Reply

Glenn W. Brier

U. S. Weather Bureau, Washington, D. C.

2 June 1960 and 22 June 1960

I am indeed grateful for the comments of Shapiro and Macdonald as it gives me the opportunity to present some additional material and to discuss some pertinent aspects of the statistical test used. I regret that these were omitted from the *Journal* paper because of my failure to foresee all the questions that might be raised. Their suggestion that any association in the three series is due entirely to non-anomalies in precipitation is an interesting and rather crucial one, but is not supported by all the facts, as will be shown.

In addition to selecting the twenty highest and twenty lowest ranking peaks and troughs in each precipitation series another group of twenty minor peaks and twenty minor troughs was chosen. However, these were classified as non-anomaly or (0) days in the investigation reported since their amplitudes were relatively small and, at that time, it seemed reasonable that only the extreme peaks or troughs could have any conceivable physical reality. Fortunately, with practically no additional effort, these data are available in a summarized form which can permit the reader to make his own judgment regarding the nature and strength of any non-randomness indicated by the data. Now the most relevant comparisons from both physical and statistical considerations are those between series 1 and 3 and between series 2 and 3. These comparisons are made by the use of contingency tables shown in table 1. The symbols P, p, O, t, and T used for column headings refer to major peaks, minor peaks, other days, minor troughs and major troughs respectively. Series 3 is used as a reference base for comparison since it can be considered as the long-term "normal" or hypothetical population, resulting from over 50 years of record. For reasons pointed out in the *Journal* paper, series 2 was 14 months long, and this accounts for the greater numbers in table 1b. The numbers in parentheses on the main diagonal of each of these tables are the expected numbers, based upon the marginal totals.

Examination of these tables reveals a number of interesting points. Referring first to the main diagonals of the tables, the discrepancies of the observed numbers over the expected numbers are *not* limited to class (0). Next, referring to the first column of table 1a, there are 16 cases of a major peak in series 3 associated with a major peak of series 1, compared with 8 cases where a major peak in series 3 is associated with a major trough in series 1. Likewise, considering the same column, there are 9 cases of a major peak appearing with a minor peak compared with 8 cases with a minor trough. Continuing in this manner, the following comparisons and differences can be listed and summarized:

$16 - 8 = 8$	$17 - 12 = 5$
$9 - 8 = 1$	$5 - 4 = 1$
$11 - 5 = 6$	$14 - 7 = 7$
$7 - 7 = 0$	$11 - 9 = 2$
$8 - 9 = -1$	$12 - 12 = 0$
$13 - 16 = -3$	$9 - 7 = 2$
$12 - 11 = 1$	$20 - 5 = 15$
$9 - 5 = 4$	$12 - 13 = -1$

These 16 differences have an average value of 2.94, which departs significantly from zero even if one considers the degrees of freedom as few as 5 (rather 15). Actually, of course, there are 16 degrees of freedom available in each table or a total of 32 for both. Furthermore, it is noted that in both tables a non-

TABLE 1. Contingency table showing (a) the association between series 1 and series 3 and (b) the association between series 2 and series 3.

(a)

		Series 3 P	p	O	t	T	Total
Series 1	P	16 (9.8)	11	12	9	11	59
	p	9	7 (8.1)	18	16	5	55
	O	20	24	61 (51.4)	12	21	138
	t	8	7	24	13 (9.6)	9	61
	T	8	5	22	8	12 (8.7)	55
	Total	61	54	137	58	58	368

(b)

		Series 3 P	p	O	t	T	Total
Series 2	P	17 (11.1)	14	24	12	5	72
	p	5	11 (9.8)	29	7	13	65
	O	30	25	66 (64.5)	30	15	166
	t	4	9	31	9 (10.3)	12	65
	T	12	7	21	12	20 (10.6)	72
	Total	68	66	171	70	65	440

anomaly (0) day in series 3 is more frequently matched up with either a minor peak or minor trough than with a major peak or major trough. Clearly, there is a definite tendency for peaks to appear with peaks, troughs with troughs, *etc.* It was, of course, impossible to demonstrate these effects with only the material presented in the original paper.

Before making a similar comparison between series 1 and series 2 it is instructive to consider the following simple example to illustrate some of the statistical characteristics of the test that has been used. In the following table consider series S as a signal sequence while M_1 and M_2 are two separate messages containing the signal S but with some noise or random error added.

Series S	0	1	1	1	0	0
Series M_1	0	0	1	1	1	0
Series M_2	1	1	1	0	0	0

Inspection of these series shows that S and M_1 are associated in a positive sense, S and M_2 in a positive sense, but M_1 and M_2 have a negative correlation of the same magnitude. This results from the numerals (1) in M_1 being displaced one position to the right of their positions in S while the corresponding displacement in M_2 is one position to the left. Of course, it is well known in correlation theory that with 3 variables the simple correlations r_{13} and r_{23} can both be significantly positive while r_{12} can be zero or even *significantly* negative. Thus two garbled messages M_1 and M_2 can contain the essential parts of signal S without having any information content about each other. To use a biological analogy, two first cousins may show some resemblance to their common grandfather without showing any resemblance to each other. Actually it can be shown that the test for singularities used here is more powerful and robust (in technical statistical terminology) than standard correlation or contingency tests but this does not seem the appropriate place to discuss such statistical theory in depth.

