CHAPTER 2

STATISTICAL ANALYSIS OF WAVE RECORDS

R.R. Putz
University of California
Berkeley, Calif.

ABSTRACT

The establishment of quantitative relationships between recorded wave-system characteristics and other phenomena requires numerical description of the wave record. Current concepts applied to time histories of wave activity at a point are discussed. Characteristic statistical regularities found in wave measurements are described. Examples given show the application of statistical techniques to the description of wave systems in terms of the distribution of spectral energy as well as the distributions of "individual" wave heights and periods. Results from the prediction, from one point to another, of surface time histories illustrate the application of approximate spectral information.

INTRODUCTION

In the study of waves and their interaction with their environment, there has been a need for effective description of observed wave systems. Convenient data (Snodgrass, 1951, 1952) is provided by the recorded time history, taken above some fixed point on the bottom, of the fluctuating pressure below the surface of or the water level at the surface. It will be understood that the length of a wave record selected for analysis is long relative to the periods of the fluctuations of interest, but short compared to the variations in meteorological conditions.

The questions we are concerned with are: What information is contained in a wave record? And how may this information be conveniently extracted? We shall first describe the statistical regularities suggesting a particular mathematical model (Tukey, 1950; Pierson 1952a) found useful for interpreting wave-record data. The two complementary aspects of this model each involve a distribution - one a statistical distribution of wave-record ordinates, the other a spectral distribution of energy. Theoretical relations connecting various model parameters will be illustrated with actual data. Although the experimental results to be presented have been obtained with pressure waves in the ocean for the most part, many of the methods and results are applicable to other kinds of data.

In the lower half of Fig. 1 is shown a short segment of a typical wave record, giving the pressure fluctuation 64 feet below the surface off the California coast. The total time interval shown represents about four and one-half minutes, during which from twenty to twenty-five waves pass the recording point. It will be seen that on the time-history curve the apparent slope and curvature, as well as the ordinate, vary irregularly from one instant to the next and bear little obvious relation to one another. The most noticeable feature may be the tendency for curvature and ordinate, measured from its average level, to have opposite signs.

The upper half of Fig. 1 shows a sample segment generated by the theoretical model in which suitable parameters were chosen to yield an artificial wave record comparable to the observed one.

DISTRIBUTIONS OF ORDINATES AND DERIVED QUANTITIES

Suppose a horizontal line is drawn cutting the wave-record curve at an arbitrary level. The fraction of the time the curve spends below such a line is a function (Birkhoff and Kotig, 1953; Putz, 1953b) increasing with the height of the line known as the distribution function for the curve. A plot of this function for a typical twenty-minute wave record is shown in Fig. 2. Here the number on the vertical scale represents the arbitrary chart level, the horizontal scale, the percent of time that the curve lies below that level. The notion of the probability of finding the curve below a given level may be introduced if one thinks of choosing at random an instant of time on the chart. The location of the plotted points on an approximate straight line is characteristic (Putz, 1953b) and results here from the choice of the horizontal scale which is the familiar Gaussian-distribution or normal-probability scale.

The interesting thing about wave records is the apparent ability of the normal probability scale to rectify very nearly not only the distribution curves for the ordinate and the first derivative, which have been experimentally checked(Rudnick, 1951; Putz, 1953b), but also, according to the theory, the distribution curves for derivatives of all orders. In each case the mean value below which the curve or derivative spends just fifty percent of its time will be zero. For the ordinates, this will be true because of our convention of measuring them from their mean. Since in this way one point is fixed on each curve, the straight-line plots for the various derivatives of the wave-record will differ among themselves only in their slopes. A convenient measure of the slope is the difference between the height of the line at the 84.1 percent level and its height at the 50 percent level. This distance, called the standard deviation, or root-mean-square (r.m.s.) ordinate, when measured for the k^{th} derivative, is denoted by σ_k .

The ordinate distribution concept may be extended by considering more than one ordinate at a time. Given n arbitrarily-selected chart levels, the n-dimensional generalization (Cramer, 1946) of the Gaussian distribution is then applicable to the fraction of the time that, simultaneously, n ordinates, chosen at different time instants on the wave record will lie below their corresponding chart levels. The assumption that for each finite n, the selection of n ordinates actually results in a multidimensional Gaussian distribution is known as the multinormal hypothesis. Such a distribution is characterized by a set of parameters known as the covariances (Cramer, 1946). These parameters may be readily interpreted in terms of the degree to which the various ordinates determine, or are determined by, each other. Computation (Rice, 1944-5) tells us that the covariance of any two quantities so distributed is proportional to cos($\pi \cdot P$), where P is the probability that the two quantities have opposite algebraic signs.

