TABLE I.—*Meteorological extremes and averages, San Jose, Calif.*

Data and length of record (years).	January.	February.	March.	April.	May.	June.	July.	August.	September.	October.	November.	December.	Annual.
Temperature (17):													
Highest recorded	76	76	88	90	102	101	103	102	103	97	85	72	103
Lowest recorded	22	25	30	33	35	38	43	42	40	31	27	23	22
Average maximum	57.0	60.3	63.8	68.2	71.0	76.4	80.1	79.7	78.5	73.4	65.2	57.5	69.3
Average *minimum*	38.4	40.9	42.1	43.7	45.9	49.4	52.7	52.2	50.2	46.2	40.9	38.4	45.1
Average monthly and annual	47.7	50.6	53.0	56.0	58.4	62.9	66.4	65.0	64.4	59.8	53.0	48.0	57.2
Average number of days with—													
Maximum 32° or below	0	0	0	0	0	0	0	0	0	0	0	0	0
Maximum 90° or above	0	0	0	0	*	3	3	2	3	*	0	0	11
Minimum 32° or below	6	2	*	0	0	0	0	0	0	*	2	5	16
Precipitation (inches):													
Average (17)	3.83	2.58	2.51	0.56	0.41	0.09	0.01	0.01	0.55	0.49	1.14	3.06	15.24
Greatest monthly and yearly	12.38	7.02	7.75	1.95	2.69	0.46	0.09	0.08	6.33	1.71	4.10	6.39	22.75
Least monthly and yearly	0.10	0.09	0.31	0	0	0	0	0	0	0	0.13	0.43	6.52
Greatest in 24 hours	4.56	2.65	2.60	0.78	1.24	0.36	0.08	0.08	4.47	0.90	1.73	2.77	4.56
Greatest amount in—													
1 hour (12)	0.65	0.45	0.50	0.31	0.21	0.08	0.06	0.08	0.57	0.24	0.39	0.36	0.65
5 minutes (12)	0.17	0.12	0.18	0.08	0.06	0.02	0.02	0.02	0.13	0.08	0.17	0.14	0.18
Number of days with 0.01 inch or more	12	10	9	4	3	1	*	*	2	4	6	11	64
Sunshine and cloudiness:													
Sunshine, per cent of possible	56	57	64	74	75	85	86	83	81	75	69	56	72
Average cloudiness (17)	5.4	5.3	4.7	3.4	3.2	2.1	1.8	2.1	2.4	3.1	3.7	5.3	3.5
Number of days—													
Clear	11	10	14	18	20	24	28	27	22	19	16	11	220
Partly cloudy	6	7	7	7	7	5	3	4	6	8	7	7	74
Cloudy	14	11	10	5	4	1	0	0	2	4	7	13	71
Relative humidity:													
Average at noon (5)	64	65	62	50	50	46	50	49	50	53	58	71	56
Average at 5 *p. m.* (16)	68	63	61	55	54	51	52	53	53	54	61	70	58
Wind (velocity in miles per hour):													
Average velocity (16)	6.1	6.1	5.9	6.3	6.3	6.2	6.1	5.7	5.1	4.8	4.9	5.6	5.8
Maximum velocity (for 5 minutes)	46	48	36	44	42	34	29	36	27	34	35	43	48
Prevailing direction	SE.	SE.	NW.	NW.	NW.	NW.	NW.	NW.	NW.	NW.	NW. and SE.	SE.	NW.

*Less than 1.

Cloudiness, recorded to tenths. Temperature in degrees Fahrenheit.

GROUP DISTRIBUTION AND PERIODICITY OF ANNUAL RAINFALL AMOUNTS.

By ROBERT E. HORTON, Consulting Hydraulic Engineer.

[Voorheesville, N. Y., August, 1922.]

INTRODUCTION.

The longest existing continuous rainfall record is that at Padua, Italy. Inasmuch as this record is of great value, on account of its length, for studies relating to variations in mean rainfall, it is given herewith in convenient form and in English units, in Table No. 1.

Examining a long rainfall record, such as that of Padua, it will be noticed that a large proportion of the total number of years occur in groups of high or low years, such that all the years in a group are either above the mean or else below the mean. This tendency to grouping of like years occurs even where there is no visible indication of an orderly or cyclic arrangement or periodic recurrence of groups. Now if the occurrence of successive like years was a matter of pure chance, then there would be in any record a certain probable number of groups of 2, 3, 4, 5, etc., successive like years. Inasmuch as the occurrence of groups of like years, especially low years, is of great importance in the application of rainfall data, it becomes of interest to inquire how the actual grouping of like years compares with that which would result from a chance distribution of wet and dry years.

The reasoning here applied to rainfall records may also be applied to other hydrologic data, such as the distribution of groups of cold and warm years, the occurrance of groups of low or high run-off in streams, the occurrence of groups of years of large or deficient yield of crops, etc. In order to make the discussion general, using rainfall records for purposes of illustration, values of the data under discussion will be described as "events"; a series of events which occur in the same way, as, for example, rainfall years all above the mean, or the results of the tossings of a coin where all cases are heads, will be described as "like events"; events which occur in opposite ways, as the alternate tossings of heads and tails with a coin, or the occurrence of wet and dry years in succession, will be described as "unlike events"; events which correspond to the occurrence of heads in the tossing of coins, or to years of rainfall, stream yield, crop yield, temperature, etc., greater than the mean, will be called "plus events," and the opposite will be called "minus events." Any series of n consecutive events, whether like or not, may be called a group. A series of n events which are all like, and which events are both preceded and succeeded by at least one unlike event, will be described as an "n group" of events. A series of events arranged in the order of their occurrence will be described as a "record." The difference in value of an event from the mean value will be called its "departure." In the discussion of the accuracy of record means, it is sometimes convenient to use the fiction "true mean," meaning thereby the result which would be obtained from a record of indefinitely great length containing no observational errors, as distinguished from the mean of any series of m events. The illustrations here given are mainly derived from rainfall, for which the letter p is commonly used to designate the mean. In order to make this discussion more general, and to reserve the letter p to designate probability, the letter M will be used to designate the mean, whether of rainfall or some other series of events.

