A SIMPLE FORM OF PERIODOGRAM

By Dinsmore Alter

Schuster's introduction of a method of systematic search for hidden periodicities and cycles opened a new field for the investigator of statistical data. The beauty of his method in its analogy to analysis of light, and the great reputation of its author, combined to give it universal acceptance and to blind statisticians to its faults.

In more recent years at least three new mathematical and two mechanical forms of periodogram analysis have been proposed, each of which exhibits certain advantages over the original one. The use of the term *periodogram* for these forms is an extension of Schuster's original definition which used as abscissae quantities proportional to the squares of the amplitudes of the sine terms found in the data for the various trial periods. He wrote: "It is convenient to have a word for some representation of a variable quantity which shall correspond to the spectrum of a luminous radiation. I propose the word *periodogram* and define it more particularly in the following way:

Let $\frac{1}{2}Ta = \int_{t_1}^{t_1+T} f(t) \cos ktdt$ and $\frac{1}{2}Tb = \int_{t_1}^{t_1+T} f(t) \sin ktdt$

where T may for convenience be chosen equal to some integer multiple of $\frac{2\pi}{k}$, and plot a curve with $\frac{2\pi}{k}$ as abscissae and $r = \sqrt{a^2 + b^2}$ as ordinates; this curve, or better, the space between this curve and the axis of abscissae, represents the periodogram of $f(t)$."

The following appear to be the essential criteria for a satisfactory form of periodogram:

1. It must exhibit plainly any repetition of form in the data regardless of how irregular the shape of the repeated interval may be. In doing this it must exaggerate the amplitude of the main terms at the expense of the lesser ones.
2. The calculation of the indices must be short. In a periodogram from many data the indices sometimes are computed for several hundred trial periods.
3. There should be a geometrical interpretation of the index used.
4. The frequency distribution of the index must be known.
5. Combining or smoothing the data should modify the index in a manner which leaves an obvious interpretation.

The Schuster periodogram has the following disadvantages:

1. Only sine terms of large amplitude are exhibited. A perfect repetition of an extremely irregular form of data would not be indicated in any way.
2. The calculations are long.
3. There is a considerable uncertainty in the length of the period found. Those methods of analysis which use harmonics as well as the fundamental have much less of this uncertainty.

The correlation periodogram has advantages in each of these points over the Schuster. However, even with it the calculations are fairly long. Furthermore, the modification of the coefficient introduced by grouping or smoothing is not a linear one.

The periodogram described here is a slight modification of one for which a preliminary note was published in 1933. Additional features have been studied and its applications to many data have shown its ease of calculation. This calculation has been reduced still more by a mechanical method which renders it practicable to contemplate the possibility of studying many data hitherto prohibited by excessive cost.

Consider data x_0, x_1, x_2, $\cdots x_i$, $\cdots x_{(n-1)}$. Let l be any integer less than n. Form the sum of the absolute values of $x_i - x_{(i-l)}$, designated by $\sum | x_i - x_{(i-l)} |$. Define $A = \sum_{i=l}^{n-1} \frac{| x_i - x_{(i-l)} |}{n - l}$, l takes the values of the various trial periods and is called the *lag*. A, therefore, is the mean error between prediction that data will be repeated after a lag of l and the fulfillment of the prediction. Such an index has a meaning that is immediately of use to a meteorologist or other investigator. Coefficients such as the Schuster and the correlation coefficient, although valuable statistically, are of less immediate interest.

The standard deviation of these errors of prediction follows at once from standard formulae under assumption of normal distribution.

$$\sigma = 1.25\, A$$

The distribution of σ, as computed from the absolute values of data, has been studied by Helmert and by Fisher. Davies and E. S. Pearson have compared the various methods of estimating σ. For the large number, $(n - l)$, pairs of data used for a periodogram point, this method becomes almost as precise as the usual one which would square the values of $(x_i - x_{i-l})$. For $(n - l)$ as small as 50, the standard deviation of the standard deviation by this method is only seven percent larger than by the other one. Fisher has shown that

$$\sigma_\sigma \rightarrow \frac{\sigma}{\sqrt{n - l}} \sqrt{\frac{\pi - 2}{2}} \qquad \text{as } (n - l) \rightarrow \infty$$

This may be written as

$$\sigma_\sigma \rightarrow \frac{1.068\,\sigma}{\sqrt{2(n - l)}}$$

The distribution approaches normal rapidly and for all values of $(n - l)$ that would be used in periodogram calculation certainly may be considered as normal. It will be very seldom that a value of $(n - l)$ much smaller than 200 will be used.