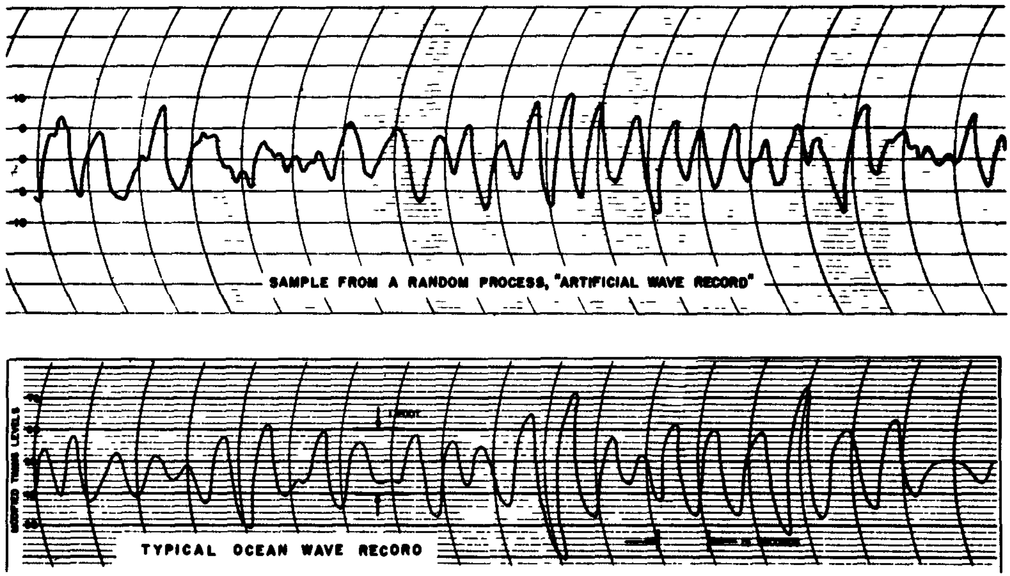


Fig. 1

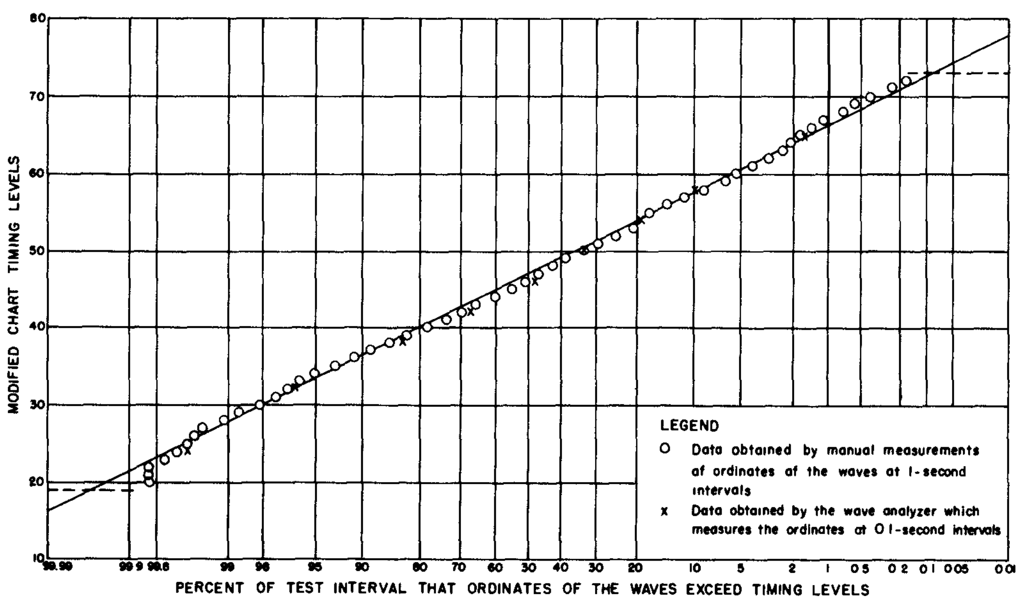


Fig. 2. Cumulative distribution function of ordinates.

