POWER-SPECTRUM ANALYSIS OVER LARGE RANGES OF FREQUENCY

By Capt. H. L. Griffith, U.S.A.F., H.A. Panofsky, and I. Van der Hoven

Pennsylvania State University[1,2]

(Manuscript received 21 July 1955)

ABSTRACT

It is shown that a spectrum covering a large range of frequencies can be found by superposing estimates of spectra of means over periods of different lengths. The method is illustrated by the power spectrum of temperature at University Park, Pennsylvania, covering periods from 2 to 7300 days. The spectrum is characterized by a major peak at four days and several minor ones, the reality of which is uncertain.

1. Introduction

The search for cycles in meteorological records is slowly giving way to "power spectrum" analysis. The reason is that, aside from the diurnal and annual variations, there is little physical reality to regular cycles. In a given record, cycles of certain periods may stand out accidentally — yet forecasts based on such cycles are always unsuccessful.

In spectrum analysis, a record is not considered to be made up of a finite number of oscillations with discrete frequencies, but of a large number of small oscillations with a continuous frequency distribution. The "power spectrum" itself measures the distribution of the variance of the variable over the various frequencies or periods.

The computation of power spectra, based on a theorem by Wiener [1], in this country often follows a procedure recommended by Tukey [2]. Given N observations, autocorrelation coefficients or autocovariances are computed with lags 0 to m. Next, these $m+1$ coefficients are subjected to a cosine transform, which is finally smoothed by a three-term moving average with weights $\frac{1}{4}$, $\frac{1}{2}$ and $\frac{1}{4}$. The resulting terms form $m+1$ estimates of the power spectrum. This procedure has been varied in different ways; however, a description of the variations goes beyond the scope of this article.

The shortest period about which a spectrum estimate is available is twice the interval between observations. The number of lags, m, determines the resolution of the spectrum. The smaller m, the poorer the resolution. The degree of resolution is particularly important at low frequencies. Consider, for example, a set of 1000 observations of daily mean temperatures. Further, let us assume that we have chosen $m = 6$. Then, the first spectrum estimate will apply to a frequency of zero cycles per day, the second to 1/12 cy/day, the third to 2/12 cy/day, and so forth up to 6/12 cy/day. The second estimate does not measure the power spectrum right at 1/12 cy/day, but is effectively an average estimate from 0 to 1/6 cy/day. In other words, all the variations of temperature with periods larger than 6 days have been lumped into a single estimate. Frequently, however, we are interested in estimating the probabilities of oscillations with periods of a week, two weeks, or even two months. We can achieve a range of periods from 2 days to 2 months only by choosing $m = 60$, or larger. This immediately increases the amount of computation tremendously. Further, if we choose $m = 60$, we actually obtain more resolution at high frequencies than would actually be needed. For example, the last spectrum estimate applies to 0.500 cy/day, and the next to the last estimate to 0.492 cy/day, or to periods of 2.00 and 2.03 days. Therefore, it is desirable to find a scheme that gives more low-frequency resolution than $m = 6$, and less high-frequency resolution than $m = 60$, and which at the same time entails less computation than the straight-forward analysis with $m = 60$. These aims can be accomplished by estimating different portions of the spectrum separately and piecing them together afterwards. In particular, in the example above, we might first estimate the spectral distribution over the higher frequencies by choosing $m = 10$. This gives good resolution for periods between 2 and 10 days. Next, we might form 5-day means (non-overlapping) and analyze their spectrum with $m = 10$. This would yield spectrum estimates with good resolution between periods of 10 and 50 days. If we should like more information about lower frequencies, we might compute spectra for means over even longer periods.

When spectra from observations over different periods are pieced together, the problems of the effects of "averaging," and of "aliasing," on the spectrum are encountered.

[1] Capt. Griffith participated while assigned to Pennsylvania State University by the U. S. Air Force Institute of Technology as a student in advanced meteorology. He is now on duty with the Air Weather Service.

[2] Mr. Van der Hoven's present affiliation: U. S. Weather Bureau.