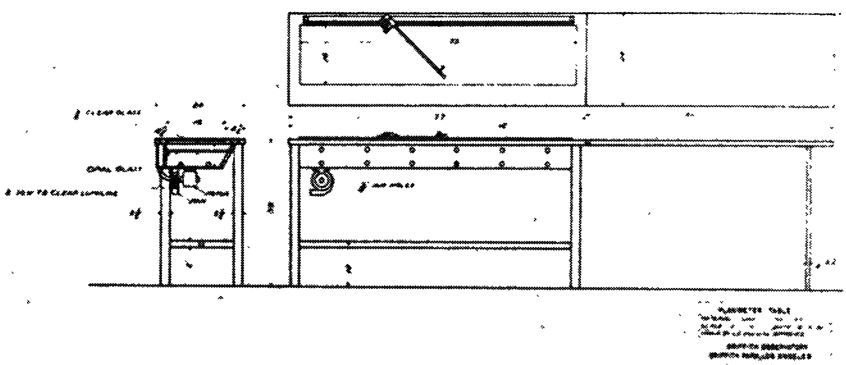
The data may be printed on two strips of adding machine tape held together by clips so as to match data separated by a lag l. In arranging them for investigation, it usually is most convenient to make all numbers positive. The computer subtracts mentally and puts the difference into an adding machine, which gives him A almost immediately.

For some computers, and especially where the numbers are large, another method of obtaining A may save time or lead to less numerical mistakes. The computer will form the sum of all his data. He will, as for the other form of computation, put these on two pieces of adding machine tape that he lays side by side. However, instead of putting the difference of the pairs into the machine, he will, in each case, put in the smaller datum of the pair. Then,

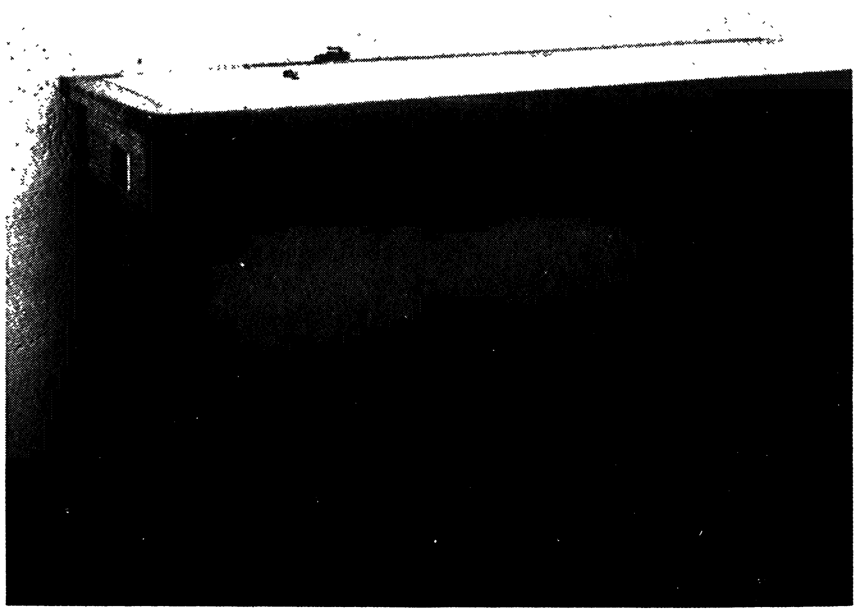
$$(n - l)A_l = 2 \sum \text{all data} - [\sum \text{1st } (n - l) + \sum \text{last } (n - l) \text{ data}] - 2 \sum \text{smaller}$$

The derivation of this equation is obvious. In computing by this method the subtotaler on the machine can be used to make the strip of sums of the first $(n - l)$ data and of the last $(n - l)$ for all values of l. The first term on the right hand side is a constant, the last is twice the sum of the smaller numbers chosen in the pairs. I have computed by both methods, and where the numbers are small, I prefer the former. Where they are large, I prefer the latter. However, when one must use comparatively untrained computers, he will find less mistakes made if the computer does not make the subtractions.