A very convenient property of these multiple ordinate distritutions, and one which is physically reasonable except for relatively long records, is the stationary property (Lee, 1949; Putz 1953a), namely that the distribution of any set of ordinates depends only upon the differences in time between the corresponding abscissas. We then have distributions with equal covariances whenever the corresponding two ordinates are separated by the same time lag, τ, i.e., the covariance γ between any two ordinates selected at times t and $t+\tau$ depends only upon τ. The covariance is thus a function of the lag τ, the function $\gamma(\tau)$ being known as the covariance function, (Mann, 1953), and, when divided by its maximum value $\gamma(o)$, as the correlation function, $\rho(\tau)$, the latter being the familiar product-moment correlation coefficient. Once assumed for the ordinates, the stationary and multinormal properties follow for derivatives of all orders.

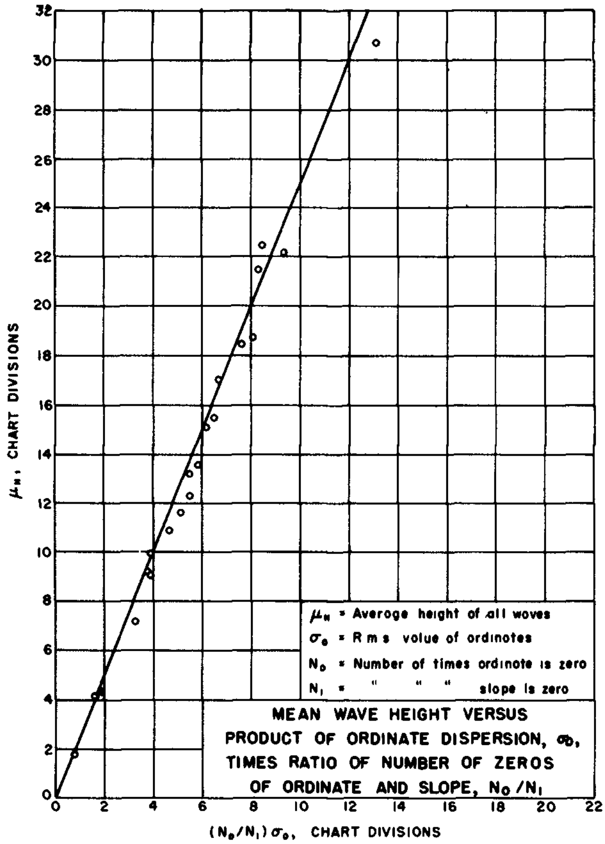
WAVE HEIGHTS

Interpretations for many of the parameters of these underlying probability distributions may be found in the distributions of certain quantities which may be graphically measured on the wave record.

If N_k denotes the number of times the k^{th} derivative passes through the zero level, and if T is the total length of the wave record in seconds, then the average number of zero-level crossings (up or down) per second will be N_k/T. A natural definition of the mean (undirectional) zero-crossing frequency is then $f_k = N_k/(2T)$ in cycles per second, or $\omega_k = 2\pi N_k/(2T) = \pi N_k/T$ in radians per second. Theory predicts that $\omega_k = \sigma_{k+1}/\sigma_k$, which serves to relate the r.m.s. values of the ordinate and of the first and second derivatives to the observed mean zero-crossing frequencies of the ordinate and the first derivative, when k is taken to be zero and one.

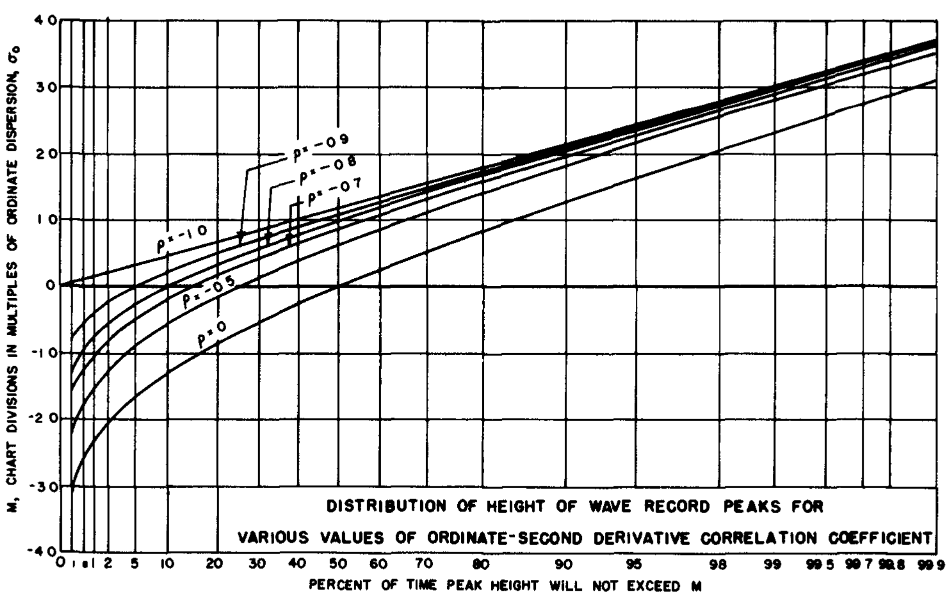
Further relations appear if we consider simultaneous values of ordinate and second derivative. The value of the coefficient of correlation between these two quantities is given by $\rho_o = \cos(\pi \cdot P_o)$, where P_o is the fraction of the time that they have opposite signs. Theory (Rice, 1944-5) predicts further that $\rho_o = -(\omega_o/\omega_1) = -(N_o/N_1)$. A satisfactory experimental check of the corresponding relation $\cos(\pi \cdot P_o) = -(N_o/N_1)$ has been obtained by locating points of inflection on a few twenty-minute records.