TABLE No. 1.—*Annual precipitation, Padua, during 176 years, 1725–1900, deduced from tables of Dr. Julius Hann.*

Year.	Inches.	Departure.	Ratio to mean.	Per cent departure.
1725	29.90	−4.01	0.88	−0.12
1726	24.78	−9.13	0.73	−0.27
1727	45.78	+11.87	1.35	+0.35
1728	53.11	+19.20	1.57	+0.57
1729	36.41	+2.50	1.07	+0.07
1730	34.32	+0.41	1.01	+0.01
1731	34.12	+0.21	1.01	+0.01
1732	32.07	−0.84	0.95	−0.05
1733	34.63	+0.72	1.02	+0.02
1734	37.86	+3.95	1.12	+0.12
1735	30.54	−3.37	0.90	−0.10
1736	31.20	−2.71	0.92	−0.08
1737	23.87	−10.04	0.71	−0.29
1738	28.09	−5.82	0.83	−0.17
1739	25.41	−8.50	0.75	−0.25
1740	22.46	−11.45	0.66	−0.34
1741	24.19	−9.72	0.72	−0.28
1742	39.04	+5.13	1.15	+0.15
1743	28.25	−5.66	0.83	−0.17
1744	35.70	+1.79	1.05	+0.05
1745	37.94	+4.03	1.12	+0.12
1746	41.61	+7.70	1.23	+0.23
1747	25.57	−8.34	0.76	−0.24
1748	41.45	+7.54	1.22	+0.22
1749	36.17	+2.26	1.07	+0.07
1750	32.35	−1.56	0.96	−0.04
1751	42.28	+8.37	1.25	+0.25
1752	37.86	+3.95	1.12	+0.12
1753	39.37	+5.46	1.16	+0.16
1754	27.78	−6.13	0.82	−0.18
1755	43.38	+9.47	1.28	+0.28
1756	39.16	+5.25	1.16	+0.16
1757	31.44	−2.47	0.93	−0.07
1758	43.77	+9.86	1.29	+0.29
1759	36.21	+2.30	1.07	+0.07
1760	34.95	+1.04	1.03	+0.03
1761	44.37	+10.46	1.31	+0.31
1762	22.54	−11.37	0.66	−0.34
1763	37.31	+3.40	1.10	+0.10
1764	42.12	+8.21	1.24	+0.24
1765	38.69	+4.78	1.14	+0.14
1766	33.13	−0.78	0.98	−0.02
1767	35.54	+1.63	1.05	+0.05
1768	30.06	−3.85	0.89	−0.11
1769	41.80	+7.89	1.23	+0.23
1770	54.06	+20.15	1.60	+0.60
1771	41.84	+7.93	1.23	+0.23
1772	61.58	+27.67	1.82	+0.82
1773	44.48	+10.57	1.31	+0.31
1774	29.94	−3.97	0.88	−0.12
1775	36.09	+2.18	1.07	+0.07
1776	37.15	+3.24	1.10	+0.10
1777	46.65	+12.74	1.38	+0.38
1778	34.12	+0.21	1.01	+0.01
1779	32.54	−1.37	0.96	−0.04
1780	31.99	−1.92	0.94	−0.06
1781	37.04	+3.13	1.09	+0.09
1782	32.35	−1.56	0.96	−0.04
1783	32.50	−1.41	0.96	−0.04
1784	31.16	−2.75	0.92	−0.08
1785	36.17	+2.26	1.07	+0.07
1786	40.86	+6.95	1.21	+0.21
1787	33.57	−0.34	0.99	−0.01
1788	32.46	−1.45	0.96	−0.04
1789	29.47	−4.44	0.87	−0.13
1790	23.52	−10.39	0.69	−0.31
1791	26.75	−7.16	0.79	−0.21
1792	26.40	−7.51	0.78	−0.22
1793	37.31	+3.40	1.10	+0.10
1794	41.02	+7.11	1.21	+0.21
1795	35.34	+1.43	1.04	+0.04
1796	33.29	−0.62	0.98	−0.02
1797	28.45	−5.46	0.84	−0.16
1798	34.20	+0.29	1.01	+0.01
1799	44.17	+10.26	1.30	+0.30
1800	39.49	+5.58	1.17	+0.17
1801	35.58	+1.67	1.05	+0.05
1802	43.73	+9.82	1.29	+0.29
1803	45.00	+11.09	1.33	+0.33
1804	44.52	+10.61	1.31	+0.31
1805	35.18	+1.27	1.04	+0.04
1806	39.04	+5.13	1.15	+0.15
1807	46.61	+12.70	1.38	+0.38
1808	29.08	−4.83	0.86	−0.14
1809	42.91	+9.00	1.27	+0.27
1810	38.34	+4.43	1.13	+0.13
1811	26.04	−7.87	0.77	−0.23
1812	40.54	+6.63	1.20	+0.20
1813	28.72	−5.19	0.85	−0.15
1814	33.06	−0.85	0.98	−0.02
1815	28.29	−5.62	0.84	−0.16
1816	21.79	−12.12	0.64	−0.36
1817	22.66	−11.25	0.67	−0.33
1818	32.58	−1.33	0.96	−0.04
1819	34.00	+0.09	1.003	−0.003
1820	20.29	−13.62	0.60	−0.40
1821	26.44	−7.47	0.78	−0.22
1822	18.91	−15.00	0.56	−0.44
1823	30.69	−3.22	0.91	−0.09
1824	30.06	−3.85	0.89	−0.11
1825	26.00	−7.91	0.77	−0.23
1826	39.01	+5.10	1.15	+0.15
1827	33.14	−0.77	0.98	−0.02
1828	24.31	−9.60	0.72	−0.28

TABLE No. 1.—*Annual precipitation, Padua, during 176 years, 12–57 1900, deduced from tables of Dr. Julius Hann*—Continued.