The calculation of A is much shorter than that for the indices even of the correlation and variance periodograms. It may, however, be shortened even more by a mechanical arrangement. $(n - l)A_l$ is the area between two histograms of the data matched after a lag l. These may be carefully graphed on a large scale and two such graphs superposed over a table with a translucent illuminated top. On the edge of this table is the track to guide a rolling planimeter. A, as computed by this means, is accurate to approximately one-half of one percent of its value, a much more exact value than is needed. The details of such a device as constructed for the Griffith Observatory are shown by the accompanying photograph and diagram. The dual saving of time by the method and by its mechanical application have resulted in the adoption of a much more ambitious program of meteorological research than previously was contemplated.



Scale Diagram of Planimeter Device



Planimeter Device for Mechanical Calculation

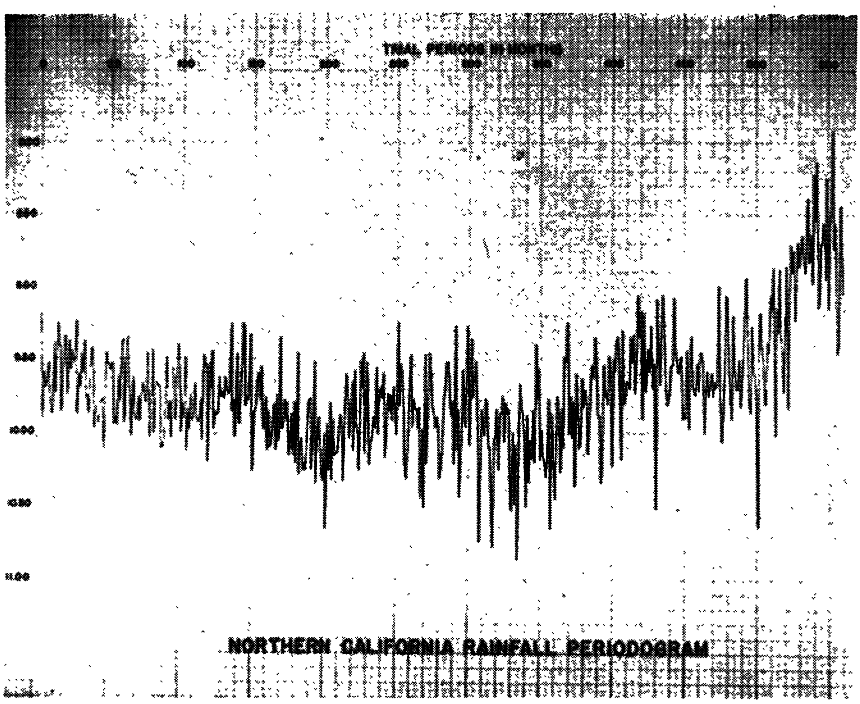
The form taken by the periodogram is important. Consider the simplest case, data which follow a sine curve.

$$y_i = a \cos\left(\frac{2\pi i - c}{p}\right)$$

$$y_i - y_{i-l} = 2a \sin\frac{\pi l}{p}\left[\sin\frac{2\pi(\frac{1}{2}l - i) + c}{p}\right]$$

The term in brackets takes values distributed around the circle and the part outside is a constant for any one lag. The bracket term sums approximately to $\frac{2(n - l)}{\pi}$, since we consider all terms as of one sign only.

$$\therefore A_l = \left|\frac{4a}{\pi}\sin\frac{\pi l}{p}\right|$$



If the absolute values were not considered in the expression for A_l, the periodogram would be a sine curve of period $2p$. The lack of sign gives a cusp curve with the cusp at lags p, $2p$, etc. Such a form is advantageous in that the periodogram gives sharp peaks at multiples of the periods which may exist.

The effect of the periodogram in exaggerating the principal terms at the expense of the smaller ones may be obtained most easily by equating σ as obtained by the linear and the quadratic formulae.

