causes: (*a*) If there are N numbers, there should be found $f(x)N$ of which the absolute deviation is equal to or greater than x; theory gives the value of $f(x)$; (*b*) the value of the expression $\frac{2N\Sigma\delta^2}{(\Sigma\delta)^2}$ should be 3.14159 . . . Now meteorological data may satisfy both these tests without at all fulfilling other conditions equally demanded by theory; we have here a good illustration of the oft-repeated warning against drawing conclusions from summary coefficients alone, such as the mean. In the present instance, the *order* in which the numbers appear is of great significance, and the following relation must also hold:[4]

If the deviations from the mean are to be likened to fortuitous errors, then the ratio of the mean variability to the mean deviation must be equal[5] *to* $\sqrt{2}=1.414$ . . . The variabilities and deviations are taken without regard to sign.

Drawings from a sack containing balls, on each of which was marked an observed daily temperature, would give a succession vastly different from the succession actually observed: Long series of increasing or decreasing values would be less frequent in the drawing than in the observing, and the mean variability would be greater in the former; in fact the ratio of mean variability to mean deviation in the case of series of daily temperatures turns out to be but little more than half the theoretical value; chance would give the deviations which are observed, but would not give the succession which is observed. *Yet both the actual and the chance successions satisfy the two tests mentioned above.*

It has been pointed out by Besson (*op. cit.*) that if a variable is taking on random values, it does not follow that the succession of the signs of the variations will obey the laws of chance; Goutereau points out further that the deviations from the mean may not be fortuitous even if they follow the Law of Gauss.—*Edgar W. Woolard.*

[4] Ch. Goutereau: Sur la variabilité de la température, *Annuaire de la Soc. Mét. de France*, 54, 122–127, 1906.
[5] The demonstration, by Maillet, is given by Goutereau, *op. cit.* The absolute difference between a number and the next consecutive number is the variability.

551.54 : 551.501

THE VARIATE-DIFFERENCE CORRELATION METHOD.

For correlating daily changes of barometric height at Halifax and Wilmington, Miss Cave[1] made use of a formula, devised by Pearson, giving the correlation coefficient between the differences of successive daily readings at the two stations; and remarked that this formula would apply to any case in which it was desired to correlate the difference of one pair of quantities with the difference of another pair; no comments on where this procedure might be desirable were offered, however. Later, Hooker[2] independently pointed out that the correlation coefficient between two variables, for each of which a series of observations is available, is a test of similarity of the two phenomena as influenced by the totality of the causes affecting each of them; when, therefore, the observations extend over a considerable period of time, certain difficulties arise which find no precise parallel in the case where the whole of the observations refer to the same moment of time: If a diagram be drawn, showing by curves the changes of the two variables during the period under consideration, some relation will often be suggested between the usually smaller and more rapid alterations while at the same time the slower "secular" changes may or may not exhibit any similarity. If, then, the correlation coefficient be formed in the ordinary way, employing deviations from the mean, a high value will be obtained if the "secular" changes are similar (this value being almost independent of the similarity or dissimilarity of the more rapid changes), but a value approximating to zero if the "secular" changes are of quite dissimilar character even though the similarity of the smaller rapid changes be extremely marked; deductions drawn from ordinary correlation coefficients may be very erroneous. In order to get rid of the spurious correlation arising from the fact that both variables are functions of the time, the correlation coefficient may be formed between the *variations*, or first differences, of the quantities, instead of between the quantities themselves. After this method had been in rather extensive use for some time, Pearson pointed out that it was valid only when the connection between the variables and the time was linear.

The name Variate-Difference Correlation was given by Pearson[3] to a generalization of the preceding artifice, in which it was demonstrated[4] that if the variables are randomly distributed in time and space, the correlation between the variables and that between the corresponding *nth* differences will be the same; and that when this is not the case, we can eliminate variability which is due to position in time or space, and so determine whether there really is any correlation between the variables themselves, by correlating the 1st, 2d, 3d, * * *, *nth* differences: *when the correlations between the differences remain steady for several successive orders of differences we may reasonably suppose we have reached the true correlation between the variables.*

The complete theory of the method was worked out by Anderson[5] and subjected to critical examination by Pearson (*op. cit.*), who found that, as usual, the theoretical formulæ were only roughly approximated to in practice unless a great number of observations were at hand.

