551.577.35 (54) : 551.509.33

A statistical study of the persistency of rain days during the monsoon season at Poona

V. K. RAMABHADRAN

Meteorological Office, Poona

(*Received 10 December 1952*)

ABSTRACT. A 60-year record of June-September precipitation data for Poona has been used to determine the frequency of rainspells of various lengths. It has been found that the frequencies conform closely to a logarithmic series $\Sigma(ax^r/r)$. A method of predicting rain-days based on past rain, making use of the observed frequencies has been illustrated. Also, the use of the constant x of the logarithmic series in persistency forecast is discussed. Finally the skill score of a persistency forecast of a rain-day following a rain period of 1, 2,...days duration has also been determined. Persistency itself may perhaps be construed as a meteorological factor and it may be possible to improve the accuracy of rain forecasts by applying the persistency theory in addition to the existing synoptic techniques.

1. Introduction

Meteorologists have found that weather sequences at any location are not entirely unrelated to one another and that certain types of weather sequences have a tendency to persist. Emphasis is placed in recent years on the objective methods of forecasting and it is natural, therefore, to look upon the theory of persistency of weather as providing an objective basis. A good deal of work on these lines has been done since 1916 when Newnham (1916) considered the persistence of Wet and Dry weather. Later, in 1929 Gold advanced a method of examining persistence of one type of weather and subsequently Cochran (1938) extended Gold's investigation to the case when the weather types have unequal probability. The work of Gold and Cochran was based on the distribution theory of runs applied to the combined data of wet and dry sequences. Jorgensen (1949) has investigated the persistency of rain and no-rain periods during the winter at San Francisco and defines a 'skill score' for persistency forecasts. Recently Williams (1952) has considered the sequence of wet and dry days in relation to the logarithmic series. The purpose of this paper is to present the results of the study of 60 years data (1891-1950) of daily rainfall at Poona during the monsoon season from June to September on the basis of the work of Jorgensen and Williams.

2. Procedure

The daily rainfall at Poona was used to tabulate the rainy days—a day receiving one cent or more being considered a 'rain-day' in this paper. From these data, the frequency of rainspells of various lengths was computed—

(*i*) for the four months June, July, August, September

(*ii*) for the season June-September and

(*iii*) for the progressive periods 1891-1900, 1891-1910......1891-1950.

These are given in Table 1 (a), 1 (b) and 1 (c) respectively. It may be emphasized that the spells discussed here represent continuity of rain days and not continuous rain during the whole period. Further, while assigning spells to different months in Table 1 (a), if a spell started in one month and ended in the next month, it was assumed that the spell had 'broken' at the end of the first month and that a different spell had started at the beginning of next month. In Table 1 (b) are also given the cumulative totals of the frequencies (from the longest to the shortest spell) from which the proportional frequencies of rain spells of various lengths followed by one or more days, two or more days etc were determined. It is seen from Table 1 (a) that the shorter sequences are more frequent in the months of June and September and that sequences

TABLE 1(a)

Observed frequency of rain period of various lengths during the four monsoon months

Length of rain period in days	Months			
	June	July	August	September
1	92	55	72	115
2	65	42	54	59
3	32	20	24	34
4	16	17	27	22
5	14	20	14	18
6	12	13	12	10
7	4	12	14	4
8	6	8	5	9
9	2	4	4	
10	7	9	12	1
11	1	7	4	2
12	1	7	7	
13		3	4	
14		3	2	
15	1	3		
16		2	1	1
17		2	5	
18		1		
19		4	1	
20		1		
21		1	2	
22				
23		2		
24		1		
25		1		
26				
27		2		

TABLE 1(b)

Observed frequency of rain periods, frequency calculated from formulae and the frequency cumulated from the longest to the shortest