We may now consider the distribution of wave heights. For this purpose we shall take as fundamental the heights of the maximum points on the wave record curve, i.e., the points at which the second derivative is negative and the first derivative is zero. The theory applies also to minimum points if the wave-record curve is reflected about its mean level. If these peak heights, measured as directed distances from the mean level, are averaged, the resulting quantity μ_M is just one-half the average trough-to-crest wave height μ_H (Putz, 1952a). The theoretical relation is $\mu_H = (2\pi)^{\frac{1}{2}}(-\rho_o)\,\sigma_o$, illustrated for twenty-minute wave records in Fig. 3, where μ_H, measured in chart divisions, is plotted against $(N_o/N_1)\,\sigma_o$, also in chart divisions. The straight line shown has the slope $(2\pi)^{\frac{1}{2}}$ given by the theory. It is seen that the average of the so-called individual wave heights (Putz, 1953b) can provide



μ_H = Average height of all waves
σ_0 = R m s value of ordinates
N_0 = Number of times ordinate is zero
N_1 = " " " slope is zero
MEAN WAVE HEIGHT VERSUS PRODUCT OF ORDINATE DISPERSION, σ_0, TIMES RATIO OF NUMBER OF ZEROS OF ORDINATE AND SLOPE, N_0/N_1
μ_H, CHART DIVISIONS
$(N_0/N_1)\sigma_0$, CHART DIVISIONS

Fig. 3.



DISTRIBUTION OF HEIGHT OF WAVE RECORD PEAKS FOR VARIOUS VALUES OF ORDINATE-SECOND DERIVATIVE CORRELATION COEFFICIENT
$\rho = -1.0$
$\rho = -0.9$
$\rho = -0.8$
$\rho = -0.7$
$\rho = -0.5$
$\rho = 0$
M, CHART DIVISIONS IN MULTIPLES OF ORDINATE DISPERSION, σ_0
PERCENT OF TIME PEAK HEIGHT WILL NOT EXCEED M

Fig. 4.