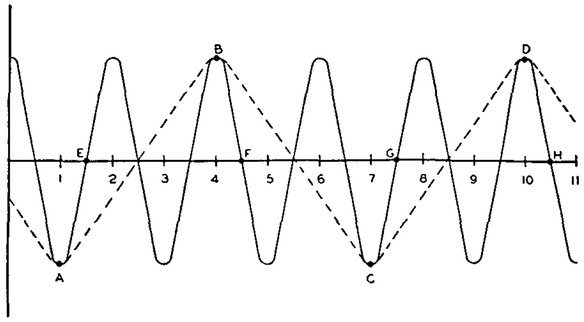

FIG. 1. Production of "aliasing" (schematic).

2. Effect of averaging

If observations are averaged, the resultant time series is devoid of high frequencies. In other words, averaging acts like a filter which eliminates the high frequencies from the record. In optical terms, averaging acts like a red filter. The nature of the "filtering function" is well known [3]. If $S_x(n)$ is the spectrum of x as function of the frequency n, the spectrum of the arithmetic mean $\bar{x}$ averaged over period T_0 is given by

$$S_{\bar{x}}(n) = S_x(n) \left(\frac{\sin \pi n T_0}{\pi n T_0} \right)^2. \qquad (1)$$

It is clear that the spectrum of x can be derived from the spectrum of $\bar{x}$ with the aid of (1).

3. Aliasing

Whereas the averaging process reduces the spectral intensity at a given frequency, "aliasing" removes energy from a high frequency and makes it appear at a lower frequency. This effect is due to the finite spacing between observations, which makes it impossible to observe oscillations shorter than twice the period between observations. In cases where important high-frequency variations occur, these will still appear in the spectrum, but at a lower frequency. For example, consider fig. 1. Here we have represented a fictitious time series with a period of 2 days. Now, suppose we take a reading only every 3 days, *e.g.*, at points A, B, C and D. The line connecting these points would have a period of 6 days! Obviously, the original record contained no cycle with this period. Rather, it was introduced artificially by the large spacing between observations. If we had made observations at E, F, G and H, (again at 3-day intervals), no fictitious oscillation would have resulted. This means the importance of aliasing is greatly influenced by sampling fluctuations. In general, the effect of aliasing is most pronounced at the high-frequency end of any spectrum.

4. Combined effect of aliasing and averaging

Fig. 2 summarizes a number of computations made for the spectrum of vertical velocity at Brookhaven, N. Y., 1 September 1952, for a period a little over an hour in length. The vertical velocities were based on Esterline-Angus traces made by a bi-directional vane and a Friez *Aerovane*. The instruments have been described previously [4]. For this period, (10/3)-sec averages were estimated from the trace by eye. The crosses and triangles in fig. 2 show the spectrum as determined from these observations by standard procedures. The circles indicate the spectrum obtained from (100/3)-sec means found from the original observations after application of (1). The fit is good only for the first six and the last point. Three of the spectrum estimates are high, presumably due to "aliasing."

This and other examples show that good estimates of portions of spectra can be made from averages by

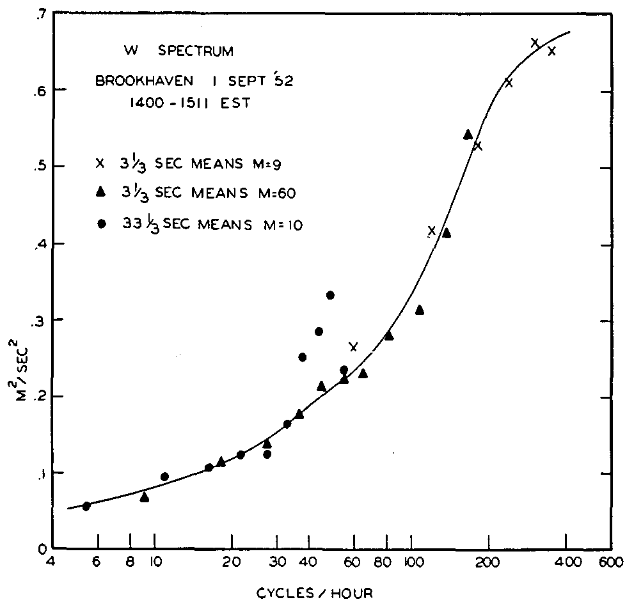

FIG. 2. Comparison of spectra of averages of different duration.