Year.	Inches.	Departure.	Ratio to mean.	Per cent departure.
1829	31.52	−2.39	0.93	−0.07
1830	23.52	−10.39	0.69	−0.31
1831	31.13	−2.78	0.92	−0.08
1832	26.52	−7.39	0.78	−0.22
1833	37.39	+3.48	1.10	+0.10
1834	19.74	−14.17	0.58	−0.42
1835	35.93	+2.02	1.06	+0.06
1836	37.74	+3.83	1.11	+0.11
1837	37.51	+3.60	1.11	+0.11
1838	32.35	−1.56	0.96	−0.04
1839	31.87	−2.04	0.94	−0.06
1840	24.19	−9.72	0.71	−0.29
1841	28.09	−5.82	0.83	−0.17
1842	27.54	−6.37	0.81	−0.19
1843	29.63	−4.28	0.87	−0.13
1844	36.76	+2.85	1.09	+0.08
1845	49.84	+15.93	1.47	+0.47
1846	37.71	+3.80	1.11	+0.11
1847	29.47	−4.44	0.87	−0.13
1848	33.21	−0.70	0.98	−0.02
1849	26.79	−7.12	0.79	−0.21
1850	37.35	+3.14	1.10	+0.10
1851	40.58	+6.67	1.20	+0.20
1852	29.43	−4.48	0.87	−0.13
1853	40.74	+6.83	1.20	+0.20
1854	28.29	−5.62	0.84	−0.16
1855	44.88	+10.97	1.33	+0.33
1856	39.99	+6.08	1.18	+0.18
1857	25.73	−8.18	0.76	−0.24
1858	31.76	−2.15	0.94	−0.06
1859	31.87	−2.04	0.94	−0.06
1860	34.83	+0.92	1.03	+0.03
1861	21.00	−12.91	0.62	−0.38
1862	44.44	+10.53	1.31	+0.31
1863	31.09	−2.82	0.92	−0.08
1864	31.13	−2.78	0.92	−0.08
1865	24.39	−9.52	0.72	−0.28
1866	28.80	−5.11	0.85	−0.15
1867	33.96	+0.05	1.003	+0.003
1868	35.97	+2.06	1.06	+0.06
1869	42.67	+8.76	1.26	+0.26
1870	27.50	−6.41	0.81	−0.19
1871	28.05	−5.86	0.83	−0.17
1872	42.87	+8.96	1.27	+0.27
1873	36.21	+2.30	1.07	+0.07
1874	28.37	−5.54	0.84	−0.16
1875	32.35	−1.56	0.96	−0.04
1876	40.66	+6.75	1.20	+0.20
1877	38.26	+4.35	1.13	+0.13
1878	34.91	+1.00	1.03	+0.03
1879	33.57	−0.34	0.99	−0.01
1880	29.43	−4.48	0.87	−0.13
1881	26.95	−6.96	0.80	−0.20
1882	32.03	−1.88	0.94	−0.06
1883	27.62	−6.29	0.82	−0.18
1884	30.26	−3.65	0.89	−0.11
1885	35.14	+1.23	1.04	+0.04
1886	32.90	−1.01	0.97	−0.03
1887	35.78	+1.87	1.06	+0.06
1888	25.53	−8.38	0.75	−0.25
1889	37.94	+4.03	1.12	+0.12
1890	27.15	−6.76	0.80	−0.20
1891	26.44	−7.47	0.78	−0.22
1892	39.04	+5.13	1.15	+0.15
1893	22.85	−11.06	0.67	−0.33
1894	22.18	−11.73	0.66	−0.34
1895	37.67	+3.76	1.11	+0.11
1896	46.73	+12.82	1.38	+0.38
1897	30.93	−2.98	0.91	−0.09
1898	38.73	+4.82	1.14	+0.14
1899	23.21	−10.70	0.69	−0.31
1900	39.12	+5.21	1.15	+0.15
Mean	33.91	....	....	....

PROBABILITY OF GROUPS OF WET AND DRY YEARS—NOTATION.

M = true mean of a series of values of a given class of events.

M_m = mean of any series of m events.

n = number of like events in a group of like events.

Δ = departure of any event from the mean.

P_1 = probability of the occurrence of any given event.

p_n = probability of occurrence of a group of n-like events.

m = number of events in a series.

P_e = probability of exceedance of an event.

T_e = average exceedance interval.

N_g = probable number of like groups of n events in a record comprising m events.

N_n = probable number of events concentrated in groups of n-like events each in a record of m events.

If a rainfall record covering m years contains d years with precipitation less than the mean, and w years with precipitation greater than the mean, then the probability of occurrence of dry years is $\frac{d}{d+w}$ and the probability of occurrence of wet years is $\frac{w}{d+w}$.

If it is assumed that wet and dry years are equally probable, then the probability $p_1 = 1/2$ for either a wet or a dry year. In accordance with a well-known theorem of probabilities, the probability of the occurrence of n-like events in succession, where n trials are made, is p^n, or in the case assumed $(\frac{1}{2})^n$. In this theorem it is assumed that the events are independent; in other words, if an urn contains equal numbers of black and white balls, the probability that if six balls are drawn, they will all be black, is $\left(\frac{1}{2}\right)^6$.