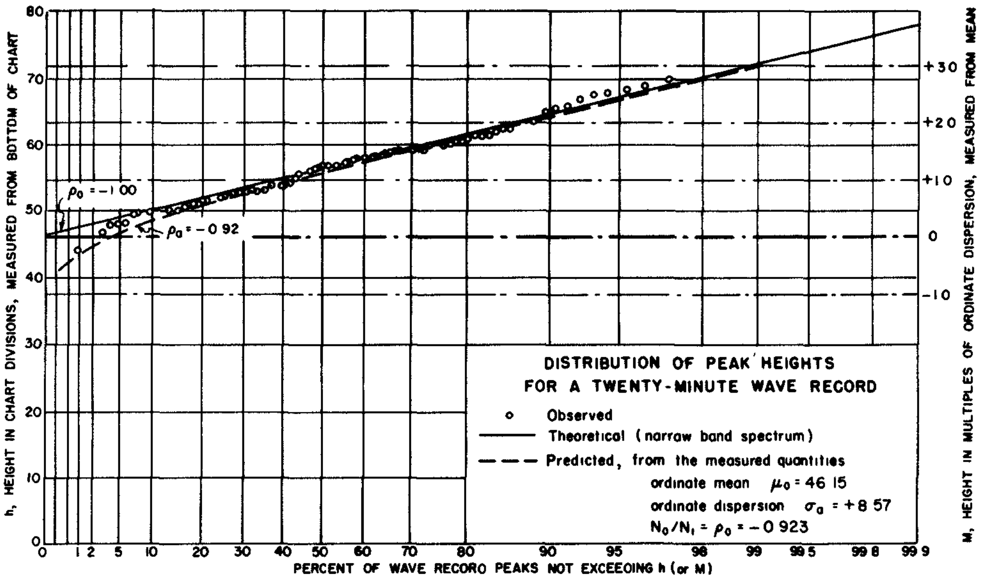

DISTRIBUTION OF PEAK HEIGHTS
FOR A TWENTY-MINUTE WAVE RECORD
Observed
Theoretical (narrow band spectrum)
Predicted, from the measured quantities
ordinate mean μ_0 = 46 15
ordinate dispersion σ_a = +8 57
$N_0/N_1 = \rho_0$ = −0 923
ρ_0 = −1 00
ρ_a = −0 92
h, HEIGHT IN CHART DIVISIONS, MEASURED FROM BOTTOM OF CHART
M, HEIGHT IN MULTIPLES OF ORDINATE DISPERSION, MEASURED FROM MEAN ORDINATE
PERCENT OF WAVE RECORD PEAKS NOT EXCEEDING h (or M)

Fig. 5

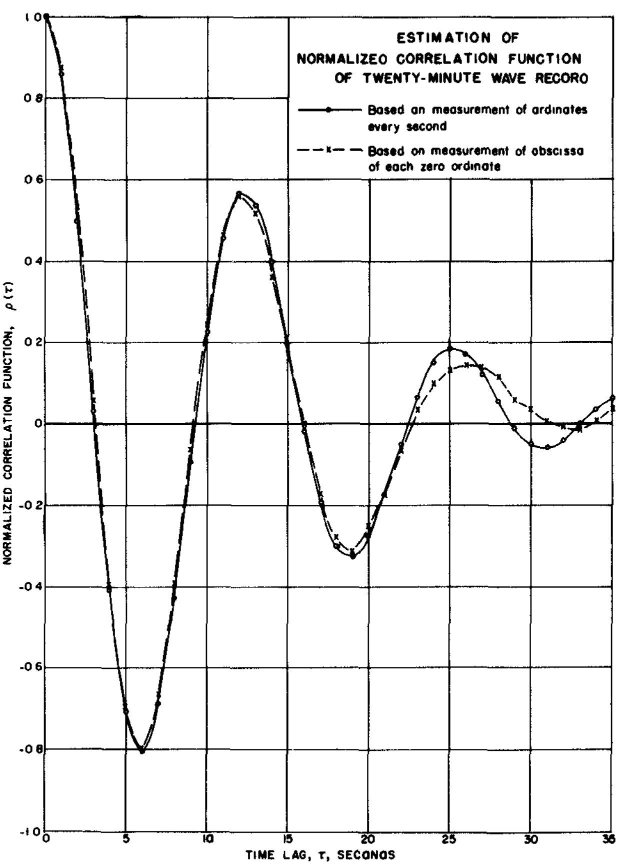

ESTIMATION OF
NORMALIZED CORRELATION FUNCTION
OF TWENTY-MINUTE WAVE RECORD
Based on measurement of ordinates every second
Based on measurement of obscissa of each zero ordinate
NORMALIZED CORRELATION FUNCTION, $\rho(\tau)$
TIME LAG, τ, SECONDS

Fig. 6.

a good estimate of the r.m.s. ordinate when combined with a count of the number of times the ordinate and the slope pass through zero.

The entire distribution of peak (or trough) heights (Rice 1944-5) given by the theory, depends only upon σ_0 and ρ_0, and is shown in Fig. 4. The vertical scale corresponds to the chart level measured from the mean in σ_0- units equal to the r.m.s. ordinate. The percent of peak heights not exceeding this vertical level appears on the horizontal, the scale in this case being chosen to correspond to the so-called Rayleigh distribution (Longuet-Higgins 1952; Lawson and Uhlenbeck; 1950, Knudtzon, 1949) which the peak heights follow more and more closely as the parameter ρ_0 tends to the value minus one. It may be observed that the probability of a low peak situated below the mean level is just $\frac{1}{2}$ $(1+\rho_0)$, which tends to zero as ρ_0 tends to minus one, corresponding to a relatively narrow-band spectrum.