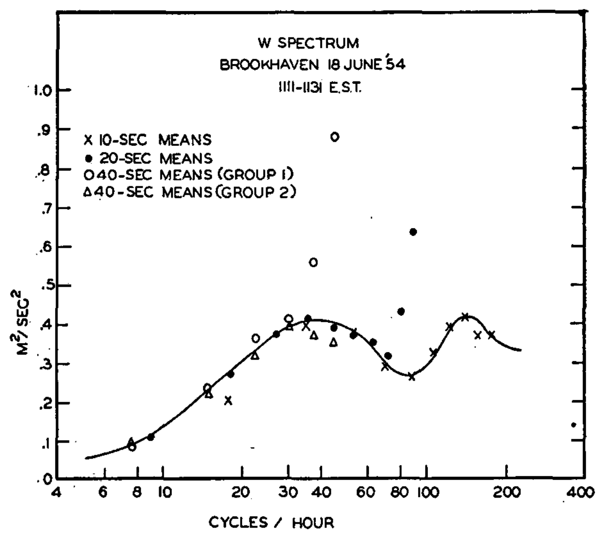

FIG. 3. Comparison of spectra of averages of different duration.

applying (1) and dropping the last 20 to 40 per cent of the estimates. Although it might not seem worthwhile to make the computation and keep only about 60 per cent of the results, it should be noticed that the first estimates cover a much larger range of periods than the last, namely, from 10 min to about 2 min as opposed to about 1.5 to 1 min in the case of fig. 2.

Of course, elimination of the error due to aliasing would be preferable to omissions of the spectrum estimates themselves. As has been noted, the effect of aliasing tends to be extremely variable and greatly influenced by sampling variations. This can be seen, also, from fig. 3, which shows a spectrum of 10-sec average vertical velocities at 91-m height at Brookhaven. Superimposed are two spectra of 40-sec average vertical velocities, both corrected for the filtering properties of the average. The two sets of 40-sec average vertical velocities differed in the following way: in the first case, the first four values were averaged, then the 5th to 8th, the 9th to the 12th, *etc.*; the second group started with an average over the 3rd to 6th observations, followed by an average over the 7th to 10th, *etc.* Examination of fig. 3 indicates that aliasing led to gross over-estimates of the spectrum at 45 cy/hr for the first set of means, but no such over-estimate was noticed for the second set. Incidentally, the spectrum was computed also for 20-sec averages. The result, also shown in fig. 3, again shows a strong effect of aliasing. In summary, at present the best method to prevent erroneous spectral estimates due to aliasing seems to consist of omitting the spectral estimates at the highest frequencies obtained from each set of averaged observations. The actual number of estimates to be omitted seems to be quite variable. It appears easiest to choose the means in such a way that their spectra overlap sufficiently to make possible an estimate of the effect of aliasing by inspection. For example, the observations summarized in fig. 3 would suggest that two spectral estimates made from 20- and 40-sec means should be omitted.

5. Spectrum of University Park daily temperatures

Fig. 4 shows the spectrum of daily mean temperatures at University Park, Pa., over a range of periods from 7300 to 2 days. All computations were made on electric desk calculators. Most of the original observations were daily mean temperatures (minimum and maximum averaged), covering 20 yr from 1934 to 1953, inclusive. From these observations, the normal annual temperature variation was removed by subtracting a sinusoidal curve with period of 12 months, fitted by least squares to the mean monthly temperatures (obtained from a 66-yr record). In addition, annual mean temperatures from 1887 to 1954 were analyzed.

From the daily temperatures, averages were formed over 1, 2, 3, 6, 14 and 42 days. Only the long-period, low-frequency part of the spectrum was based on all 68 years of record. The high-frequency end was based only on one or two years' data, to eliminate