Length of rain period in days	Observed frequency	Frequency from logarithmic series	Frequency from Cochran's formula	Cumulative total of observed frequency	Cumulative total of logarithmic frequency
1	312	360·0	848	950	945·4
2	196	163·7	456	638	585·4
3	104	99·2	245	442	421·7
4	74	67·7	132	338	322·5
5	54	49·3	71	264	254·8
6	39	37·3	38	210	205·5
7	32	29·1	21	171	168·2
8	25	23·1	11	139	139·1
9	8	18·7	5·9	114	116·0
10	19	15·3	3·2	106	97·3
11	13	12·6	1·7	87	82·0
12	12	10·5	<1	74	69·4
13	7	8·84		62	58·9
14	3	7·47		55	50·1
15	5	6·34		52	42·6
16	9	5·41		47	36·3
17	6	4·62		38	30·9
18	3	3·98		32	26·2
19	4	3·42		29	22·3
20	6	2·96		25	18·8
21	6	2·56		19	15·9
22	1	2·22		13	13·3
23	3	1·93		12	11·1
24	1	1·68		9	9·17
25	1	1·47		8	7·49
26		1·29		7	6·02
27	4	1·12		7	4·73
28		0·99		3	3·61
29	1	0·87		3	2·62
30		0·76		2	1·75
31		0·67		2	0·99
>31	2	0·32		2	0·32

of longer duration (9 and more days) are more frequent in the rainiest months of July and August as should be expected.

During the 240 months (June-September) of the 60-year record there were 3969 rain-days out of a total of 7320 days giving the probability of ·54 for a day of rain on chance. The probability is respectively .40, .73, .63 and .40 for the months June, July, August and September. Also the probability* based on chance of rain on two consecutive days is $(.54)^2$, for three days is $(.54)^3$ etc, assuming that the occurrence of rain on any day is independent of its occurrence on the previous day or days.

The data showed that the longer sequences (*i.e.*, exceeding 10 days) generally start in middle of July and extend to middle of August depending on their length. There were two sequences lasting longer than a month the first in 1896 of 33 days duration from 20 July to 21 August and the second in 1914 lasting 40 days from 5 July to 13 August.

3. Frequency of rain periods of various lengths represented by a logarithmic series

Cochran (1938) has studied the problem of runs for two weather types and gave the formula

$$f_{r,m} = N \left\{ p^r q \left\{ 2 + q\,(m - r - 1) \right\} \right\}$$

for determining the frequency of rain period of various lengths expected on chance. In this equation p represents the probability of rain (*i.e.*, .54), q is $1-p$ (=.46) and $f_{r,m}$ is the frequency of runs of length r out of m. Also m is the number of days in a season (*i.e.*, 122) and N the number of years of data (*i.e.*, 60). Frequencies calculated from this equation are given in Table 1 (b).

TABLE 1(c)

Observed frequency of rain period of various lengths during the six progressive periods

Length of rain period in days	Progressive periods					
	1891-1900	1891-1910	1891-1920	1891-1930	1891-1940	1891-1950
1	62	107	147	200	260	312
2	45	80	105	137	163	196
3	16	35	60	71	89	104
4	19	31	43	60	63	74
5	8	18	28	39	48	54
6	3	7	13	20	31	39
7	3	8	18	21	25	32
8	5	9	15	18	23	25
9	2	4	6	6	8	8
10	3	7	8	10	14	19
11	1	3	4	9	12	13
12	3	4	6	8	10	12
13		1	1	3	4	7
14	1	2	2	3	3	3
15	1	2	4	4	4	5
16		1	3	6	7	9
17	1	3	4	5	6	6
18		2	2	3	3	3
19	2	2	2	2	3	4
20		1	2	2	5	6
21			1	2	4	6
22				1	1	1
23	2					3
24			1	1	1	1
25					1	1
26						
27				2	3	4
28						
29						1
30						
31						
>31	1		2			2

* The probability of two consecutive days of rain can be got from the total number of two consecutive rain-days. Thus if $S_1, S_2, \ldots\ldots S_n$ are the frequencies of rain spells of 1, 2, ……. n days the probability of two consecutive rain days $= \left\{ S_2 + 2S_3 + \ldots\ldots + (n-1)S_n \right\} / (T-1)$ where T is the total number of days. In the present case $p = 3019/7319 = ·41$. But this definition was not adopted since the elements entering in this probability are not independent. Hence the probability $(·54)^2$ was used instead