The data may be written as the sum of cosine terms

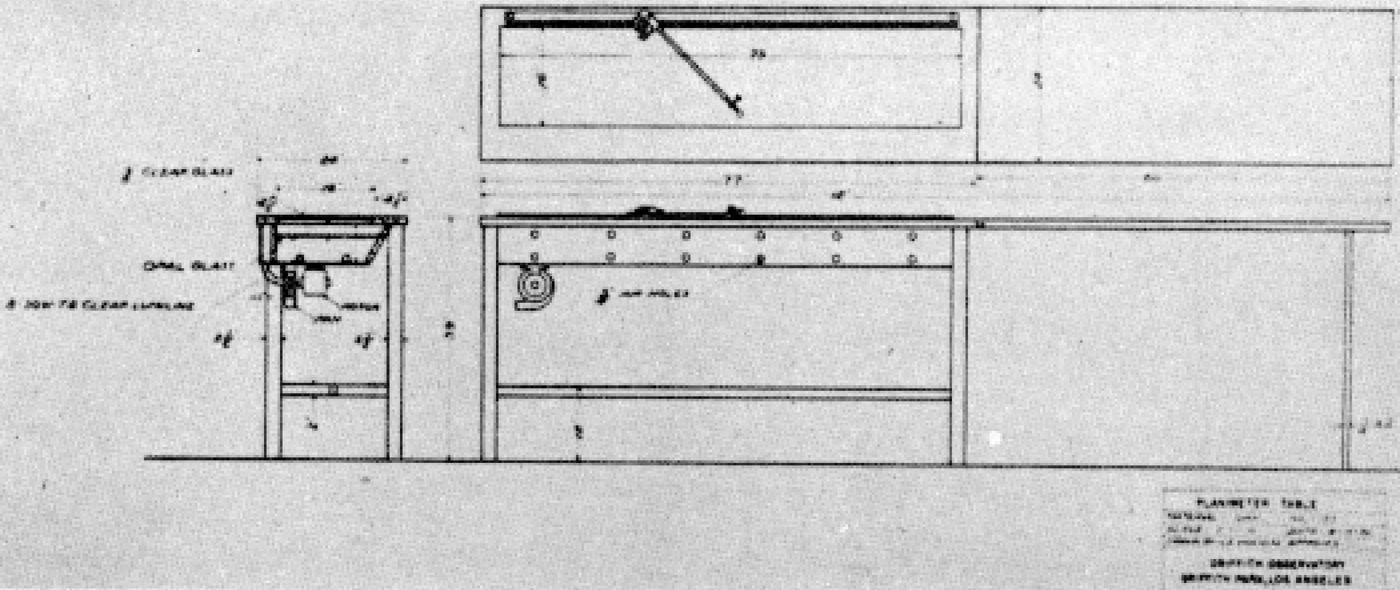
$$y_i = a \cos\left(\frac{2\pi i - \varphi_a}{p_a}\right) + b \cos\left(\frac{2\pi i - \varphi_b}{p_b}\right) + \cdots + c_i$$