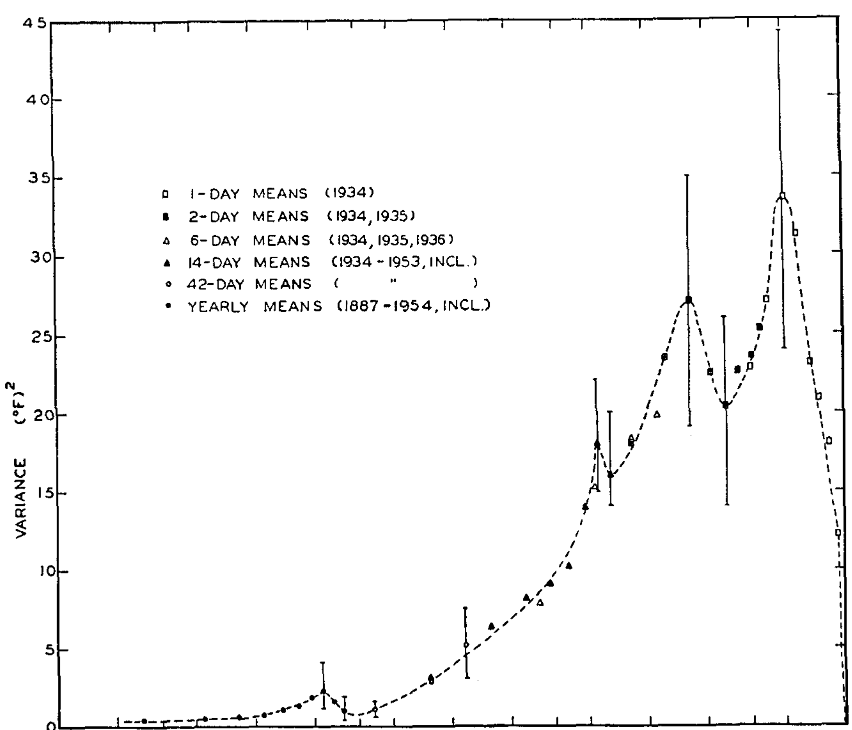

FIG. 4. Spectrum of mean temperature at University Park, Pa., giving, also, 5 and 95 per cent fiducial limits.

TABLE 1. Portions of temperature spectrum.

Years	No. of days averaged	Periods analyzed (days)	m	Deg. of freedom
1887–1954	365 or 366	730–7300	10	13
1934–53	42	84–504	6	57
1934–53	14	28–252	9	115
1934–36	6	12–72	6	60
1934–36	3	6–54	9	81
1934	2	4–48	12	30
1945	2	4–48	12	30
1934	1	2–24	12	60

excessive computation. It is assumed that the spectrum estimates resulting from this short period are representative of the high-frequency estimates for the whole 68-yr period. Table 1 summarizes the parameters for the various portions of the spectrum shown in fig. 4. For these, the filtering properties of averages have been removed and the last point in each portion omitted.

Fig. 4 shows the temperature spectrum on a logarithmic scale of frequency or period. In order that area on the graph represent the variance produced by oscillations with periods in the ranges indicated along the abscissa, the ordinate is the original spectrum estimate multiplied by frequency. Since the units of the ordinate are then variance, and independent of the frequency scale, this representation has the convenient property that the different portions of the spectrum can be combined without the use of conversion factors.

Fig. 4 shows individual points as well as an eye estimate of the smoothed power spectrum of temperature. A surprisingly large fraction of the variance is produced by short-period fluctuations.

The most important property of the temperature spectrum is the large peak at four days. Inspection of the original records showed, indeed, that fluctuations with this relatively high frequency are quite common, perhaps reflecting the average time between frontal passages.

The spectrum also shows that a 5-day mean chart would successfully cut off fluctuations of cyclonic scale and preserve the variation with larger time scale, for which the peak at 12 days appears to be typical.

The spectrum shows several secondary peaks and valleys. Their reality can best be judged by inspection of the 95 and 5 percent fiducial limits of the spectrum (shown in fig. 4) as well as the spectrum itself. Tukey [2] proved that spectrum estimates are distributed according to χ^2/f, where f is the degrees of freedom, defined by $(2N - \frac{1}{2}m)/m$. The number of degrees of freedom for each portion of spectrum is given in table 1. Fig. 4 indicates that the various secondary peaks and valleys may be due to sampling variations, and that the spectra obtained from a longer record would not show these features. Of course, this test cannot disprove the reality of the peaks.

REFERENCES

1. Wiener, N., 1930: Generalized harmonic analysis. *Acta Math.*, **55**, 117–258.
2. Tukey, J. W., 1949: In *Symposium on "Applications of autocorrelation analysis to physical problems."* Washington, Office of Naval Research, 47–67.
3. Jeffreys, H., and B. Jeffreys, 1950: *Methods in mathematical physics.* London, Cambridge Univ. Press, 708 pp.
4. Mazzarella, D. A., 1952: An all-weather remote-recording bivane. *Bull. Amer. meteor. Soc.*, **33**, 60–66.