Problems in Statistical Analyses of Geophysical Time Series[1]

H. R. Seiwell

Woods Hole Oceanographic Institution, Woods Hole, Massachusetts

AS AN EXAMPLE OF GEOPHYSICAL TIME SERIES we concern ourselves with sea surface wave data, the preparation of this article having been prompted by a recent paper on ocean waves by G. E. R. Deacon (*3*). Deacon and his co-workers have carried out extensive investigations of sea surface wave phenomena at Pendeen, England, from which they conclude that a heterogeneous system of waves is transmitted to a distance as a group and can there be analyzed. Records of wave height against time at fixed locations are mechanically analyzed by a periodogram technique, and the sea surface wave patterns are then represented as a number of frequencies of varying amplitudes. In actual practice, the Pendeen wave records, each of approximately 20 minutes' duration, are subjected to analysis on a specially constructed ocean wave frequency analyzer. The procedure is believed to isolate the constituent elements, whose sum composes the series of the sea surface wave pattern.

The wave period (or wave frequency) analysis of the Pendeen records applies a periodogram technique to data of finite scope as it comes from nature. The physical significance of the results of this analysis is believed to be shown by the general outline of the resulting spectrum. In particular, significance is attached to the occurrence of a "prominent frequency band," which persists through a number of consecutive spectra. This is interpreted to mean that waves between the indicated period limits are physically present in the sea surface wave patterns, discrete frequencies being indicated by individual peaks.

As an example of the interpretation, we refer to published wave spectra of Pendeen sea records dated 0500 to 1900, July 1, 1945 (*1*). Each of the spectra shows a prominent frequency band containing 25 or more individual peaks between 7 and 14 seconds (approximately). According to the Pendeen interpretation, this indicates the presence of a band of surface waves with periods between 7 and 14 seconds that has traveled from a distant storm-generating area, each individual wave (identified by the peak in the spectrum) having a velocity proportional to its period. The physical postulate is important, and, as alternate ones exist, we shall briefly examine the extent to which one may reasonably formulate a physical hypothesis from results of a mathematical analysis of a geophysical time series.

The problem of searching for hidden periodicities in natural and economic time series is an old one for which no satisfactory general solution appears to have yet been found (*2*). Mathematical methods, based on theory of infinite series, generally require some modification when applied to natural phenomena of finite scope, because the system, though safe for the problem for which it was developed, may become unsafe in other uses. Investigations of time series have, in particular, been influenced by the classical work of Schuster (*8*), who introduced probability theory into a method for determining hidden periodicities, developed for a specific purpose. The method of Schuster or, more generally, one of its modifications (for instance, that of Whittaker and Robinson [*12*]) has been used in many types of times series analysis, and in such a fashion as to suggest the physical occurrence of periods and cycles in numerous economic and natural time series. In contrast, we have the opinion of statisticians that the method of periodogram analysis as a means of searching for periodicities appears rarely, if ever, to have been successful in geophysical and economic work. Although exact for infinite series, it appears to give insufficient information when applied to observational data of finite scope.

The methodology of searching for periodicities has, in particular, been discussed at length by Bartels (*2*), Wilson (*13, 14*), Yule (*15*), and Kendall (*4, 5, 6*).

[1] Contribution No. 505 from the Woods Hole Oceanographic Institution and the Office of Naval Research, Contract N6ONR-277/I.

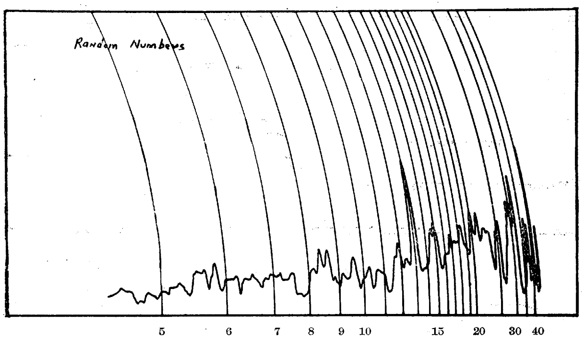

FIG. 1. Frequency spectrum of Tippett's random numbers from periodogram analysis by the ocean wave analyzer. (Period of waves in seconds.)

