Rainfall Singularities

E. J. HANNAN

Australian National University, Canberra City, Australia

8 January 1962

Bowen (1953) has suggested that the seasonal pattern of world rainfall is marked by certain singularities. By this is meant that on certain days the mean rainfall is very much greater than on neighboring days. He has suggested that these days of extraordinarily high mean rainfall are caused by the passage of the earth through meteor showers the dust from which, when it falls to appropriate levels, provides rain forming nuclei. Some considerable support for this theory was provided by Brier (1961) who examined three daily rainfall records for non-overlapping time periods and showed that there was some association between the three, by means of a test which was designed to be mainly influenced by forms of association due to the coincidence of the dates of peaks in the three series. Thus Brier's test appears to confirm the existence of common rainfall singularities in the series and since one of the series was composed of records from a large number of stations with a world wide distribution this in turn suggests an extraterrestrial cause for the phenomenon and thus tends to support Bowen's theory.

The material for Brier's test was constituted by three series of 365 observations:[1]

series 1—United States Rainfall 1952–57,
2—United States Rainfall 1958, and
3—World Rainfall 1880–1950.

These will be referred to as the "daily means" series below. The third series was constructed by Bowen while the construction of the other two is described in Brier's paper (1961). Brier replaced each of these three series by a new series obtained by scoring 1 for each of the 20 greatest rainfall peaks and also for the days immediately following and preceding the peak, -1 for each of the 20 deepest troughs and also -1 for the days immediately preceding and following, and finally scoring 0 for every other day. These three series constructed by Brier will be referred to as Brier's series below. Brier then considered the $41^2=1681$ arrangements for the three series relative to each other, obtained by fixing series 1, permuting series 3 cyclically by up to 20 days in either direction and lagging series 2 by up to 20 days in either direction (December 1957 and January 1959 were adjoined to series 2 to enable this to be done). For each such arrangement of the three series he computed a score by counting the number of occasions on which

(a) three units occur on the same day,
(b) three of -1 occur on the same day,
(c) three zeros occur on the same day,
(d) two of 1 and one of 0 occur on the same day,
(e) two of -1 and one of 0 occur on the same day.

The score for the initial arrangement (zero lag) of the three series was 178 and only 22 of the 1681 scores were as great or greater than this. Thus, using this score as a measure of association, significance is obtained at the $(22/1681)\times 100=1.31$ per cent level.

The highest score (187) was found when series 3 was not permuted and series 2 led the other series by 1 day. If the other arrangements equally as near or nearer to the zero lag arrangement are considered as equivalent, *a priori*, to the arrangement at which this greatest score was got, then significance is obtained at the $(7/1681)\times 100=0.4$ per cent level.

Thus there is considerable evidence for some form of association between the three series. The nature of the test and of the results obtained (see Brier, 1961, Table 2 and Fig. 3) suggests that this association is due to a sharply phased agreement between the peaks and troughs in the three series.

There seems to be only one possible explanation for the results apart from the rainfall singularities explanation and this is the possibility that there is some periodic oscillation in the world weather or in the weather of stations which influence markedly the three series (i.e., U. S. rainfall). Brier in a personal communication suggested that this could not be so, mainly because the peak in the test score which he had computed was marked and the score fell away quickly as the average displacement of the series 1 and 3 relative to 2 increased, which would hardly be the case with a conventional seasonal oscillation (of not very short period). To check this the present author considered Brier's series 1 and 3 alone. These two were considered because the averaging involved in their construction would reduce the level of "noise" superimposed on the effects being sought. Attention was restricted to two series only, to reduce the labor, which was done by hand. For the same reason the score computed was obtained more simply by counting the number of pairs (1,1) and $(-1,-1)$ for the two series, one score being so computed for every cyclic permutation (so that 365 scores were obtained). The resulting series of 365 scores shows a marked oscillation with a frequency of about 18 cycles per year (cpy). This series is shown in the form of sums of successive sets of 5 in Table 1 below.

[1] The series used are actually three day running means of the actual daily values.