Fig. 5 shows a typical observed distribution of peak heights for a twenty-minute wave record. The straight line represents the asymptotic Rayleigh distribution, while the curve represents the distribution of peak heights actually predicted from measured estimates of σ_0 and ρ_0, as well as of the mean ordinate μ_0. The vertical level is indicated in chart divisions on the left and in σ_0-units on the right.

ENERGY DISTRIBUTION

The statistical analysis of the periods between the zero crossings, or between the peaks on a wave record is closely related to the distribution of energy over the frequency or period spectrum. This assignment of energy to a continuum of superimposed elementary waves is specified by the so-called spectral distribution function or spectrum. The spectrum provides a means of deriving and interpreting the parameters of the distributions of wave-record ordinates and derivatives entering into the distribution of wave heights.

The distribution of energy is of fundamental importance, of course, for the study of the generation, propagation and the effects of waves. While information regarding the frequency of occurrence of wave heights or periods of given magnitudes appears occasionally to be directly useful for design purposes, a knowledge of the spectrum can be used whenever the frequency-response function corresponding to a general filtering action is known.

The usual spectrum analysis of a given wave-record sample results in the assignment of an amplitude and, in addition, a phase angle to each of a certain set of frequencies. However, the direct dependence of such phase information on the time origin to which the spectrum analysis is attached, makes it unsuitable as a possible parameter for a stationary random process. Thus it is natural to take as fundamental the remaining aspect of the spectrum -- i.e., the amplitude, or equivalently, the square of the amplitude, the so-called power spectrum (Rice, 1944-5).

There is a function obtainable from a sample wave record known as the sample covariance function which has for its spectrum just the power spectrum of the wave record itself. In many cases the power spectrum of a given sample wave record is most naturally obtained by first computing the covariance function for the wave record which was observed. However, any problem whose solution involves the prediction of the possible sample wave records which might have been observed before or after the given one will naturally lead to the association of a probabilistic model with each sample wave record. This model may be thought of as comprising a whole set of possible, infinitely long, wave records, each having the same power spectrum or covariance function. The identification of this common covariance function with that described earlier in connection with stationary multinormal ordinate distributions is the step which supplies the relation between these probability distributions and the spectral-energy distribution. The specification of the covariance function, together with the stationary multinormal assumption, defines a so-called stationary Gaussian random (or stochastic) function (or process). (Mann, 1953; Lee, 1950).

The ability of such a theoretical model to describe observed wave records is illustrated by comparing a sample taken from such a random process with a typical wave record. The upper curve in Fig. 1 was generated after first specifying three fixed numbers defining the spectrum and then consulting a table of random numbers (Wold, 1948).

If the mathematical moments of the spectrum exist (Putz, 1953a), certain of these can be shown to be equivalent to the r.m.s. values of the derivatives of the original wave record, while certain ratios of these moments are essentially the coefficients of correlation between the derivatives. Thus the r.m.s. ordinate σ_0 is the zero-order moment of the spectrum or the total spectral energy, while ρ_0 is related to the second-order spectral moment, or more precisely, to a certain measure of the relative spectral bandwidth when the latter is small.

WAVE PERIODS

It has been seen that the instants of time at which the wave record passes through zero or through a peak enter, by way of the relative spectral bandwith, into a relation between the wave heights and the r.m.s. ordinate. For the analysis of the spectrum, the differences between two successive zero-crossings may be considered the fundamental individual wave periods. According to the theory, for a long record the list of these individual wave periods in the order of their occurrence is completely equivalent to the spectrum. Peak-to-trough or trough-to-peak time intervals, corresponding to the periods between successive zero-crossings of the wave-record derivative, while more sensitive to the influence of noise, are capable in principle of yielding the same information. On the other hand, peak-to-peak or trough-to-trough time intervals (Snodgrass, 1952), corresponding to the periods between alternate zero-crossings of the derivative, would be expected to contain less information.

The significance of the zero-crossings for our model lies in the fact that the abscissas of these points separate the intervals during which the wave record has constant sign. Using these abscissas, we may estimate the correlation function for a given wave record by first determining the fraction $P(\tau)$ of the time that the original wave record and the wave record advanced by a time interval τ have opposite sign. It is from the correlation function $P(\tau) = \cos[\pi P(\tau)]$ that the spectral information is obtained.

