 1-Mc/s and a 10-Mc/s output. Both digital outputs are on when the input signal is inside the amplitude interval Δx , at any selected amplitude. The PD is obtained by counting the number of cycles per second from either digital output. The counting interval can be varied as the occasion demands.

Another output available on the PDA is the pulse output. This particular output gives one pulse each time the signal goes into the amplitude interval, or two pulses for signals exceeding the set-in level (one pulse when signal passes through on the way up and another pulse when it is on the way down).

Some of the important controls on the PDA are the input potentiometer (to determine and set input level), the "X" 10-turn calibrated potentiometer (to select the amplitude around which the probability density is to be measured), and the output-damping-sweep time control (to select the averaging time and sweep speed). Other controls deal with the calibration of the PDA and are better described in the PDA manual.¹

APPENDIX B:

DETERMINATION OF NUMBER OF DATA POINTS OF EACH SAMPLE

The application of the χ^2 test to our type of data was desired as a means of testing the "goodness of fit" of the data to certain theoretical distributions and, in particular, the fit to the gaussian distribution. There is some difficulty in applying this test directly because N , the total number of points in our sample, is not known. However it is possible to obtain an estimate of N which can be used. The estimate of N is also used in determining whether the moments of the distribution, that are calculated by the method of moments, are significant at a given probability level. The estimate of N is found as follows.

From reference 1 use

$$\sigma_n^2 = \frac{(3/8\pi)^{\frac{1}{2}}}{\int_0^T w(x) \Delta x} \quad (\text{B-1})$$

where f_0 is the input bandwidth of the signal, T is the sampling interval or a time constant which gives an equivalent averaging, and $w(x)$ is the value of probability density obtained from the PDA. The σ_n^2 is the normalized variance of the estimate of the probability density, and is not to be confused with the variance of the distribution of amplitudes of the input signal. In reference 1, N is given as

$$N = \frac{1}{\sigma_n^2 \Delta x} \quad (\text{B-2})$$

Here N represents the total number of times that the input signal enters the amplitude interval Δx at the (x) of interest. After combining equations B-1 and B-2 we have

$$N = \frac{\int_0^T w(x)}{(3/8\pi)^{\frac{1}{2}}} = 2.907 w(x) \quad (\text{B-3})$$

The total number of points in the entire sample is what is needed and N represents only that number in the interval Δx . However, we have normalized data and therefore

$$\sum w(x) \Delta x = 1 \quad (\text{B-4})$$

$$N_T = \sum N \Delta x = \frac{\int_0^T \omega}{(3/8\pi)^{\frac{1}{2}}} \quad (\text{B-5})$$

Calculations performed with this equation agree well with some data taken using the pulse output of the PDA. Figure B1, curve B, shows the results of equation B-3 for $T = 1$ second, and $w(x) = w(0) = 0.4$, or

$$N = 1.16 \int_0 \quad (\text{B-6})$$

The theoretical curve is compared with actual measurements taken from the PDA pulse output using a counter with a 1-second sampling interval. The slight difference between the two is due to an error in the actual cutoff frequency of the filter. The effective bandwidth of the filter is a little larger, about 1.12 times the set value.