TABLE 1. Number of occurrences (1,1) and (−1,−1) in Brier's series 1 and 3 for all possible cyclic permutations of series 3 relative to series 1. Sums of successive sets of 5 scores.*

Mean lead Days	Score	Days	Score	Days	Score
3	119	123	100	243	121
8	71	128	77	248	103
13	93	133	66	253	75
18	109	138	128	258	94
23	153	143	124	263	122
28	85	148	109	268	98
33	79	153	70	273	69
38	91	158	77	278	97
43	106	163	81	283	104
48	87	168	89	288	79
53	57	173	77	293	58
58	96	178	98	298	70
63	118	183	164	303	107
68	67	188	88	308	85
73	79	193	80	313	94
78	102	198	98	318	65
83	129	203	148	323	143
88	72	208	93	328	105
93	44	213	77	333	100
98	104	218	85	338	70
103	109	223	138	343	103
108	70	228	93	348	109
113	45	233	88	353	91
118	99	238	60	358	72
				363	110

* There are errors in these computations as is evident from the fact that 58 of the 365 values for Brier's series 1 are −1, and 59 are +1, while the corresponding totals for series 3 are 58 and 60. Thus the total of the scores for all lags should be 6904 whereas it is 6836 in the table above. The 80 values for series 3 lagged with respect to series 1 from −15 to +65 days were checked and a total discrepancy of 33 was found. The errors are evenly distributed over the figures in Table 1 and will not affect the general nature of the conclusions.

The contribution from the harmonic with frequency 18 times a year is very large, the total corrected sum of squares for the 73 figures in Table 1 being 42,263, of which 57 per cent is explained by this one harmonic. The next major peaks on either side of the peak at zero lag are in the neighborhood of 22 lags and −15 lags for series 3 relative to series 1, the peak at −15 being smaller than that at zero lag, while that at +22 is larger. The peak at +22 lag will not have shown up in Brier's test of course.

Two questions now need to be discussed: (1) Is this oscillation due to the treatment of the data; for example, is it merely a reflection of the choice of the particular number of 20 peaks and troughs? (See Brier, 1962.) (2) If it is not due to the treatment of the data but rather to a tendency for the two series to oscillate with a frequency near 18 times a year, does this tendency correspond to the presence of a true periodicity (i.e., the presence of a finite spectral mass precisely at this frequency for the two series), or does it represent rather merely a peak in the two spectra?

(1) The present author can see no theoretical reasons why the treatment of the data should have induced the oscillation. The transition from the daily means series to what we have called Brier's series was effected by a markedly non-linear transformation. If the transformation were instantaneous (in the sense that the new value at time t was a function only of the initial value at a time t and not of any lagged values), then one would feel that the new series should, if anything, have a "whiter" spectrum (i.e., one nearer to uniform). (See Grenander, 1959.) In fact the transformation is not instantaneous since the decision as to whether the score at time t shall be 1, −1 or 0 depends on the size of the peak at that point relative to other peaks. The impression remains, however, that the treatment of the data has not induced the oscillation. In any case the simplest thing to do is to look at the constituent series. The results are displayed in Table 2 below, which was obtained by using the 73 sums of consecutive sets of 5 days in all cases.

TABLE 2. Fraction of total corrected sum of squares due to harmonic with frequency 18 times a year.

	Series 1	Series 3
Daily means	0.117	0.178
Brier's series	0.131	0.173

In the case of series 1 a large part of the total sum of squares (27 per cent) is due to an annual term which was removed by Brier before calculating his series and the figure for series 1 for the daily means series has therefore been calculated after removing this seasonal component. The general impression is that the transformed data is not markedly different from the original data insofar as the presence of a component with frequency 18 cpy is concerned. The testing of the significance of the components at the frequency being studied needs further discussion, which will be given under (2) below.

The product of the two figures in Table 2 for the daily means series is 0.021. All of such products have not been computed but the few corresponding to figures shown in Table 3 below (omitting the 18 cpy component) average 0.0005. Omitting the annual term, there would be 34 of these, giving a total of 0.017 if this average is maintained over other frequencies. The ratio (0.021)/(0.021+0.017)=0.55 represents the proportion of the total variance of the 365 figures obtained by forming the lagged series of cross products (permuting cyclically) of the two daily means series. It was mentioned above that 57 per cent of the variation in the (related) series given in Table 3 was due to the frequency 18 cpy. A more detailed analysis would check all of this but it seems probable that this cycle can be completely explained in terms of the peak in the spectrum of the two daily means series at 18 cpy.