An alternative approach is based on the assumption that weather persists, *i.e.*, the longer a spell of a particular type of weather the more likely it is to last another day. Williams (1952) applies the logarithmic series which is suited to describe a series with much characteristics. Let S_1, S_2, ...S_n represent the frequency of rain spell lasting over 1, 2, ...n days respectively. If S is the total frequency and T the total number of rainy days then

we have setting $S_r = \frac{\alpha x^r}{r}$

that $S = \Sigma S_r = -\alpha \log_e (1-x)$ ....(1)

and $T = \Sigma r S_r = \frac{\alpha x}{1-x}$ ......(2)

The quantity x is less than unity and we see from these equations that

$$S = \alpha \log_e \left(1 + \frac{T}{\alpha}\right) \text{ and } x = \frac{T}{T+\alpha}$$

giving α and x in terms of the known quantities S and T. Methods of solving these equations are given by Williams (1947). Yule (1944) gives

$$\alpha = \frac{T_1^2}{T_2 - T_1}$$

Where $T_1 = \Sigma r S_r$ the first moment and $T_2 = \Sigma r^2 S_r$ the second moment of the logarithmic series.

The frequency of rain spells was calculated from the logarithmic series (α, x) fitted to the data using the table given by Williams (1947). These are included in Table 1(b) along with the cumulative values. It may be seen that the frequencies calculated from Cochran's formula (based on chance distribution of rain days) are far from the observed frequencies. On the other hand, the frequencies obtained by fitting the logarithmic series (based on the persistency theory of rain) compare favourably with the observed, particularly the frequencies of rain of 3 and more days duration. This strengthens the general belief that the phenomena of rain persists. The graph of the logarithmic series is given in Fig. 1.

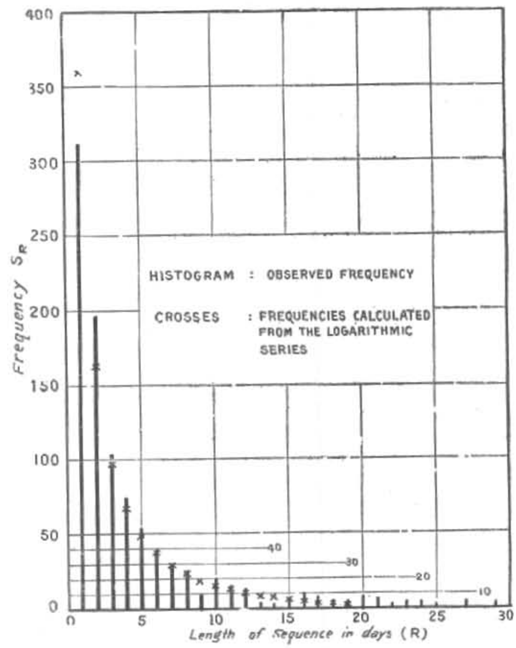

Fig. 1. Logarithmic series fitted to observed frequency

Also in order to see how the constants x and α change in the six progressive periods and also during the course of the monsoon, the logarithmic series was fitted to data of Tables 1(a) and 1 (c) and the results are given in Table 2 (a) and 2(b) along with values of T, S, p, x and α. In Table 2 (b) the values for the different decades are given in brackets. It is observed that the average number of rain-days per sequence and the constants x and α increase* with the length of the period considered. The constant α is in fact additive whereas the constant x increases because it theoretically depends on T/S

* There was change in the raingauge site in the years 1879, 1905, 1927 and 1930. It is possible that the increase in number of rainy days per sequence is partly due to the change in the site

(the average number of rain-days) and is given by

$$\frac{T}{S} = \frac{x}{(1-x)\ (-\log_e \overline{1-x})}\ .$$

solving equations (1) and (2).

4. The percentage frequency of rain periods followed by the same type of weather

The logarithmic series possesses as shown above theoretical properties of persistency. But for our work we need a quantitative definition and it is as follows : Persistency is the percentage frequency in excess of chance that a period of R days of a particular type of weather is followed by M or more successive days of the same weather beginning L days later.

From the accumulated values given in Table 1(b) (row 5), the percentage of times a period of rain is followed by rain-days can be calculated. For instance out of 950 cases of a rain period of at least 1-day duration, 638 cases or 67 per cent were of at least 2-days duration, 442 cases or 47 per cent were of at least 3-days duration etc. Similarly of the 638 cases lasting 2 days or more 442 or 69 per cent lasted at least 3 days, etc. Corresponding percentage frequencies obtained in this manner are given in Table 3 upto $R=25$ and $M=8$. Further frequencies being small were not considered. The percentage frequencies (*i.e.*, probability) of rain period continuing for 1, 2,........ days on a chance basis are also given in the table.