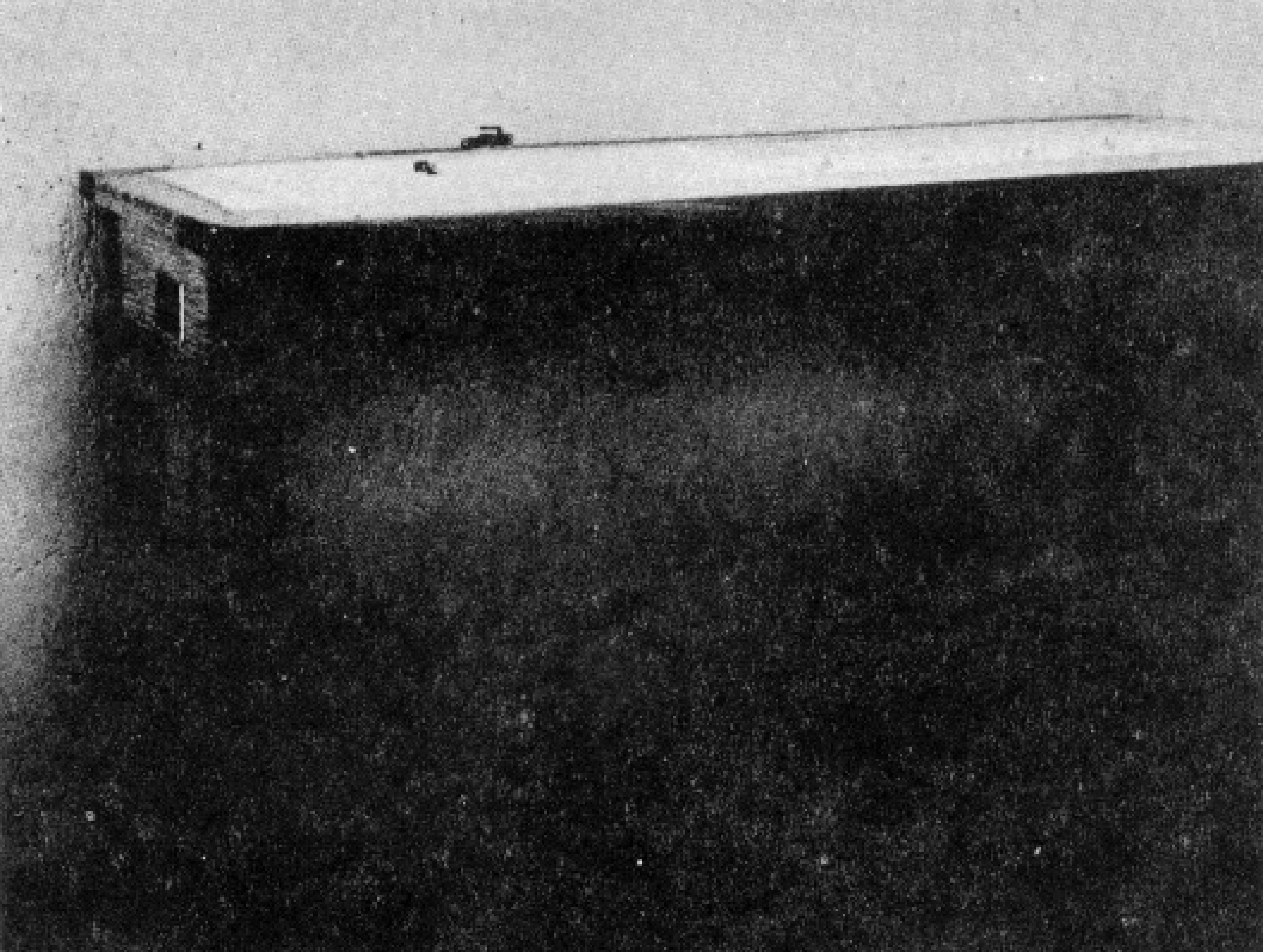
$$y_i - y_{i-l} = 2a \sin\frac{\pi l}{p_a}\left[\sin\frac{2\pi(\frac{1}{2}l - i) + \varphi_a}{p_a}\right] + \cdots + (c_i - c_{i-l})$$