This has no regard to the question whether the next subsequent ball drawn would be black or white. The question of the occurrence of groups of n wet or n dry years in a rainfall record is of somewhat a different nature, and is not subject merely to the conditions that there shall be some number n of like years in succession, but also to the condition that after these events have occurred the next event shall be of the opposite kind.[1]

Treating the black balls in the example of the urn as wet years, and the white balls as dry years, the probability that the first two years of a series will be wet years is 1/4, and the probability that the next subsequent year will be dry is 1/2; therefore the probability that the two wet years will occur, followed by a dry year, is 1/8. Similarly, the probability of occurrence of three successive like events is $\left(\frac{1}{2}\right)^3$, but the probability that these three will occur and that they will be followed by an unlike event is $(\frac{1}{2})^4$. One group of three like events—plus, for example—having occurred, then, in order that these may be succeeded by another group, the next succeeding group must be made up of minus events. The chance that the first event of the succeeding group will be minus is $\frac{1}{2}$; the chance that three minus events will occur in succession is $(\frac{1}{2})^3$; and the chance that these will occur and will be followed by a plus event, which is necessary to constitute a like group of three events, is $(\frac{1}{2})^4$. In other words, when a group of three like events has occurred, the probability of its being succeeded by another group of three like events of the opposite kind is—

$$P_n = (\tfrac{1}{2})^{n+1} = (\tfrac{1}{2})^4 = \frac{1}{16}$$

The same reasoning may be applied to any number of years, with the result that the relative probabilities of occurrence of wet or dry years in groups of n years each, is $(\frac{1}{2})^{n+1}$ if wet and dry years are equally probable, or if wet and dry years are not equally probable, then we have—

$$P_d = \tfrac{1}{2}\left(\frac{d}{d+w}\right)^{n+1}$$
$$P_w = \tfrac{1}{2}\left(\frac{w}{d+w}\right)^{n+1} \quad (1)$$

Writing the formula[2] in the form to give the probability of either a plus or a minus group of n years, where both are equally probable

$$P_n = (\tfrac{1}{2})^{n+1} \quad (2)$$

or

$$\log P_n = (n+1) \log \tfrac{1}{2} = -0.698970\ (n+1)$$
$$= -0.70\ (n+1) \text{ approx.} \quad (3)$$

P_n is less than 1.0 and its logarithm is negative.

If N_g = probable number of occurrences of a group of n like events in a record of m events, then—

$$N_g = mP_n = m\ (\tfrac{1}{2})^{n+1} \quad (4)$$

Also the length m of record in which there will be on an average one group of n-like events is obtained from the relation—

$$1 = m_n\ (\tfrac{1}{2})^{n+1}$$

or

$$\log \frac{1}{m_n} = 0.698970\ (n+1)$$

or

$$\log m_n = 0.70\ (n+1) \text{ approx.} \quad (5)$$

The probable number of events which will be concentrated in groups of n-like events each is—

$$N_n = mN_g = mn\ (\tfrac{1}{2})^{n+1} \quad (6)$$

For example, the length of record in which a like group with $n = 15$ will occur once on an average is 15,259,000,000.

Using the logarithmic expression, equation (3), it is evident that the graph of this formula will plot as a straight line on semi-logarithmic paper. Such a plotting is shown in Figure 1, from which the relative probability of occurrence of n-like groups up to $n = 19$ can be read accurately in spite of small values of P.

The author has been accustomed to use the term "exceedance frequency" or "exceedance probability" in the peculiar sense of the frequency or probability of a given value of an event being *either equalled or exceeded.* The term will be used in the same sense here. Referring to column (2) of Table No. 2, it will be noted that the sum of the probability series for a given value of n, and for all greater values, is equal to the probability for $(n-1)$. In other words, the exceedance frequency for any value of n may be expressed by the formula—

$$P_e = P_{(n-1)} \quad (7)$$

For example, the probability of occurrence of a group of either nine or any greater number of like years in a given record is the same as the probability of occurrence of a group of just eight like years, namely, $\frac{1}{512}$.

[1] Woodward: *Probability and theory of errors.*

[2] After this paper was submitted for publication, the author's attention was called to the fact that Mr. H. W. Clough (Mo. WEATHER REV., March, 1921, pp. 125 (5) and 128 (5)), had given the formula for the number of n-like groups in a record of m events, which can be transformed into an expression identical with formula (2) given in this paper. The author has not been able to find this formula in any other published work on the theory of statistics and it was worked out independently in the preparation of this paper. Mr. Clough does not state the source from which he obtained his formula nor give its derivation. He is, however, authority for the statement that the same method was used by Buys Ballot in a study of temperature (*Met. Zeit.*, 1881, p. 404). The object of the present paper is to present the matter of practical application of this important formula, with special reference to rainfall, in a manner readily intelligible to those not versed in the advanced theories of statistics.

TABLE 2.—*Probability of groups of n-like events either plus or minus—Probable number of n groups in 1,000 events for various values of n number of events included in each class of n groups—Actual numbers of events in 1,000 coin tossings.*

[$P_s = (\frac{1}{2})^{n+1}$ plus and minus events equally probable.]

n	Relative probability of an n group.	Probable number of n groups in 1,000 events.	Number of events in 1,000 fallings in n groups.	Actual number in 1,000 coin tossings. Series I. Groups.	Series I. Events.	Series II. Groups.	Series II. Events.	Mean. Groups.	Mean. Events.
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
1	0.250	250	250	263	263	238	238	250	250
2	.125	125	250	120	240	117	234	118	237
3	.0625	62.5	187.5	55	165	60	180	58	172
4	.03125	31.25	125	27	108	34	136	30	122
5	.01562	15.62	78.1	18	90	15	75	16	83
6	.007812	7.81	46.86	8	48	10	60	9	54
7	.003906	3.91	27.37	5	35	5	35	5	35
8	.001953	1.95	15.60	4	32	3	24	3	28
9	.000977	.977	8.793	1	9	2	18	1	11
10	.000488	.488	4.88	1	10			1	5
11	.000244	.244	2.70						
12	.000122	.122	1.46						
13	.0000611	.0611	.794						
14	.0000306	.0305	.427						
15	.0000152	.0152	.228						
16	.00000763	.00763	.122						
17	.00000382	.00382	.0694						
18	.00000191	.00191	.0344						
19	.000000955	.000955	.0181						
20	.000000477	.000477	.00954						

In the accompanying Table 2, assuming wet and dry years equally probable, column (2) shows the relative probability of occurrence of groups of years all alike, either wet or dry, each group containing the number of years indicated by n in column (1). Column (3) shows the probable number of groups of n years each which would occur in a record of 1,000 years. The resulting numbers of years which would probably be included in groups of n-like years each in a record of 1,000 years is shown in column (4). As a test, two series of 1,000 coin tossings were made under conditions insuring chance occurrence of heads (plus) and tails (minus). The results for each series and the mean of the two are shown in Table 2.