TABLE 2(a)

Value of T, S, p, x and α for the 4 months June to September

Months	Number of rainy days (T)	Number of sequences (S)	Average per sequence	p	x	α
June	726	253	2·87	·403	·841	137·7
July	1356	240	5·65	·729	·941	84·7
August	1166	264	4·42	·627	·916	106·4
September	721	275	2·62	·401	·817	162·0

TABLE 2(b)

Value of T, S, p, x and α for the six progressive periods

Progressive period	Number of rainy days (T)	Number of sequences (S)	Average number of rainy days per sequences	Probability of a day of rain (p)	x	α
1891-1900	653	178	3·67	·535	·890	80·8
1891-1910	1254(601)	330(152)	3·80(3·95)	·514(·493)	·896(·902)	146·0(65·5)
1891-1920	1918(664)	479(149)	4·00(4·46)	·524(·544)	·903(·917)	204·8(59·7)
1891-1930	2597(679)	637(158)	4·08(4·30)	·532(·557)	·906(·914)	269·8(64·8)
1891-1940	3288(691)	795(158)	4·14(4·37)	·539(·566)	·908(·915)	333·3(64·0)
1891-1950	3969(681)	950(155)	4·18(4·39)	·542(·558)	·909(·916)	395·8(62·6)

(Values for the different decades are given in brackets)

Table 3 shows that the probability that the phenomena of rain at Poona will recur is greater than would be expected on chance. For instance, if what has taken place the day before be ignored, the probability of rain which is ·54 rises abruptly to ·67 if we know it has rained. The probability continues to rise gradually till 9 successive days of rain but behaves somewhat irregularly afterwards. The probability of a rain period continuing for at least one more day is high following a rain period of 9, 14, 18 and 22 days duration with a value of 90-94 per cent as compared to 54 per cent expected on chance. A closer examination of Table 3 shows that the probability of rain in general has four maxima corresponding to R=9, 13-14, 17-18 and 22.

5. Forecasting of rainy days based on persistency

Persistency as defined above measures the tendency for weather to continue in excess of chance and thus quantitatively is equal to the probability of continuance obtained from the data minus the random probability. This difference can be worked out for various values of R and M from Table 3. Fig. 2 represents graphically the persistency calculated for various values of R with $M= L=1$.

From the figure it may be seen that persistency has maxima following a rain period of 9, 14, 18 and 22 with a value of 36-40 per cent. Practical interpretation of this result is that if 'today' is the 9th day of a rain period the total probability that it will rain 'tomorrow' is 93 which on subtraction of 54 per cent random probability for a rain-day, gives 39 per cent for the value of persistency. Similar results for other values of M can be derived directly from Table 3 using the percentages given in the right hand margin. In practice it would be necessary to forecast rain for the second or third day after it has rained on a particular day and we have no knowledge of rain on intervening days. We have then to take L=2, 3 etc. From a further examination of records the persistency curve M=1, L=2 in Fig. 3, was drawn. This curve does not show any prominent peaks like the curve M=1, L=1. It rises upto R=8 and is steady afterwards at a value 26-27. Interpretation of this is that if 'today' is the 8th day of a rain period at Poona, the probability that it will rain 'the day after tomorrow' without any knowledge of what might happen 'tomorrow', is 81 per cent as obtained from the data. Reducing this figure by 54 per cent which is the probability of rain due to chance we get 27 per cent for the persistency.

6. Skill score of a persistency forecast

A measure of success of a forecast is fundamental to any technique of objective forecasting. Such a measure is called a SKILL SCORE and is given by

$$\lambda = \frac{C-E}{N-E}, \quad \text{where}$$

C=the number of correct forecasts

E=the number of forecasts likely to be correct due to chance

N=the total number of forecasts.