The actual result of comparing series 1 and series 2 is summarized by the following differences which were determined in the same way as those previously presented:

$$-3, \quad 2, \quad -4, \quad -9, \quad 8, \quad -5, \quad -8, \quad -11.$$

These 8 differences have an average value of -3.75, giving a value of $t = 1.70$ which, although not significant, strongly suggests that series 1 and 2 are each more closely related to series 3 (in the positive sense) than they are to each other. Now the two highest scores reported in the study (186 and 187) result when series 1 and 2 are displaced approximately one day in *opposite* directions from series 3 and a detailed examination of the data disclose numerous cases where peak days and trough days are adjacent without any class (0) in between. The findings here are therefore consistent with the type of effect shown earlier in the illustrative example. It can be shown that the combination of these effects will be to reduce the magnitude of the score, thus making it more difficult to detect singularities if they exist.

A measure of the overall association between series 1 and 2 is shown by the column averages computed from the complete table of scores from which table 3 in the *Journal* paper is an excerpt. The row averages indicate the relation between series 1 and 3 and the averages of the diagonals (from lower left to upper right) indicate the relation between series 2 and 3. These averages are shown in table 2 as well as the

TABLE 2. Average scores (reduced by 100) according to several methods of summarization.

Lag	Rows	Columns	Main diagonals	Opposite diagonals
−20	53	54	52	56
−19	56	51	51	56
−18	57	51	53	57
−17	50	52	54	57
−16	47	52	58	57
−15	46	49	60	57
−14	48	48	59	58
−13	50	52	56	59
−12	50	59	53	59
−11	51	59	55	57
−10	49	57	58	57
− 9	52	53	59	56
− 8	58	54	61	56
− 7	68	54	63	55
− 6	68	53	64	56
− 5	64	53	58	55
− 4	58	53	51	56
− 3	61	59	48	56
− 2	61	62	50	57
− 1	62	65	58	56
0	61	61	64	56
1	65	62	65	55
2	62	61	57	55
3	59	61	48	54
4	53	58	44	54
5	53	64	47	53
6	53	68	48	55
7	54	68	51	55
8	54	58	56	56
9	53	52	62	55
10	57	49	68	56
11	59	51	64	55
12	59	50	62	56
13	52	50	57	56
14	48	48	56	57
15	49	46	51	56
16	52	47	50	57
17	52	50	50	57
18	51	57	54	57
19	51	56	54	57
20	54	53	60	56

averages for the opposite diagonals for purposes of comparison. In all cases the average score between ±1 lag appreciably exceeds expectation compared with the "nonsense" values listed in the fourth column. There is no reason for this to happen with 3 series that are random with respect to one another. As Shapiro and Macdonald point out, the contribution of class c is important to these scores but in no way does this vitiate the general test of significance nor does it deny the significant positive relationships indicated by table 1.

Innumerable other things could be done with the data but excessive attention to detail *a posteriori* can become dangerous since one is likely to find interesting peculiarities or apparent contradictions even in random data if one searches long enough. An experiment could be designed to test the bias of a coin by tossing it say 1000 times, and after having found a significant tendency for Heads, one could look for a subset of say 50 throws in which Tails dominated and then proceed to "prove" the coin was unbiased or even biased in the other direction. Shapiro and Macdonald are of course very much aware of such dangers but it seems worth emphasizing here for the general reader because of the tendency of some investigators or critics to use a few isolated cases to prove (or disprove) a point. We must not get so close to the trees that we can't see the forest. Small samples or even case histories are often useful in providing clues or suggesting ideas but they must be used cautiously for drawing inferences or conclusions.

Shapiro and Macdonald object to the statement " . . . a *strong* tendency for precipitation anomalies . . . to occur on specific calendar dates." The resolution of this controversy would seem to depend upon the definition of the word *strong*. Perhaps a more appropriate phrase would be " . . . a *remarkable* tendency . . . " *i.e.*, remarkable that any physical cause should be sufficiently pronounced to produce a signal strong enough to be recognizable through all the "noise" of a few years of precipitation data. This does not imply that the variations in precipitation are necessarily dominated by this cause, but no *weak* tendency would produce such results. Some empirical sampling experiments in which artificial "singularities" are introduced into two oscillatory time series that are random with respect to each other show that the magnitude of a common signal introduced in both series must exceed a standard deviation before it can be detected in series comparable to those used here and with this particular test. In view of the variability in the precipitation series discussed here, this suggests that real physical variations of say 2 or 3 per cent are likely to go undetected and it becomes necessary to consider possible physical mechanisms that can produce variations in precipitation of 10 or 20 per cent or more if they are to be detected at the significance level reported here.

I appreciate the interest of Shapiro and Macdonald in this subject and their efforts which have prompted me to make this attempt at clarification and hope that it has resulted in a better understanding of both the data and the characteristics of the significance test. However, in view of the foregoing, there appears to be no "strong" reason for making any important modifications in the conclusions previously stated.