In general the numbers of groups and numbers of events falling in each n group agree well with the results given by the formulas. A peculiarity of formula (1) is that the actual number of years comprised in groups of two like years as indicated by the formula should be the same as the number of years occurring singly. This of course results from the fact that the probability of two like years in succession is one-half that of an isolated high or low year, whereas the number of years in a group of two is double that in "groups" of one year.

TABLE 3.—*Actual numbers of groups of wet and dry years in various precipitation records.*

	Padua, 1725 to 1900 (175 years); mean, 38.91.			Havana, 1859 to 1914 (56 years); mean, 48.69.			New Bedford, 1814 to 1913 (100 years); mean, 46.47.			Cincinnati, 1850 to 1914 (65 years); mean, 39.23.			Albany, 1826 to 1918 (93 years); mean, 37.68.			San Diego[1] 1850–51 to 1917–18 (68 years); mean, 9.72.		
M=	+	−	Σ	+	−	Σ	+	−	Σ	+	−	Σ	+	−	Σ	+	−	Σ
1	17	19	36	6	6	12	10	9	19	9	9	18	4	7	11	11	11	22
2	8	8	16	2	0	2	5	4	9	5	2	7	4	3	7	0	1	1
3	8	3	11	2	2	4	1	0	1	0	3	3	1	2	3	1	4	5
4	2	1	3	0	3	3	1	3	4	1	0	1	1	1	2	5	1	6
5	2	0	2	0	1	1	0	1	1	0	1	1	2	1	3	0	1	1
6	0	6	6				1	0	1	0	0	0						
7	0	1	1				0	0	0	1	0	1						
8							0	2	2	0	1	1						
9							1	1	2				1	0	1			
10	1	0	1										1	0	1			
11				1	0	1												
17													0	1	1			

[1] Seasonal, July-June, inclusive.

TABLE 4.—*Relative numbers of groups of n-like years either plus or minus in various precipitation and other records, reduced to basis of 100 years record.*

n	Probable number of groups.	Relative sun spot Nos. 1750–1912 (162 years).	Padua precipitation.	Havana precipitation.	New Bedford precipitation.	Cincinnati precipitation.	Albany precipitation.	San Diego seasonal precipitation.	San Diego temperature 1852–1901 (50 years).
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
1	25	1.23	20.44	21.42	19	27.70	11.88	32.30	22
2	12.5	0	9.09	3.57	9	10.78	7.56	1.47	2
3	6.25	1.85	6.25	7.14	1	4.62	3.24	7.35	6
4	3.12	3.08	1.70	5.36	4	1.54	2.16	8.82	6
5	1.56	4.94	1.14	1.78	1	1.54	3.24	1.47	
6	0.78	2.47	3.40		1				
7	0.39	1.23	.57		0	1.54			2
8	0.20	0.62			2				
9	0.10	0.62			2		1.47		2
10	0.05	1.23	.57				1.47		
11	0.02	0.62		1.78					
17							1.47		
Σ	50	20	43.16	41.05	39	47.72	32.49	51.41	40.00
n_a	2.00	5.00	2.32	2.44	2.56	2.10	3.08	1.94	2.50
C_p	1.00	2.50	1.16	1.22	1.28	1.05	1.54	0.97	1.25

As a test of the applicability of the theoretical formula here deduced to rainfall records, the actual number of occurrences of n-like groups, both plus and minus, in several rainfall records, have been tabulated as shown in Table 3. The results are more readily comparable in the form in which they are presented, together with certain other data, in Table 4, in which the actual numbers of occurrences of n-like groups have been reduced to the basis of equivalent numbers in a record of 100 events. To the data given in Table 3 there have been added, in Table 4, the theoretical numbers of n-like groups (column (2)) the actual numbers of n-like groups in Wolf-Wolfers relative sun-spot numbers (column (3)) and the occurrence of groups of cold and warm years in succession in the temperature record at San Diego, Calif. (column 10). Comparing columns (2) and (4), Table 4, it will be noted that there is a slight deficiency of groups for $n=1$ and $n=2$. There is a considerable excess of groups for $n=6$ in the Padua rainfall record and some deficiency in the occurrence of groups for most other values of n. The formula would indicate the probable occurrence of 0.78 of a group of six like years on an average per 100 years of record, where actually 3.4 or nearly five times as great a number of groups of six like years occurred. This seems to indicate a somewhat pronounced tendency to repetition of six-year periods in the Padua rainfall, and six being the nearest integer to the length of one-half of the 11.3-year sun-spot cycle, a connection between the two might be inferred. There is, however, no great regularity in the time of occurrence of the six-year groups in Padua, as will be readily noted from Figure 1. A comparison of the actual with the probable distribution of groups of like years, regardless of their times of occurrence, by the methods hereby given, may, however, lead to indications of a periodicity subject to discontinuities, as suggested by Turner.[3]

In these studies like years at the beginning and end of a record have been counted as complete like groups, although it is not known whether the terminal groups in such cases were really complete or not; furthermore, fractional groups can not occur, although for large values of n only a fractional group is indicated in general on the average for 100 years of record. As a result, one

[3] Turner: Discontinuities in meteorological events. *Quar. Jour. Roy. Met. Soc.*, 41: 315-352, 1915.

finds oftentimes a single group with a relatively large value of n and the total omission of several groups with still larger values of n, the result being that, as for example in the case of the Padua precipitation, the total occurrence of groups with n greater than 7 is not far different from that indicated by chance. In the case of New Bedford and Albany precipitation, there seems to be a tendency for a recurrence of groups of about 9 years' average duration, with a frequency 15 to 20 times as great as that indicated by the laws of chance. Referring to Table 3, it will be noted that the Albany record contains one group of 17 like years preceding the year 1919. The record for 1919 shows that this was also deficient relative to the 93-year mean, which would make this group comprise 18 years, and it may not yet be completed. However, analyzing the data for 46 years, since the establishment of the United States Weather Bureau station at Albany, this group becomes subdivided into four like groups, the longest of which comprises 12 low rainfall years.