Fig. 6 shows a comparison of the estimate of $\rho(\tau)$ obtained for a twenty-minute record by measuring zero-crossing abscissas and the ordinary estimate (Tukey, 1949; Pierson & Marks) obtained by measuring ordinates every second. This agreement between correlation functions has been found sufficiently close to result in a relatively small difference between the corresponding estimates of the wave-record power spectrum from the two methods. It may be observed that in addition to measuring 1200 ordinates on the wave record, nearly 1200 multiplications and 1200 additions are required to obtain each point of the solid-line curve, while for each point of the dotted line curve it is sufficient to measure about 200 abscissas on the wave record and perform about 400 additions and 200 subtractions. While either method leads to a substantial saving in labor over a straightforward numerical Fourier analysis of the original wave record (James, Nichols and Phillips, 1947); the zero-crossing method for the correlation function, requiring essentially no multiplication, may be carried out with only an adding machine.

The degree of reliability of the zero-crossings exhibited by this one experimental check on a twenty-minute wave record may be of some theoretical interest. The close correspondence, particularly for small time lags, seen in Fig. 6 is additional evidence for the general mathematical model we have described, and for the multinormal property in particular. It constitutes evidence also for the sampling adequacy of a wave record twenty minutes in length. The indications are that the so-called individual wave periods, suitably defined, can furnish reliable information about the Fourier spectrum of the wave record.

This kind of approximate spectral information has been used to predict, by a least-squares method (Putz, 1952b), the behavior of ocean waves at one time and place from their behavior at a time and place nearby. The resulting prediction is shown in Fig. 7, where the upper plot represents the time history of the observed surface elevation at a point 1570 feet seaward of the point at which prediction was to be made. The lower part of the figure shows the observed and the predicted wave record at the shoreward point fifty seconds later. The fact that the accuracy of prediction represented here compares favorably with that based upon a more exact spectrum analysis is a further indication of the reliability of the information provided by the zero-crossings.

SUMMARY

There exists an approximate mathematical model for wave records which may be profitably exploited until refinements in it are found necessary. This model specifies simple verifiable relations which may be

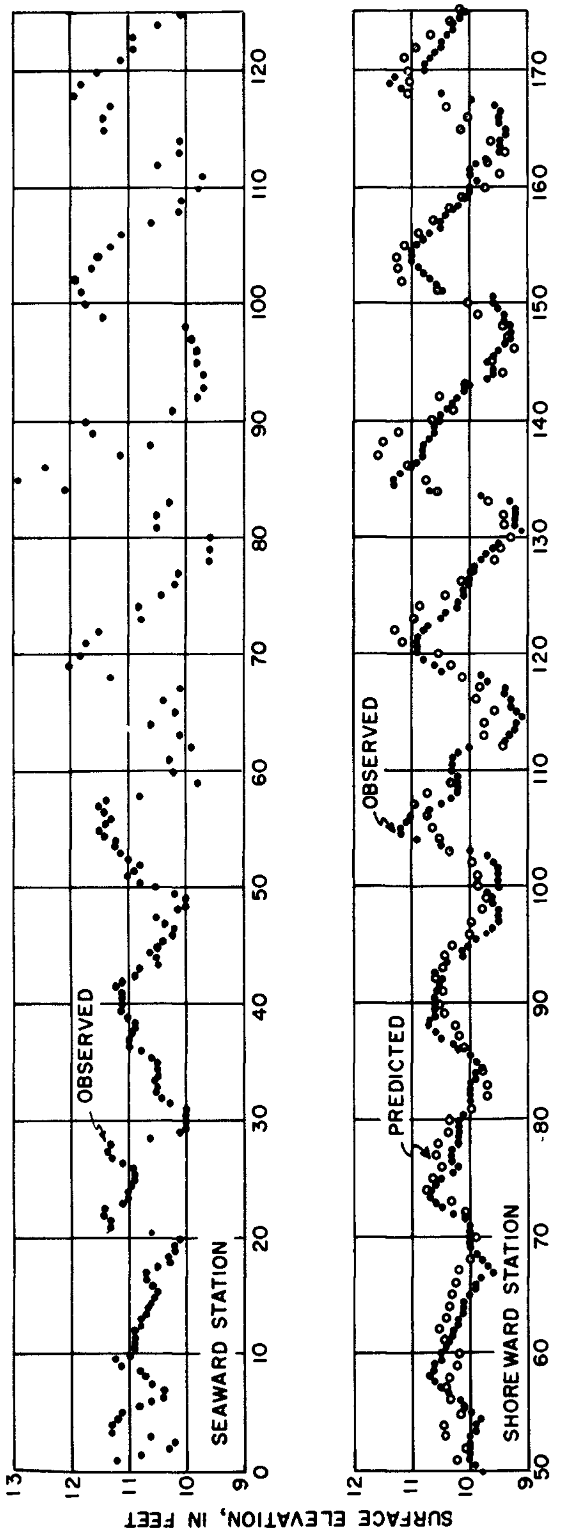


Fig. 7. Least-squares surface prediction over 1570 ft. distance (Four-term kernel based upon approximate spectral information from both positions).