(2) At the time of writing only the actual daily means for the two series are available to this author. Obviously

more could be done with the actual daily values from which these series were constructed. The computing was done by hand so that in all cases the 73 observations were used. The classic test of significance for the presence of an harmonic component, when the frequency is prescribed *a priori*, is obtained by forming the F ratio obtained in the usual way for a regression analysis (see Fisher, 1944). For the series exhibited in Table 1 we have, of course, chosen the harmonic with the largest amplitude and this test is not appropriate. An appropriate test, *on the null hypothesis of randomness for the series*, has also been given by Fisher (1929). This test is obtained by computing the ratio exhibited in Table 2. For 73 observations this must be greater than 0.17 for significance at the 5 per cent level. However, this test is not appropriate here for a variety of reasons. In the case of the series exhibited in Table 1, for example, it is not reasonable to set up as a null hypothesis that the series is random. In this case, of course, a test is hardly needed. There can be little doubt of the reality of an effect which explains 57 per cent of the variation in a series. In the case of the series studied in Table 2 this objection arises also and, of course, in addition we have not now chosen the harmonic component with the largest amplitude by direct examination of the series so that the test is also too strict. For a discussion of the problem of testing for a true periodicity in a time series the reader is referred to Hannan (1961). Here it will suffice to say that a decision as to whether a periodictiy exists must be founded upon an examination of the spectrum of the series in the neighborhood of the frequency corresponding to the apparent periodicity. For the present series the question is complicated by the fact that only the daily means are available. Thus the two series used cannot be regarded as parts of time series of infinite extent. If we regard them as being obtained from stationary time series their spectra will be discrete, the spectral mass (i.e., part of the total variance) at the frequency j (cpy) being n^{-1} times an average of the spectral mass in an interval of width 1 cpy centered at the frequency j for the original series from which the daily averages were obtained. Here n is the number of years averaged. The more years which are averaged to obtain these daily averages, the more closely will the spectral average concentrate around the frequency j. If there is a discrete spectral mass in the original data at the frequency j this will be added to the mass at that frequency for the daily averages. The contributions to the total sum of squares from the harmonics with frequency near 18 cpy for the two daily means series (using averages of 5 successive days) are shown in Table 3.

In the case of series 1 the peak is quite sharp and suggests that there is a considerable concentration of spectral mass for the original series in a very narrow interval around the frequency 18 cpy. For series 3 the effect appears not so sharp, though again there is little doubt that there is a concentration of spectral mass near this frequency. There seems to be some evidence therefore for a periodicity in the data with frequency 18 cpy.

TABLE 3. Contributions to total sum of squares from certain harmonics, for daily means series. Sums of five successive days.

Frequency	Fraction of total sum of squares For series 1*	For series 3
16	0.018	0.032
17	0.003	0.085
18	0.117	0.178
19	0.014	0.033
20	0.017	0.037

* Excluding contribution from annual term.

When the daily means for 1958 (whose peaks and troughs give Brier's series 2) are examined, there is no significant component at 18 cpy. This might be expected since the effect of averaging 6 years of data to give series 1 would increase the chance of discerning an effect. When the 1958 data are combined with the 1952/57 data, the figures in the second column of Table 3 become 0.034, 0.001, 0.153, 0.014, 0.032. The average for the four frequencies other than 18 cpy would be $34(0.153/0.847)=6$ which as F with 2 and 68 degrees of freedom is significant at the 1 per cent point. The analysis made suggests that even if the data cannot reasonably be regarded as normal on the null hypothesis, the average intensity around 18 cpy is no greater than $0.847/34=0.025$ and may be less, so that the appropriate F ratio is no smaller than 6. As indicated earlier, this test is not valid as we have tested the significance of the frequency 18 cpy after examining the data. A test which would be satisfactory would be one based upon the largest product of any two figures in the same row of an extension of Table 3 to all frequencies. Though it would not be too difficult to derive such a test on the assumption of randomness (and normality) for the two series, the test would require considerable modification in the present case. The coincidence of the major peaks in the two series at 18 cpy, *each* being near to significant even when considered as the greatest in 34, leaves little doubt of the reality of the jump in the two spectra.