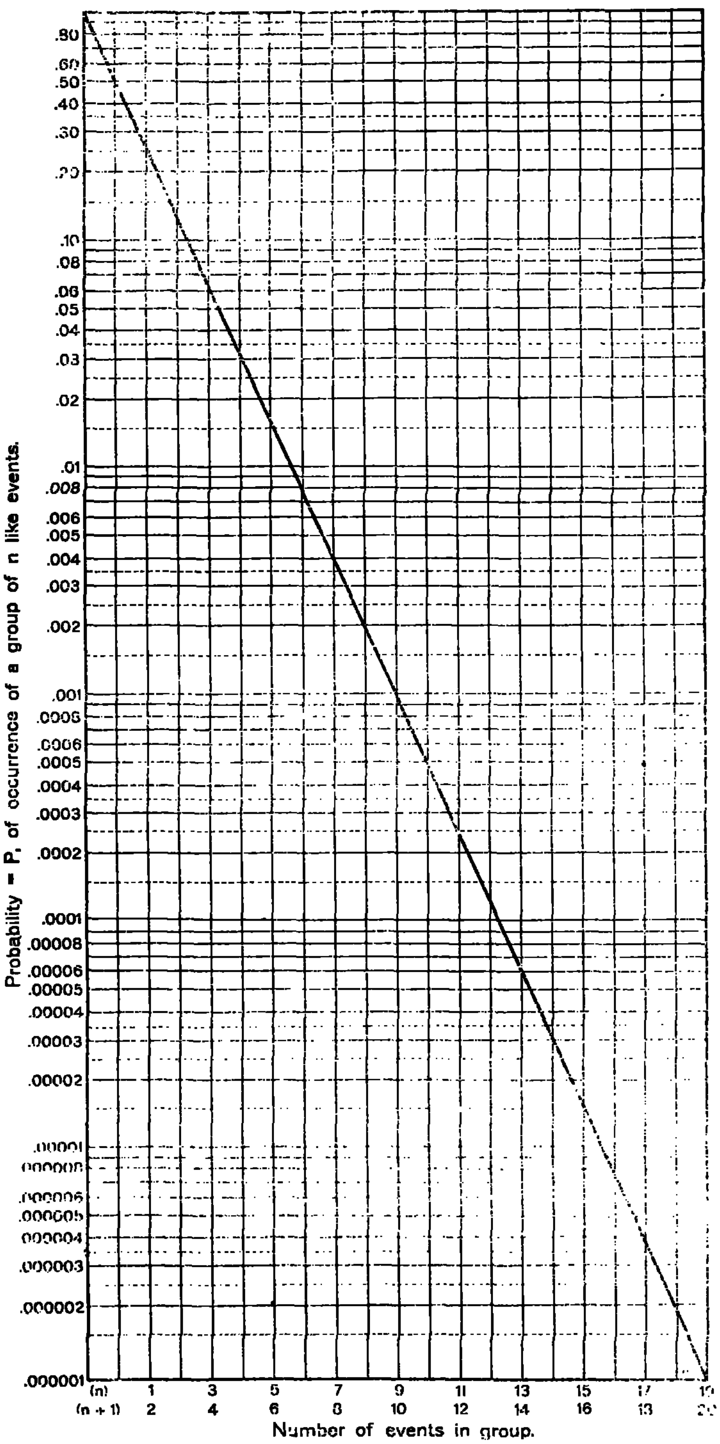

FIG. 1.—Probability of the occurrence of a group of n-like events.

In the case of the possible six-year periodicity indicated for Padua, two or three of the six-year groups would have become five-year groups by a very slight change in the long term mean, less than the probable error of the observations.

A CRITERION OF PERIODICITY.

Arthur Shuster (ref. 8) pointed out in 1902 that evidence of periodicity can be obtained by the use of Fourier's series in any record of annual rainfall or similar data, and furthermore, since a certain number of equal like groups would occur as a result of chance, the true indication of periodicity is not the total number of occurrences of like groups of n values each in a given record, but only the excess in the number of such occurrences over the probable chance occurrences. As will be seen from the above illustrations, formula (2) supplies a method of determining this excess in the case of rainfall and similar records.

It also affords a more direct and precise criterion of periodicity, to which attention will now be called.

Referring to column (2) of Table 2, it is evident that the sum of the probability series is $\frac{1}{2}$; in other words, the probable total number of like groups of events in a record of m events is $\frac{1}{2}m$, consequently the average length of all the periods of n-like events occurring in such a record, in accordance with the laws of chance, would be 2.0. If the events do not occur in accordance with the laws of chance, but are controlled by causes which render the occurrence of events more or less regularly periodic, then it is evident that the average number of events in a like group would approach one-half of the length of the complete cycle or period as a limit in the case of perfect periodicity. These facts furnish a simple criterion of the tendency to periodicity in any record, which may be calculated as follows:

Calling consecutive events all above or all below the mean like events, count the number of groups of like events in the record, including as a group each single or isolated plus or minus event. Let this number of groups be G, where m is the total number of events. Then the average number of events in a like group is $\frac{m}{G}$.

The average number for chance distribution is 2; hence the

$$\text{Periodicity criterion, } C_p = \frac{m}{2G} \qquad (8)$$

This criterion is one-half of the number of like events in an average group.

It will be seen that the periodicity criterion is simply the ratio of the average number of like events per group in a given record to two, the number of events per group for chance distribution. Its numerical value is 1.00 in the case of pure chance occurrence of the events; whereas for purely cyclic events free from all chance variation its value is equal to precisely one-fourth the length of the cycle.