Wilson, in an elaborate series of experiments with the Schuster periodogram technique, cast doubt on the validity of periods or cycles in the *Ayres' Index of American Business Activity.* More recently, the experimental approach of Kendall appears to confirm the unreliability of the Schuster and related techniques of cycle analysis as a means of discovering from natural data the fundamental laws by which they were generated. Thus, Kendall's investigations reasonably demonstrate that the spectra of periodogram analyses of data having finite scope—that is, when limited (and workable) amounts of data only are available—may not only fail to indicate the true character of the oscillation, but also show unexplained spurious peaks not related to the physical character of the phenomenon dealt with. On the other hand, we find Barber and Ursell (*1*) proposing a theoretical justification of the Pendeen periodogram technique, but without experimental demonstration of their results.

The above points are illustrated by periodogram analyses of two experimental examples (Figs. 1, 2). The periodogram spectrum of each model was obtained by use of a wave frequency analyzer basically similar to the Pendeen model (*7*). Fig. 1 is the spectrum given for Tippett's (*11*) random numbers, which were plotted and then painted on a tape to resemble a natural wave record. Fig. 2 is the analyzer spectrum of the function 2.5 sin $\frac{2\pi}{12}+\varepsilon$, arranged in the same fashion. Either spectrum resembles that obtained from similar analysis of an ocean wave record. The random number spectrum shows well-defined teeth (periods?) at 13 and 30 seconds bounding a reasonably prominent "frequency band." The second function is revealed as a 12-second period, being part of a well-defined "period band" between 11 and 17 seconds, together with other well-defined peaks for higher periods. The examples demonstrate the danger of deducing physical properties of ocean wave records from periodogram spectra.

The foregoing discussion may appear somewhat destructive in that, if the main conclusions are correct, they indicate that inferences based on periodogram analysis of ocean wave data may require some reconsideration. Methods of statistical inference, whereby we hope to determine the physical properties of data that are finite in scope (and not excessively long) in order to discover something of the physical causes, require information beyond that given by the periodogram. In certain situations, computation of the autocorrelation function and plotting of the correlogram appear to provide this additional information. It was to this end that Seiwell and Wadsworth (*10*) applied the method of autocorrelation analysis to ocean wave records (wave height variations with

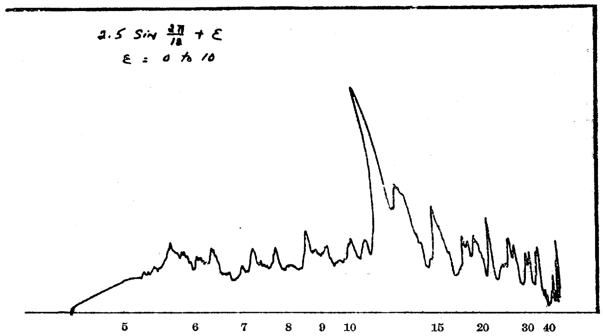

FIG. 2. Frequency spectrum of the function 2.5 sin $\frac{2\pi}{12}+\varepsilon$ from periodogram analysis by the ocean wave analyzer. (Period of waves in seconds.)

time) and, in their earliest analyses, found situations where the sea surface wave pattern was referable to one main trigonometric plus a residual with a random component. From this it was postulated that the main trigonometric was "sea swell" generated at a distance, and that the remainder of the pattern, "sea," was generated locally. This differed from current interpretations of sea surface wave patterns which, from results of periodogram analyses, were interpreted to be a combination of waves of many frequencies generated by a distant storm, and which had independently traveled from that area to the target. The idea of a single frequency, only, reaching the target at any one time from a distant storm, and on which is superimposed locally generated disturbances of the same average frequency as the fundamental, had not been considered.

Hence, differences in the two postulates concern the physical origin of sea surface roughness patterns at a target distant from the generating area. Seiwell and Wadsworth infer that the ocean acts as a filter, and that, after a certain distance from the storm-generating area, swell of a single period is all that remains of the wave pattern. This single wave has been termed the dynamic component of the sea surface roughness pattern, and its period and amplitude change with time are referred to a distant target. On the other hand, the postulate supported by Deacon and his co-workers proposes that conditions of the generating area may be sampled at a distant target and subsequently evaluated. It is the opinion of this writer that weakness in the latter postulate lies in the statistical methods on which the physical inference is based.