Obviously such a measure takes into account the residual skill of a forecaster and does not give him credit for the number of forecasts which would be correct with a purely random distribution of them throughout the season. Based on this measure a persistency forecast will have a worthwhile score. Also C for such forecasts is given by the percentage frequency in Fig. 2 for a situation considered and E by the percentages on chance given in the right hand margin of the figure. Thus $C-E$ would be the actual persistency when N is taken to be 100. Considering the most common forecast, *viz.*, M=1, we have

$$\lambda=\frac{\text{Persistency } P \text{ expressed as percentage}}{1\text{—Probability } (\cdot 54) \text{ of a rainy day}}$$

$$=\frac{P}{0\cdot 46} \text{ or } 2\cdot 17P$$

These values of λ are given in the right hand scale of Fig. 2 and it may be seen that for M=1, the skill of persistency 'rain forecasts' varies from ·22 to ·87 depending on R the length of the preceding rain period.

7. The use of the logarithmic series in the evaluation of persistency

In the practical application of the persistency theory to forecasting day to day rainfall, it may be worthwhile to calculate instead of the observed percentage frequency (given

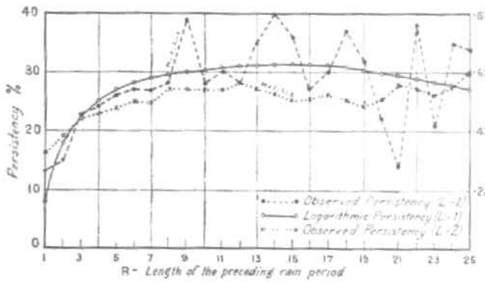

Fig. 2. Graph showing persistency of rain following a rain period of R days and corresponding skill score

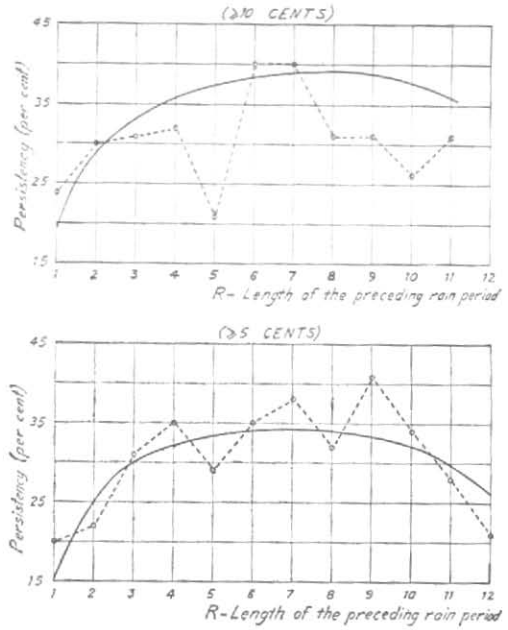

Fig. 3. Observed and logarithmic persistency of days with rain at least 5 cents and at least 10 cents

TABLE 3

Percentage frequency of occurrence of a rain period of M or more days following a rain period of R days

M—Number of rain days following a rain period of R days	R—Length of the preceding rain period in days																									Probability on chance
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	
8	12 (12)	17 (17)	20 (19)	22 (22)	23 (23)	26 (24)	30 (25)	33 (26)	33 (27)	30 (27)	32 (27)	32 (27)	31 (27)	24 (27)	23 (26)											<1%
7	15 (15)	18 (20)	24 (23)	26 (25)	28 (27)	29 (29)	32 (30)	36 (31)	41 (31)	36 (32)	35 (32)	38 (32)	40 (32)	35 (32)	25 (31)	26 (31)										1%
6	18 (18)	22 (24)	26 (27)	32 (30)	33 (32)	35 (34)	36 (35)	39 (36)	46 (37)	44 (37)	44 (38)	43 (38)	47 (38)	45 (38)	37 (37)	28 (37)	32 (36)									3%
5	22 (22)	27 (29)	32 (33)	34 (36)	40 (38)	41 (40)	43 (41)	44 (42)	48 (43)	49 (44)	54 (44)	51 (44)	52 (44)	53 (45)	48 (44)	40 (44)	34 (43)	37 (42)								5%
4	28 (27)	33 (35)	39 (40)	41 (43)	43 (45)	50 (47)	51 (49)	53 (50)	54 (51)	52 (51)	60 (52)	63 (52)	61 (52)	58 (52)	56 (52)	53 (52)	50 (51)	41 (51)	41 (50)	36 (49)	42 (47)	54 (45)				9%
3	36 (34)	41 (44)	48 (49)	51 (52)	53 (55)	54 (56)	62 (58)	63 (59)	65 (60)	58 (61)	63 (61)	70 (61)	76 (62)	69 (62)	62 (61)	62 (61)	65 (61)	59 (61)	45 (60)	48 (59)	47 (58)	62 (58)	58 (54)			16%
2	47 (45)	53 (55)	60 (60)	62 (64)	65 (66)	66 (68)	67 (69)	76 (70)	76 (71)	70 (71)	71 (72)	74 (72)	84 (72)	85 (72)	73 (73)	68 (72)	76 (72)	78 (72)	66 (71)	52 (71)	63 (70)	69 (69)	67 (67)	78 (66)		29%
1	67 (62)	69 (72)	77 (76)	78 (79)	80 (81)	81 (82)	81 (83)	82 (83)	93 (84)	82 (84)	85 (85)	84 (85)	89 (85)	94 (85)	90 (85)	81 (85)	84 (85)	91 (85)	86 (84)	76 (84)	68 (84)	92 (83)	75 (82)	89 (82)	88 (80)	54%