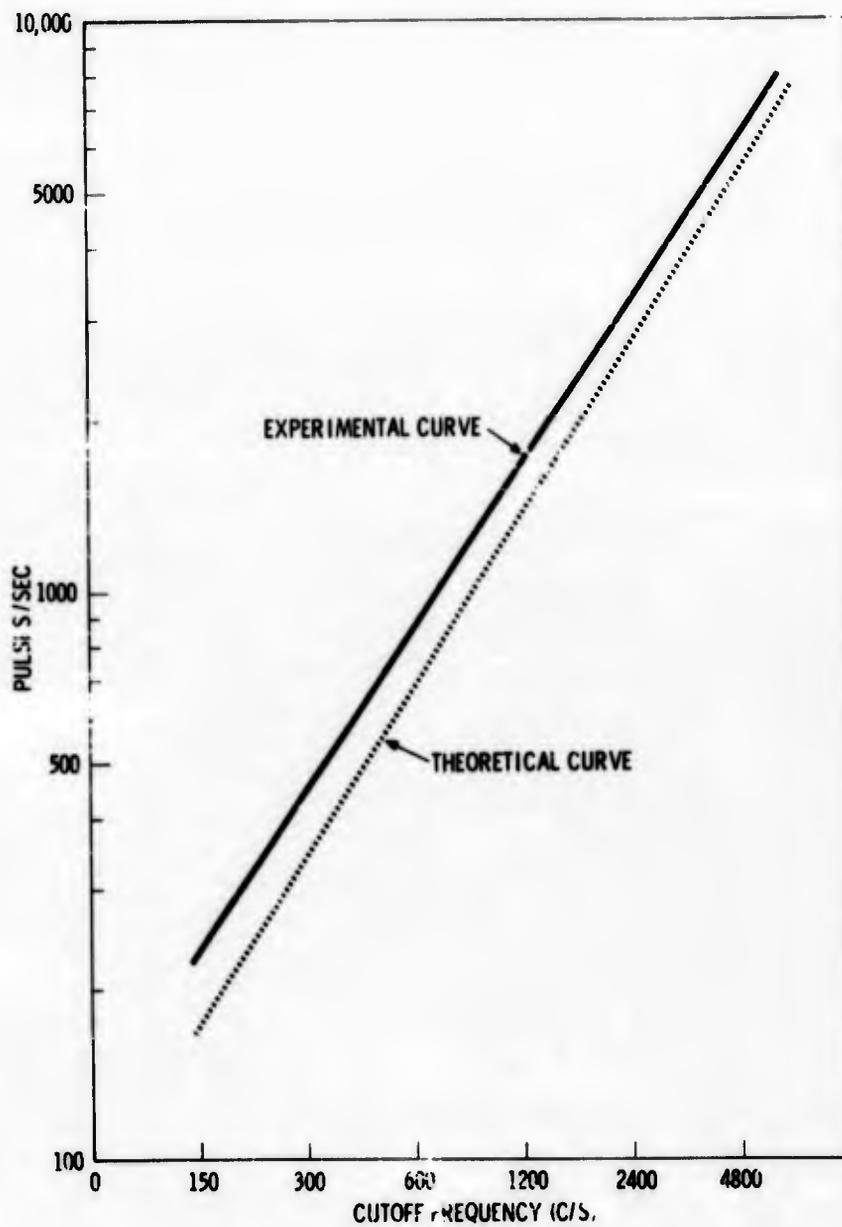


Figure B1. Illustration of the number of times that random noise goes into an interval about $x = 0$ for a PD of 0.4, vs. the cutoff frequency of low-pass filter.

Navy Electronics Lab., San Diego, Calif.
Report 1248

PROBABILITY DENSITY ANALYSIS OF OCEAN AMBIENT AND SHIP NOISE, by M. A. Calderon, 45 p., 5 Nov 64.

UNCLASSIFIED

Analysis of ocean ambient-noise samples indicated that such noise, under certain conditions, can be assumed to have a gaussian distribution of amplitudes. Similar tests showed that ship-noise samples are not gaussian. In the analytical procedure, the method of moments was found to be the most suitable for providing accurate and useful estimates of the parameters.

1. Underwater noise -
Mathematical analysis

I. Calderon, M. A.

SR 004 03 01, Task 8119
(INCL 2-4)

This card is UNCLASSIFIED

Navy Electronics Lab., San Diego, Calif.
Report 1248

PROBABILITY DENSITY ANALYSIS OF OCEAN AMBIENT AND SHIP NOISE, by M. A. Calderon, 45 p., 5 Nov 64.

UNCLASSIFIED

Analysis of ocean ambient-noise samples indicated that such noise, under certain conditions, can be assumed to have a gaussian distribution of amplitudes. Similar tests showed that ship-noise samples are not gaussian. In the analytical procedure, the method of moments was found to be the most suitable for providing accurate and useful estimates of the parameters.

1. Underwater noise -
Mathematical analysis

I. Calderon, M. A.

SR 004 03 01, Task 8119
(INCL 2-4)

This card is UNCLASSIFIED

Navy Electronics Lab., San Diego, Calif.
Report 1248

PROBABILITY DENSITY ANALYSIS OF OCEAN AMBIENT AND SHIP NOISE, by M. A. Calderon, 45 p., 5 Nov 64.

UNCLASSIFIED

Analysis of ocean ambient-noise samples indicated that such noise, under certain conditions, can be assumed to have a gaussian distribution of amplitudes. Similar tests showed that ship-noise samples are not gaussian. In the analytical procedure, the method of moments was found to be the most suitable for providing accurate and useful estimates of the parameters.

1. Underwater noise -
Mathematical analysis

I. Calderon, M. A.

SR 004 03 01, Task 8119
(INCL 2-4)

This card is UNCLASSIFIED

Navy Electronics Lab., San Diego, Calif.
Report 1248

PROBABILITY DENSITY ANALYSIS OF OCEAN AMBIENT AND SHIP NOISE, by M. A. Calderon, 45 p., 5 Nov 64.

UNCLASSIFIED

Analysis of ocean ambient-noise samples indicated that such noise, under certain conditions, can be assumed to have a gaussian distribution of amplitudes. Similar tests showed that ship-noise samples are not gaussian. In the analytical procedure, the method of moments was found to be the most suitable for providing accurate and useful estimates of the parameters.

1. Underwater noise -
Mathematical analysis

I. Calderon, M. A.

SR 004 03 01, Task 8119
(INCL 2-4)

This card is UNCLASSIFIED

UNCLASSIFIED

UNCLASSIFIED