Although it is apparent that the real physical picture of the sea surface roughness pattern will result only from extensive series of observations between generating areas and distant targets, the extreme difficulty and expense involved will, for the present, make

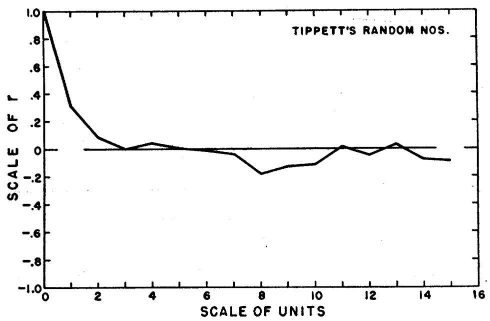

FIG. 3. Correlogram obtained from 250 of Tippett's random numbers.

it necessary to base knowledge of the physical properties of the sea surface primarily on wave observations from near-shore localities. In view of possible significance of inferences to be drawn from results of autocorrelation analysis, an experimental investigation of the autocorrelation function and the resulting correlogram, and of the power spectrum, was carried on parallel with correlogram analyses of ocean wave records from North Atlantic localities. Present indications are that, although the correlogram provides more information than the periodogram spectrum (of finite data), it is limited to determinations as to whether the basic series is random, periodic in one term only, periodic in more than one term, or completely autoregressive according to the scheme of Kendall.

In the practical computation[2] of the normalized autocorrelation function, r_k, from finite amounts of data (N), we use the formula

$$r_k = \frac{\sum_{i=1}^{N-k} \frac{X_i X_{i+k}}{N-k} - \frac{\left(\sum_{i=1}^{N-k} X_i\right)\left(\sum_{i=1}^{N-k} X_{i+k}\right)}{(N-k)^2}}{\sigma^2}$$

where σ^2 is the variance of the series. The correlogram is obtained by plotting r_k against k: r_0 is set equal to 1.

[2] Computation is now carried out on an especially designed mechanical autocorrelator.

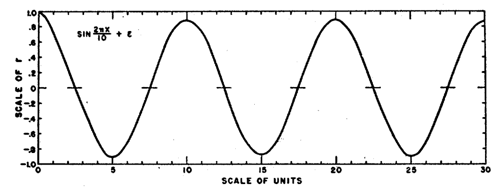

FIG. 4. Correlogram obtained from sin $\frac{2\pi}{10} + \varepsilon$.

For a random series, the correlogram is a strongly damped exponential (Fig. 3) of the form

$$r_k = e^{-\lambda k}$$

In the special case where the basic series contains a single sine wave, the correlogram is a cosine curve of that period, having equal distances between successive peaks and successive valleys (Fig. 4), and damping to a terminal amplitude, the value of which depends on the percentage variability of the basic trigonometric to the total data. On the other hand, when the basic series contains several cyclical terms, the correlogram will *not* have a constant period, distances between successive peaks and successive valleys are not constant, and damping is not consistent over more than two or three cycles. Distinction between the two types of series becomes apparent after two, and not more than four, cycles of the correlogram have been examined.[3] For this situation, additional information may sometimes be obtained by a Fourier transform of the autocorrelation function into its approximate power spectrum. In the special case of the basic data being completely autoregressive, the correlogram damps to zero after a few cycles.[4]

Correlograms of the two series previously discussed (Figs. 1, 2) are illustrated by Figs. 3 and 4. Fig. 3 is the correlogram of a series of 250 equally spaced random numbers (Tippett). The damping factor λ is about 1.2. The correlogram shows some persistence and small correlation in that the numbers themselves are not strictly random. However, there is no suggestion of periodicity, a sharp contrast to the peaks or teeth brought out in the periodogram spectrum of this same series (Fig. 1).

As a second example of a correlogram application to finite data, the function $y = A \sin \frac{2\pi}{10} + \varepsilon$ is considered. A had a value of 1, and ε was drawn from Tippett's random numbers, as before. The correlo-

[3] This is based on unpublished experimental data on file at this Institution.

[4] For discussion of properties of autoregressive series, see the various works of M. G. Kendall (*4, 5, 6*). Only one case of completely autoregressive wave data has been found by this author, this being a record taken off Scripps Pier, La Jolla, Calif., dated November 5, 1948.

gram computed from 250 equally spaced digits (Fig. 4) reveals the period of the original trigonometric. In the first cycle, it damps to a value $r_T = 0.88$ and thereafter remains nearly constant; at the third cycle $r_T = 0.86$ (within limits of computational error). If the function had been unknown and given only as a series of values, the foregoing analysis permits computation of the following properties.