Independent evidence is provided by an examination of the phasing of the two series. If there is no true periodicity in the data but merely a concentration of spectral mass in a narrow frequency range, then one would not expect any close agreement between the phasing of the oscillations in the two series. That there is some agreement is shown by Table 4 below.

These are all in fairly close agreement. Each angle is (approximately) distributed uniformly over the interval from $-\pi$ to π if there is no true periodicity. The probability that so small a difference (modulo 2π)

TABLE 4. Phase angle for harmonic with frequency of 18 cpy.

	Daily means	Brier's series
Series 1	−2.737	−2.608
Series 3	−2.372	−2.733
Series 1+3	−2.819	—

as that observed from Brier's series would arise if the two series were independent is about 0.04, while for the series 1 and 3 of daily means it is 0.12.

Though this evidence by itself is by no means conclusive, when it is considered along with that obtained from the consideration of the intensities it becomes fairly considerable. The test conducted by Brier constitutes a different examination of the same evidence and since his tests are valid tests one must agree that there is association between the two series. This association is apparently due entirely to the presence of jumps in his two spectra at or near the frequency 18 cpy, the jumps being due to a common cause so that they are in phase. A thorough examination of the spectrum for the original data from which the series were obtained would be needed to finally establish whether Brier's effect is due to this cycle. This examination would best be confined to the more homogeneous data from which series 1 and 2 were obtained and which, if carried up to the end of 1961, would give 3650 observations.

The only other series examined for the presence of the cycle at frequency 18 cpy was Sydney rainfall for the period 1859 to 1952. There appeared to be no such periodicity in this data (though a really careful analysis was not made). Bowen's series 3 has a markedly non-homogeneous appearance, the higher frequency components (with frequency around 36 times a year) being much more important in the first three months of the year than in the last nine. Thus non-homogeneity may, at least partly, arise from the combination of rainfall records from regions with very different patterns of seasonal rainfall. The variance of daily rainfall is probably strongly related to the mean in some places so that the combination of results from regions with two different seasonal patterns may result in the virtual disappearance of one of the two region's oscillations under the much greater oscillations from the other, over part of the year. This in turn could explain the slight apparent spreading of the peak, around 18 cpy, for Bowen's series. This is not much more than a guess of course.

The possibility that the component at frequency 18 cpy is an aliasing effect should not be forgotten. The nearest frequencies aliased with 18 cpy have periods, in hours, of 25.24, 22.87, 12.30 and 11.71. Tides in the atmosphere with periods of about 12.425 hr are known but the corresponding frequency would alias with about 25 cpy.

If the association which shows up in Brier's test is due to a periodicity in the data, as seems to be the case, then Brier's result appears to give no support to the isolated rainfall singularities hypothesis and therefore no support to Bowen's theory.

REFERENCES

Bowen, E. G., 1953: The influence of meteoric dust on rainfall. *Austral. J. Phys.* **6**, 490–497.

Brier, G. W., 1961: A test of rainfall singularities. *J. Meteor.*, **18**, 242–246.

——, 1962: Visual meteoritic activity and rainfall singularities. *J. Atmos. Sci.*, **1**, 56–59.

Fisher, R. A., 1929: Tests of significance in harmonic analysis. *Proc. Roy. Soc. A*, **125**, 54–59.

——, 1944: *Statistical methods for research workers.* 9th ed. London, Oliver and Boyd, p. 131.

Grenander, V., 1959: Some non-linear problems in probability theory. *The Harold Cramer Volume on Probability and Statistics*, Uppsala, Almqvist and Wiksells, p. 111–112.

Hannan, E. J., 1961: Testing for a jump in the spectral function. *J. Roy. stat. Soc. B*, **23**, 394–404.