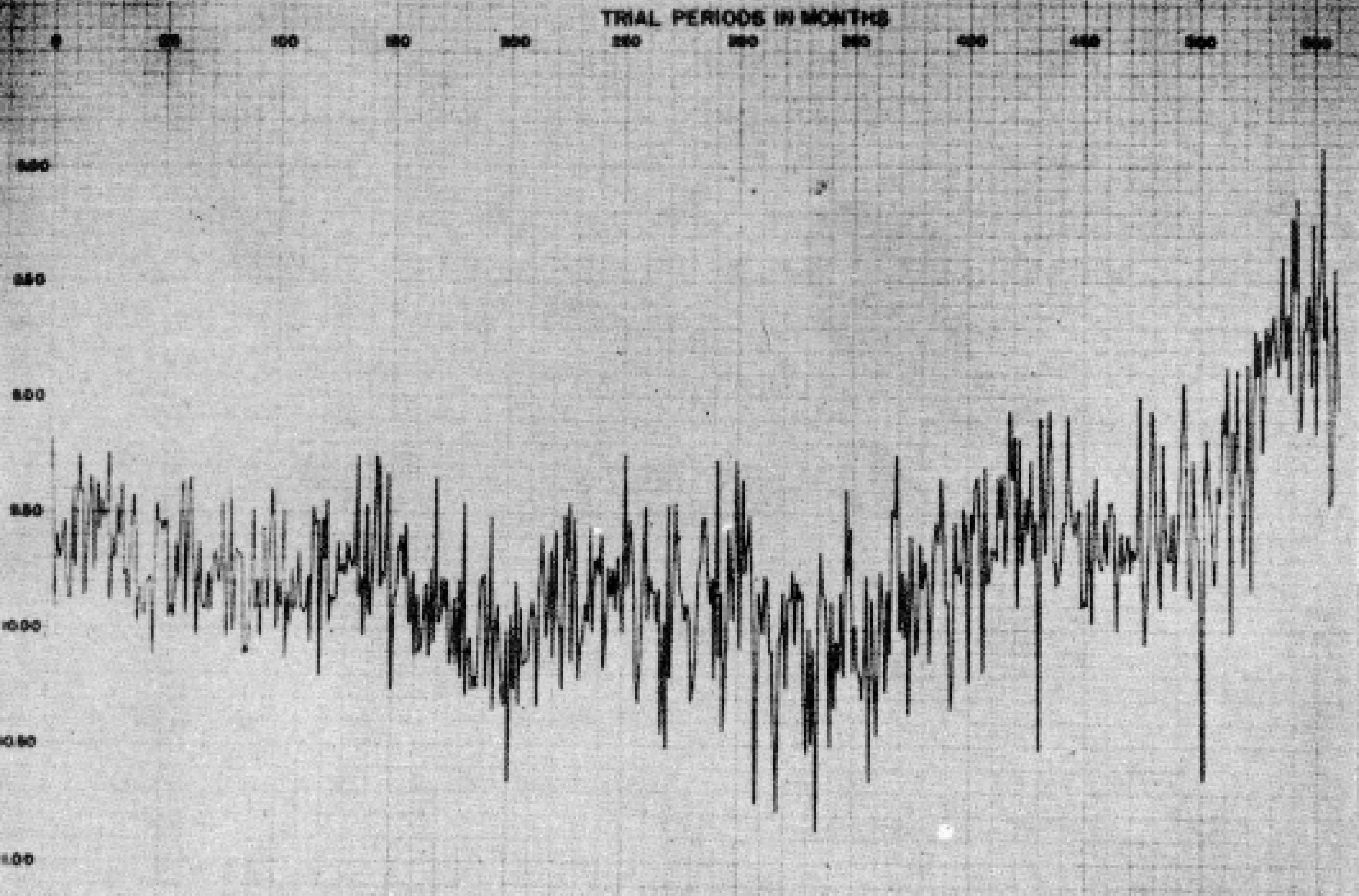
$$\sum (y_i - y_{i-l})^2 = 2(n - l)a^2 \sin^2\frac{\pi l}{p_a} + 2(n - l)b^2 \sin^2\frac{\pi l}{p_b} + \cdots + (n - l)\sqrt{2}\,\sigma_c^2$$

The sine terms contribute to A_l^2 in proportion to the squares of their amplitudes. On account of the $\sin^2\frac{\pi l}{p_j}$ factor, they contribute very little to values of A_l for which $\frac{\pi l}{p_j}$ is not very closely an even multiple of π.

This method has been applied to rainfall data of the Pacific Coast and has proved as satisfactory in practice as would be expected from the simplicity of the theory. The periodogram of rainfall stations along the northern third of the California coast is shown here, exhibiting perhaps the most definite single piece of evidence ever found for rainfall cycles. Outstanding is a cycle of about 45 years with its fourth harmonic as the secondary feature. The writer expects to publish the results of that work in the Monthly Weather Review.


CLEAR GLASS
OPAL GLASS
AIR HOLES
PLANIMETER TABLE
GRIFFITH OBSERVATORY
GRIFFITH PARK, LOS ANGELES





NORTHERN CALIFORNIA RAINFALL PERIODOGRAM