expected to hold between appropriate quantities easily measured on long wave records. The parameters occurring in the model may be estimated from individual wave-height and wave-period distributions. A few such parameters, easily interpreted in terms of the spectrum, are available for the description of a wave record.

REFERENCES

Birkhoff, G. and Kotik, J., (1953). Fourier analysis of wave trains: Proceeds. Nat. Bu. of Standards Semicentennial Symp. on Gravity Waves (1950) pp. 221-234, Govt. Print. Off. Wash. D.C.

Cramer, H., (1946). Mathematical methods of statistics; Princeton University Press, Princeton, N.J.

James, H.M., Nichols, N.B., and Phillips, R.S. (1947). Theory of servomechanisms, McGraw-Hill Book Co. Inc., New York, N.Y.

Knudtzon, N., (1949) Experimental study of statistical characteristics of filtered random noise, Tech. Report No. 115, Research Lab. of Electronics, Mass. Inst. of Tech.

Lawson, J.L. and Uhlenbeck, G.E. (1950). Threshold signals, McGraw-Hill Book Co. Inc., New York, N.Y.

Lee, Y.W. (1950). Communication applications of correlation analysis: Symposium on applications of autocorrelation analysis to physical problems (1949), Off. of Nav. Research, Wash. D.C.

Longuet-Higgins, M.S.(1952). On the statistical distribution of the heights of sea waves. Jour. of Marine Res., v. 11, pp 245-266.

Mann, H.B. (1953), Introduction to the theory of stochastic processes depending on a continuous parameter. Applied Math. Series No. 24, Nat. Bu. Standards, Wash. D.C.

Pierson W.J., Jr. (1952a), A unified theory for the analysis, propagation and refraction of storm-generated ocean surface waves, Pt. I. Res. Div., College of Engr., New York University, N.Y.

Pierson W.J., Jr. and Marks, W. (1952). The power spectrum analysis of ocean wave records. Trans. Amer. Geophys. Union, v. 33, pp. 834-844.

Putz, R.R. (1952a), Statistical distributions for ocean waves; Trans. Amer. Geophys. Union, v. 33, pp. 685-692.

Putz, R.R. (1952b). Predictability of wave transformation: Linear least squares prediction. Section G, Tech. Report Ser. 29, Issue 54, Inst. of Engr. Res., University of Calif., Berkeley, Cal.

Putz, R.R. (1953a) The analysis of wave records as random processes; Presented at Pacific S.W. Regional Meeting, Am. Geophys. Union, Feb 6-7 1953. also Tech. Report Series 3, Issue 352, IER., University of California, Berkeley.

Putz, R.R. (1953b). Ocean-wave record analysis - ordinate distribution and wave heights, Tech. Report Ser. 3 Issue 351, IER. University of Calif. Berkeley, Cal.

Rice, S.O. (1944-5), Mathematical analysis of random noise, Parts I-IV, Bell System Tech. Journal vol. 23, pp. 282-332, v. 24, pp. 46-156.

Rudnick, P. (1951) Correlograms for Pacific Ocean waves. Proceedings of the Second Berkeley Symp. on Mathematical Statistics and Probability, pp. 627-638. Univ. of Calif. Press. Berkeley, Cal.

Snodgrass, F.E. (1951) Wave recorders; Proceedings of the First Conf. on Coastal Engr. (1950), Council on Wave. Res., The Engr. Foundation, J.W. Johnson, ed., Univ. of Calif., Berkeley.

Snodgrass, F.E. (1952) Ocean wave measurements; Symposium on Oceanographic Instrumentation (1952), Nat. Research Council.

Tukey, J.W., (1950). The sampling theory of power spectrum estimates. Symp. on Applications of Autocorrelation Analysis to Physical Problems (1949). Off. of Nav. Research, Wash. D.C.

Wold, H. (1948) Random normal deviates. Cambridge University Press.