There has been no source more fruitful of fallacious statistical argument than the common influence of the time factor. The difference method of correlation is one of great promise and usefulness. The very frequent and superficial statements that such and such variables, both changing rapidly with the time, are essentially causative cease to have any foundation when the difference method is applied.[6]—*Edgar W. Woolard.*

[1] F. E. Cave-Browne-Cave; On the influence of the time factor on the correlation between the barometric heights at stations more than 1,000 miles apart, *Proc. Roy. Soc.*, 74:403–413, 1904–1905.
[2] R. H. Hooker: On the correlations of successive observations, *Jour. Roy. Statistical Society*, 68:696–703, 1905.
[3] Beatrice M. Cave and Karl Pearson: Numerical illustrations of the variate difference correlation method, *Biometrika*, 10, 340–355, 1914–15.
[4] "Student": The elimination of spurious correlation due to position in time or space, *Biometrika*, 10, 179–180, 1914–15.
[5] Nochmals über "The elimination of spurious correlation due to position in time or space," O. Anderson, *Biometrika*, 10, 269–279, 1914–15.
[6] Illustrations of the method are given by Cave and Pearson, *op. cit.*, and by G. U. Yule, *Introduction to the Theory of Statistics*, 5 ed., 1919, pp. 197–201; see also T. Okada, Some researches in the far eastern seasonal correlations, Mo. WEATHER REV., 1917, 45: 238, 299, 535.

NOTE ON PROF. MARVIN'S DISCUSSION OF "A POSSIBLE RAINFALL PERIOD EQUAL TO ONE-NINTH THE SUN-SPOT PERIOD."

By DINSMORE ALTER.

[University of Kansas, Lawrence, Kans., Apr. 26, 1921.]

I have naturally been much interested in Prof. Marvin's conclusions[1] regarding my paper.[2] I am very sorry that it is impossible for us to agree concerning the possibility of the phenomenon discussed, and especially concerning the legitimacy of the method employed. A further statement concerning some of the points raised by him may be in order.

[1] Mo. WEATHER REV., February, 1921, 49: 83–85.
[2] *Ibid.*, pp. 74–83.

In no place in the paper is there any reference to a systematic variation in the length of the sun-spot period as claimed in the opening paragraph and also later in Prof. Marvin's discussion. The figure which gives the length of the period for each year is not in any way based on such a supposition and applies equally, whether, as believed by Newcomb, the differences are accidental variations or, as by Lockyer and Clough, they are systematic. The basis of this curve is the observed epochs of maxima and minima, and its accuracy depends solely upon the accuracy with which these have been observed. I refer the reader especially to page 76 of my paper in the February number of the MONTHLY WEATHER REVIEW, where I have discussed the possible inaccuraries.

Prof. Marvin, speaking of the method of tabulation of rainfall data, says: "Exactly the same method has been used by meteorologists almost for centuries." He then proceeds later to criticize the points in which this method differs from the old. To do this he gives a table of months skipped or repeated and shows how much rainfall fell in Washington, D. C., during these months. I would make three replies to this criticism.

(A) The exact form of the method is comparatively new but is already standard. Prof. Schuster, on page 75, *Philosophical Transactions of the Royal Society*, 1906, Volume 206A, makes the first use of it that has come to my notice. In this place he says: "Thus for a period of 7¾ years the alternate rows were formed of 15 and 16 figures. This gives 31 intervals of six months for two complete periods, or, on the average, 7¾ years. In the last column alternate numbers were missing, and this column was omitted in the calculations of A and B, the number being chosen to correspond to the number of columns retained." As an example of a problem in which numbers were repeated I wish to quote from page 461 of Prof. Turner's paper "On the Fifteen Month Periodicity in Earthquake Phenomena" published in *Monthly Notices* for April, 1919. "The cycle was identified (in the B. A. Report for 1912) as of 104/7 months, which can be approximately dealt with either—

(*a*) by repeating a month at the end of seven complete sets of 15 months (7×15=105), or

(*b*) by collecting sets of 15 months in sevens without repetition and then shifting the initial month one to the right for each set.