Of course there will be some departure from the theoretical group distribution in any finite series of pure chance events. In general the longer the series of fortuitous events the more closely will the grouping agree with the calculated distribution. If the occurrence of wet years in a rainfall record, for example, is a matter of pure chance, then the periodicity criterion determined from the record may be expected to approach the minimum value 1.00 more and more closely as the duration of the record increases.

This criterion has an advantage over the use of Fourier's series in that it applies equally well to discontinuous periodicity in the sense of the term suggested by Turner (*loc. cit.*) as to periodicities repeated at regular intervals with reference to a fixed epoch. It is sensitive for the reason that if in any record there is a preponderance of groups of n-like events, the excess of events comprised in these groups must be drawn from other groups, and there will be a corresponding deficiency in the other groups, especially in the more numerous smaller groups, so that the average values of n and C_p increase rapidly as the tendency to periodicity increases.

Referring to the average values of n for the different records as given at the foot of Table 4, it will be noted that in the case of sunspot numbers, the value of C_p, periodicity criterion, is 2.5. This is not far from one-half of the average or 11.3-year sunspot period. In other words, the criterion in this case approaches closely the limiting value, although the periodicity of sunspot numbers is by no means perfect. In the case of rainfall records, the periodicity criterion is very nearly 1.0 in two instances; ranges from 1.16 to 1.28 in four instances; and exceeds 1.5 only in one instance, namely, Albany. In this case the large value is mainly due to the inclusion of the 17-year group of low years above cited. It will be noted that lack of homogeneity, in the record will tend to increase the criterion above its value 1.0 for pure chance occurrences, so that in so far as the criterion here used indicates, there is little valid evidence of definite periodicity in the rainfall records cited in Table 4, comparable with the periodicity of sunspot numbers.

Another disturbing factor in most of the rainfall records used in Table 4 arises from the fact that the totals are for calendar years. If, as is usually the case, the rainfall shows a marked seasonal variation, then seasonal rather than annual totals will usually give more consistent results. As an illustration, the San Diego records of precipitation have been analyzed on the basis of seasonal totals, and when so analyzed show a periodicity criterion very close to the minimum value corresponding to a purely chance occurrence of the distribution of wet and dry years.

The periodicity criterion can be applied to various other hydrologic data as well as to rainfall. Examples of the grouping of wet and dry years and the values of the periodicity criterion for run-off records are shown in Table 5.

The criterion of periodicity gives of course only the weighted average of the various periodicities if more than one exists in a given record. If one peridicity predominates it is indicated by the criterion. The criterion, however, fails in the case of the perfect periodicity of two years' cycle, since for that case the value of the criterion is the same for perfect periodicity as it would be for a purely chance occurrence of the events.

TABLE 5.—*Periodicity criterion—Annual run-off.*

[Water year, November, 1884–November, 1911.]

Group duration, n.	L. Cochituate, Mass., 1864–1915.[1]	Neshaminy Creek, Pa.[2]	Tohickon Creek, Pa.[3]	Perkiomen Creek, Pa.[3]	Merrimac River, Lawrence, Mass.,[3] 1880–1915.	Murray River, Mildura, Australia,[1] 1865–1918.	Probable N_a groups in 100.
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1	7	6	6	5	8	11	25.00
2	4	2	2	4	1	5	12.50
3	......	3	3	2	1	2	6.25
4	4	2	2	2	3	1	3.12
5	......	0	0	......	1	......	1.56
6	2	......	......	......	1	2	0.78
7	......	......	......	......	......	......	0.39
8	......	......	......	......	......	......	0.20
9	1	......	......	......	......	......	0.10
10	......	......	......	......	......	......	0.05
11	......	......	......	......	......	1	0.02
12	......	......	......	......	......	......	......
Years	52	27	27	27	36	54	49.97
ΣG	18	13	13	13	15	22	100.00
n_a	2.9	2.08	2.08	2.08	2.4	2.45	2.00
C_p	1.45	1.04	1.04	1.04	1.20	1.22	1.00

[1] Hall P. A. S. C. E.
[2] W. S. Papers, U. S. G. S.
[3] W. S. Paper 415.

CORRECTED CRITERION FOR SHORT RECORDS.

The preceding discussion is based on the assumption that formula (2) gives correctly the probability of the occurrence of a group of n-like years in a series of purely fortuitous events. This is true only in case the record of events is of indefinitely great length, or, in other words, if m is very large. As already pointed out, in the case of a relatively short record it is not known whether the initial and terminal groups are complete or not. In the case of such a short record what may be termed an "end correction" is required to be applied to formula (2) in order to determine the true probability of the fortuitous occurrence of a group of n-like events. After the manuscript of this paper was submitted for publication the author's attention was directed to a paper by Mr. H. W. Clough,[4] in which there is given a formula from which this end correction can be determined. To distinguish them from the approximate values for short records derived from the formulas already given, the true probability of a group of n-like fortuitous events and the corrected periodicity criterion will be designated P'_n and C'', respectively. Using this notation the formula given by Mr. Clough may be written

$$P'_n = \frac{m-n-1}{m}\left(\frac{1}{2}\right)^{n+1} \qquad (9)$$

Values of the correction factor $\frac{m-n-1}{m}$ for any chosen values of m and n can be read directly from the diagram, Figure 2. Since the correction factor is less than unity, it follows that the probable number of groups of like events in a record of m events, as given by formula (9), is less than as given by formula (2). Hence the true average number of events per group in a short record is somewhat greater than the minimum average number for a long record, as given by formula (2). Again, the true value of the periodicity criterion may be written

$$C''_p = \frac{m}{kG} \qquad (10)$$

[4] Cf. footnote 2.

where G as before is the actual average number of groups of like events in a given record, and $\bar{k}$ is the true average number of events per group in a record of purely fortuitous events of the same duration, m. The value of $\bar{k}$ replaces the value 2.00 in formula (8) and its value in general is somewhat greater than 2.00. The value of $\bar{k}$ can be determined without great labor for any duration of record, m, by applying the correction factors from Figure 2 to the calculated number of groups for different values of n for fortuitous events. Calculation of k for a record of 100 events is given on Table 6. The calculation has been carried through only for groups of 15 events or less. It is not necessary to go farther for the reason that although