Bracketted figures correspond to logarithmic series in Table 3), percentages based on the logarithmic series fitted to the observed frequencies. The theoretical percentages are determined by the value of x and can be directly obtained from the cumulative totals of logarithmic frequencies given in Table 1 (b). The theoretical percentage frequencies thus calculated correspond to $x = \cdot 909$ and are given within brackets in Table 3. It is seen from Table 3 that the theoretical persistency rises to a value of 31 (*i.e.* 85-54) after 11 days of rain and continues till the 18th day and later begins to decline gradually. The continuous line in Fig. 2 gives the persistency and skill score based on theoretical frequencies for $M=L=1$. The skill score attains a value of ·67 on the 11th day of rain.

Thus, while the observed persistency rises gradually upto $R=8$ and exhibits four maxima later on, the logarithmic (*i.e.*, theoretical) persistency shows a rise upto $R=11$, steadiness upto $R=18$ and declines later.

In this connection it was studied how far the definition of rain-day adopted in this paper accounts for the difference between these curves. Criterion of 5 and 10 cents for a rain-day was considered. The persistency curves corresponding to the data and the logarithmic series are shown in Fig. 3. The 5 cents curve has peaks at 4, 7 and 9 but not the 10 cents curve, but both the curves deviate from the smooth logarithmic curves like the 1 cent curve. The significance of the logarithmic frequencies lies in the fact that x is fairly constant over the 60 years having increased only by ·02 from ·89, its value for the 1st decade 1891-1900. The constant x may aptly be called an 'Index of Persistency' since the percentage frequencies based on x increase with x as illustrated below—

x	p_{11}	p_{21}	p_{31}
0·50	28	36	40
0·91	62	72	76

(where p_{R_1} is the probabilitiy of a day of rain after R days of rain). From Table 2(a) it is seen that the variation in x values is much less compared to the variation in the probably p. Based on x it is possible to prepare charts like Fig. 2 to indicate the persistency during the different months. But, it appears that for Poona the value of x can be taken to be ·909, the season's value itself. As mentioned previously the constant x depends on T/S and hence should be the same for stations having the same number of rain-days per sequence.

8. Acknowledgement

The author is grateful to all those who kindly helped him in the collection, analysis and interpretation of data.

REFERENCES

Cochran, W.G. (1938). *Quart. J. R. met. Soc.*, **64**, pp. 631-634.

Gold, E. (1929). *Quart. J. R. met. Soc.*, **55**, pp. 307-309.

Jorgensen, D. L. (1949). *Mon. Weath. Rev.*, **77**, 11.

Newnham, E.V. (1916). *Quart. J. R. met. Soc.*, **42**, 179, pp. 153-162.

Williams, C.B. (1947). *J. Ecology*, **34**, p. 253.

Williams, C.B. (1952). *Quart. J. R. met. Soc.*, **78**, pp. 91-96.

Yule, G. U. (1944). *The statistical study of literary vocabulary*, Cambridge University Press.