The first method (*a*) was adopted in the 1912 report. As a variant the second method was adopted here."

(B) In my paper, totals of rainfall are not the data upon which the arguments were based (although their use would have been legitimate), but ratios between two tables built, each with the same months repeated or averaged, the one using actual the other normal rainfall values. This is clearly stated on page 77 of my paper. The most serious objection that could possibly be raised is that the skipping of months lessens the weight of the argument in direct proportion as the number skipped is to the total number. Thus, if one month in six must be skipped in a certain six years' stretch of data and none in another five years' stretch, the number of months used is the same in each case and the weights of the two stretches are equal. Even this slight objection can not apply if the months are averaged instead of skipped.

(C) The method is legitimate in all cases, no matter how frequently months must be repeated or averaged, but even though one should assume the legitimacy of Prof. Marvin's criticism there would be almost no application to the conclusions of the present paper, since almost his whole argument is based on the large amount of repetition necessary in years earlier than the earliest for which we have state averages.

Prof. Marvin criticizes my application of the method of least squares, as he claims, to rainfall. Regardless of the merits of his objection to its application to rainfall data, I would call attention to the fact that I have not so applied it, but have considered only the differences and similarities of two curves already obtained without its use. The whole argument of the paper is based on the similarity of these curves obtained from different stretches of years.

Long records are certainly needed. It is for lack of long State averages that I included the word possible in the title. I would call attention, however, to the fact that the data used for the average State run through approximately 18 complete cycles—not a short record, as tabulations of physical data usually run.

ERRATA: On pp. 78-79, February REVIEW, legends for Figs. 3, 4, and 5 apply to figures marked 4, 5, and 3, respectively.

551.481 (747)

DATES OF THE OPENING OF ONEIDA LAKE, N. Y., 1869-1921.

By ERNEST S. CLOWES.

[1309 East Adams Street, Syracuse, N. Y., Apr. 30, 1921.]

Oneida Lake is the largest in area in central New York, famous for the beauty and number of its lakes. It is about 25 miles long and averages about 6 miles wide over the greater portion of its length. It is distant 25 miles from Lake Ontario, and its central point is approximately the same distance northeast of the city of Syracuse. Its northernmost point is in latitude 43° 15′.

Although the largest, it is the shallowest of the larger lakes of this region, its average depth being 45 feet and its deepest but little more than 60 feet, as contrasted with the 600-foot depth of Cayuga and Seneca Lakes. For that reason it freezes early in the winter and stays frozen usually until spring is fairly set in, although in this variable climate snowstorms after its opening are not unknown. The country immediately surrounding it is flat and marshy on the south and rolling on the north. No river of any size flows into it, but its outlet at its western end is the Oneida River, a navigable stream used as part of the route of the New York State Barge Canal.

The record of its opening in the spring given here was kept by residents of the village of Constantia, on its northern shore. The ice in the spring breaks up suddenly at the last and in the space of a few hours is blown ashore or carried out down the river, so that the opening of the lake may usually be put down as occurring on a single day. The record follows:

1869	April 21	1896	April 19
1870	April 13	1897	April 2
1871	March 15	1898	March 17
1872	April 22	1899	April 20
1873	April 26	1900	April 18
1874	April 15	1901	April 14
1875	April 17	1902	March 29
1876	April 21	1903	March 21
1877	April 19	1904	April 17
1878	March 15	1905	April 11
1879	April 24	1906	April 6
1880	March 5	1907	April 1
1881	April 23	1908	April 2
1882	March 19-20	1909	April 8
1883	April 20	1910	March 25
1884	April 5	1911	April 15
1885	April 25-26	1912	April 18
1886	April 2	1913	March 22
1887	April 24	1914	April 16
1888	April 18	1915	April 12
1889	April 12	1916	April 14
1890	March 27	1917	April 3
1891	April 13	1918	April 15
1892	April 6	1919	March 21
1893	April 14	1920	April 4
1894	March 21	1921	March 16
1895	April 19		