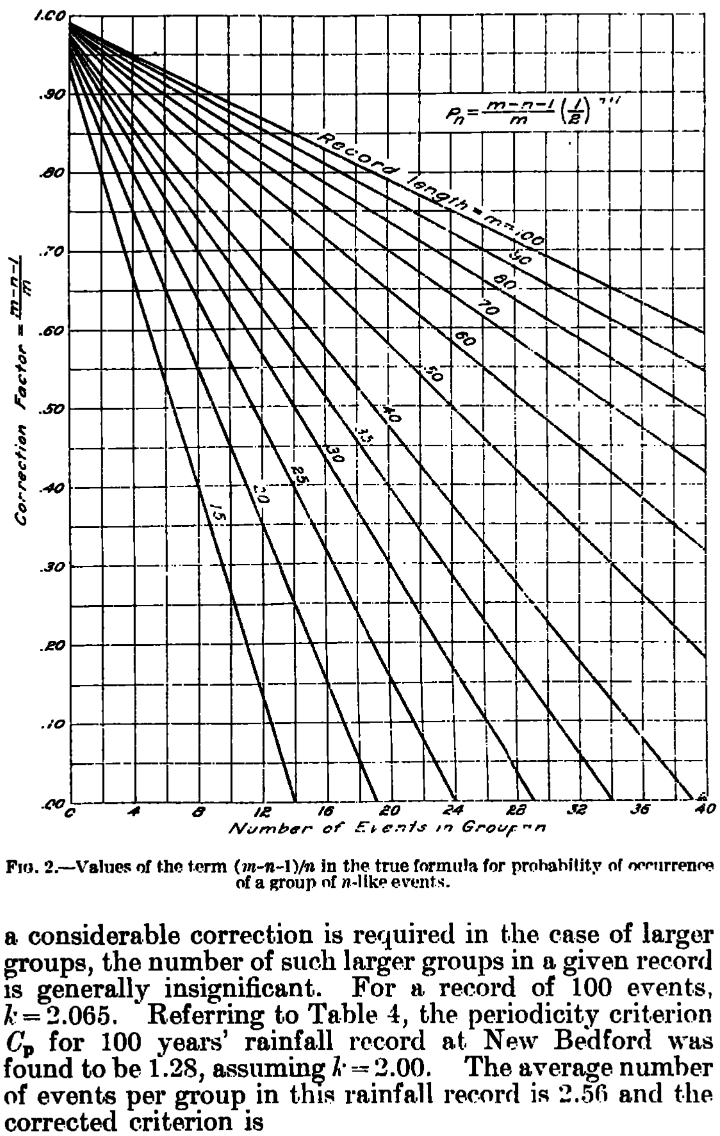

FIG. 2.—Values of the term $(m-n-1)/n$ in the true formula for probability of occurrence of a group of n-like events.

a considerable correction is required in the case of larger groups, the number of such larger groups in a given record is generally insignificant. For a record of 100 events, $k = 2.065$. Referring to Table 4, the periodicity criterion C_p for 100 years' rainfall record at New Bedford was found to be 1.28, assuming $k = 2.00$. The average number of events per group in this rainfall record is 2.56 and the corrected criterion is

$$C'_p = \frac{2.56}{2.065} = 1.24 \text{ instead of } 1.28.$$

The difference is commonly of little significance.

TABLE 6.—*Calculation of corrected average number of events per group in a record of 100 fortuitous events.*

n = number of like events per group.	$N_g = m(\frac{1}{2})^{n+1}$ for $m = 100$.	Correction factor.	Corrected value of N_g.
(1)	(2)	(3)	(4)
1	25	.98	24.500
2	12.5	.97	12.125
3	6.25	.96	6.000
4	3.12	.95	2.964
5	1.56	.94	1.466
6	0.78	.93	0.725
7	0.39	.92	0.359
8	0.20	.91	0.182
9	0.10	.90	0.009
10	0.05	.89	0.044
11	0.025	.88	0.022
12	0.012	.87	0.010
13	0.006	.86	0.005
14	0.003	.85	0.003
15	0.0015	.84	0.001
Total	49.9975		48.415

Average events per group = 100/48.42.
$\bar{K} = 2.065$.

Of course, short records are of little value as a basis of determining periodic cycles. The author believes that it is better in general to adhere to the simple, though approximate, formula $C_p = \frac{m}{2G}$ (8) in most cases for the reasons:

(1) The calculation is much simpler.

(2) This formula gives for any record of finite duration a value larger than the true value of C_p'; in other words it indicates a rather greater probability of periodicity than actually exists in the given record. If, then, the approximate formula (8) does not show definite evidence of periodicity in any given record of events, it affords rather positive evidence that no actual periodicity exists, and it is useless to search further or to analyze the data by the laborious application of the Fourier series or other advanced statistical methods. The object of the method is to quickly eliminate cases where there is no real, positive evidence of periodicity. In this sense formula (8) is perhaps more properly a criterion of non-periodicity than a periodicity criterion. What it really shows is the tendency, if any exists, toward a greater degree of grouping of like events than would probably occur fortuitously. It is true that in any individual record, especially one of short duration, considerable departures from the ideal fortuitous distribution of events can and not infrequently will occur as a result of what may be called secondary probability, but if there is true periodicity there must be a systematic departure of the grouping of events from the purely fortuitous arrangement. The criterion shows whether there is any tendency to a systematic concentration of the events in groups to a greater extent than would occur fortuitously, whether this is the result of chance or true periodicity. If, then, the application of the criterion shows there is no such group concentration there can apparently be no true periodicity in the phenomena. It may be noted, however, that the criterion fails in case there is a true periodicity with uniformly two events per group. In that case the value of the criterion is unity, although the periodicity may be perfect.