Period, $T = 10$ units

Variance cosine, $\frac{A^2}{2} = 0.500$

Total variance, $\sigma_y^2 = 0.588$

Residual variance, $\sigma_r^2 = 0.088$

Terminal amplitude, $r_T = \frac{\text{variance cosine}}{\text{total variance}} = 0.85$

Total standard deviation, $\sigma_y = 0.7668$

Total average deviation, $AD_y = .656$

$\frac{AD_y}{\sigma_y} = 0.856$

When a cosine wave of 10 units was fitted to the basic data ($y = A \sin \frac{2\pi}{10} + \varepsilon$), the least square amplitude was computed to be $A' = 0.991$, nearly identical with the original amplitude of $A = 1.00$.

Thus, if the above series had been a natural wave record, we would infer the following characteristics: The sea surface wave pattern consisted of one fundamental wave (swell) with a period of 10 seconds and an amplitude of 1 foot on which has been superimposed a series of local impulses (sea). The former, presumably generated by a meteorological situation acting at some distance from the target, is the predictable component in the data and accounts for an estimated 85 percent of the total variability of the sea surface wave pattern. It is this component that may be forecast from knowledge of wind action on the sea surface. The remainder of the data represents a series of superimposed local impulses, generated by local winds, which die out rather rapidly when the energy source is removed.

The correlogram of the above model is typical of that sometimes obtained from ocean wave data. Eight examples of such analysis have been published (*9*), in two of which (53-X and 39-L) correlogram analyses were carried out over 17 and 124 cycles, respectively. In each case the amplitude of the single trigonometric was determined by a least square fit of a cosine curve of the period isolated by the correlogram. It is suggested that a quantitative distinction between swell generated at a distance and the locally produced sea is possible. When only one period is present in the basic series, it is to be expected that it changes with time and at a rate dependent on the stability of the offshore causative meteorological situation. On the other hand, the local sea is induced by local winds and other factors tending to disturb the sea surface, and it dies out fairly rapidly.

In the foregoing, some of the problems associated with geophysical time series analysis have been briefly discussed. Hypotheses for both periodogram and correlogram analysis were originally developed for problems of infinite series, and the background mathematical theories are understood. Uncertainty in their application arises when the object is to discover from experimental or observational data (of finite scope) something about the physical laws by which they were generated. Hence, prior to extensive use of either method, it appears pertinent that mathematical models of a type known to occur in nature be studied, both experimentally and theoretically, to develop ideas that will help to define the problems and to identify the basic laws underlying particular sequences of data.

References

1. Barber, N. F., and Ursell, F. *Phil. Trans. Roy. Soc. Lond.*, 1948, **A240**, 527.
2. Bartels, J. *Terr. Mag.*, 1935, **40**, 1.
3. Deacon, G. E. R. *Quart. J. Roy. Met. Soc.*, 1949, **75**, 227.
4. Kendall, M. G. *J. Roy. Stat. Soc.*, 1945, **108**, 93.
5. ———. *Nat. Inst. Econ. Res., Occ. Papers IX*, Cambridge Univ., 1946.
6. ———. *The advanced theory of statistics*, Vol. II. London: Charles Griffin, 1948.
7. Klebba, A. A. *Ann. N. Y. Acad. Sci.*, 1949, **51**, 533.
8. Schuster, A. *Terr. Mag.*, 1898, **3**, 13.
9. Seiwell, H. R. *Proc. Nat. Acad. Sci.*, 1949, **35**, 518.
10. Seiwell, H. R., and Wadsworth, G. P. *Science*, 1949, **109**, 271.
11. Tippett, L. H. C. *Tracts for computers, No. XV*, Cambridge, 1927.
12. Whittaker, Sir Edmund, and Robinson, G. *The calculus of observations*, 4th ed. London: 1948.
13. Wilson, E. B. *Quart. J. Econ.*, 1934, 375.
14. ———. *Science*, 1934, **80**, 193.
15. Yule, G. U. *Phil. Trans. Roy. Soc.* Lond., 1927, **A226**, 267.