Indian J. Met. Geophys. (1971), 22, 67–74

551.577.31(541.2)

Fitting of a Markov chain model for daily rainfall data at Calcutta

A. N. BASU*

River Research Institute, Calcutta

(Received 14 August 1968)

ABSTRACT. A Markov chain probability model has been fitted to the daily rainfall data recorded at Calcutta. The 'wet spell' and 'weather cycles' are found to obey geometric distribution. The distribution of the number of rainy days per week has been calculated and compared with the actual data.

1. Introduction

A number of authors have analysed the distribution of rainfall and occurrence of dry and wet weather spells in different ways. Gabriel and Neumann (1957) have shown that dry and wet spells follow a geometric distribution. The same authors (1962) have found daily rainfall data to fit a Markov chain model which presents the probable "spell distribution" and other properties of rainfall occurrence patterns. The results of a statistical study of 62 years daily rainfall data recorded at Calcutta during the monsoon season (June to September) on the basis of the work of Gabriel and Neumann are presented in the present communication.

2. Geometric distribution and Markov Chain Model — Gabriel and Neumann's treatment

Let x be a positive, integral valued random variable. Then x is said to obey a geometric distribution if,

$$P_r\{x=K\} = q_1\, p_1^{K-1}, \quad K = 1, 2, 3, \ldots \tag{1}$$

where p_1 and q_1 are positive numbers such that $p_1+q_1 = 1$.

The definition of Markov chain model in terms of rainfall occurrence and some properties of the model given by Gabriel and Neumann are stated here for convenience.

It is assumed that the probability of rainfall on any day depends only on whether the previous day was wet or dry. If the event (wet or dry) of the previous day is given, the probability of rainfall is assumed to be independent of events of further preceding days. Such a probability model is known as Markov Chain Model, whose parameters are the two conditional probabilities given by—

$$p_1 = P_r\,(\text{Wet day/previous day wet}) \tag{2}$$

$$p_0 = P_r\,(\text{Wet day/previous day dry}) \tag{3}$$

The probabilities of rainfall i days after a 'wet' or a 'dry' day are —

$$P + (1-p)\,d^i \tag{4}$$

or $$P + pd^i \tag{5}$$

respectively,

where, $$d = p_1 - p_0 \tag{6}$$

and $$P = p_0/(1-d) \tag{7}$$

P being the probability (absolute) of a day being 'wet'.

A wet spell of length m is defined as a sequence of m wet days preceded and followed by dry days. A 'dry spell' is defined in the same way. A 'weather cycle' is defined as the combination of a wet spell with the immediate successive dry spell or a dry spell with the immediate successive wet spell. The first combination is called the 'wet-dry cycle' and the later 'dry-wet cycle'.

The probability of a wet spell of length K is,

$$(1-p_1)\,p_1^{K-1} \tag{8}$$

and that of a dry spell of length m is,

$$p_0\,(1-p_0)^{m-1} \tag{9}$$

According to the model described above, the lengths of wet and dry spells would be independent.

*Formerly at Agricultural Meteorology Division, Poona

So the distribution of the length of the weather cycle is,

$$f(n, p_0, p_1) = p_0(1-p_1) \frac{(1-p_0)^{n-1} - p_1^{n-1}}{1 - p_0 - p_1} \quad (10)$$

where, n = length of the cycle $= 2, 3, \ldots$

The probability of exactly s wet days among n days following a wet day is,

$$P_r\{s/n, 1\} = p_1^s (1-p_0)^{n-s} \times$$

$$\times \sum_{c=1}^{c_1} \binom{s}{a}\binom{n-s-1}{b-1}\left(\frac{q_1}{q_0}\right)^b \left(\frac{p_0}{p_1}\right)^a \quad (11)$$

where,

$$q_1 = 1-p_1, \quad q_0 = 1-p_0$$

$$c_1 = n + \tfrac{1}{2} - |2s - n + \tfrac{1}{2}|, \quad \text{if } s < n$$

$$= 0 \text{ (Summation contains this term only)},$$

$$\text{if } s = n \quad (12)$$

and a and b are the least integers not smaller than $\frac{1}{2}(c-1)$ and $\frac{1}{2}c$ respectively. Similarly, the probability of exactly s wet days among n days following a dry day is,

$$P_r\{s/n, 0\} = p_1^s (1-p_0)^{n-s} \times$$

$$\times \sum_{c_1=1}^{c_0} \binom{s-1}{b-1}\binom{n-s}{a}\left(\frac{q_1}{q_0}\right)^a \left(\frac{p_0}{p_1}\right)^b \quad (13)$$

where,

$$c_0 = n + \tfrac{1}{2} - |2s - n - \tfrac{1}{2}|, \quad \text{if } s < n,$$

$$= 0 \text{ (Summation contains this term only)},$$

$$\text{if } s = n \quad (14)$$

and a and b defined as above. The probability of s wet days among n days is given by—

$$P_r\{s/n\} = P \cdot P_r\{s/n, 1\} + (1-P)\, P_r\{s/n, 0\} \quad (15)$$

For large n, the distribution of the number of wet days tends to normality with mean and variance

$$\left.\begin{aligned} E(s) &= nP \\ \text{Var}(s) &= nP(1-P)\frac{1+d}{1-d} \end{aligned}\right\} \quad (16)$$

TABLE 1

Estimates of conditional probabilities of rainfall occurrence for different months

Preceding day		Actual day Wet	Actual day Dry	Total	Estimated Probability p_1	Estimated Probability p_0
Jun	Wet	438	335	773	·56662	
	Dry	350	737	1087		·32199
Jul	Wet	640	394	1034	·61896	
	Dry	390	498	888		·43919
Aug	Wet	682	401	1083	·62973	
	Dry	403	436	839		·48033
Sep	Wet	496	360	856	·57944	
	Dry	350	654	1004		·34861

3. Data and computations

The daily rainfall data for the period June to September recorded at Calcutta (Alipore) for the years 1902 to 1964 (excluding 1947 as the data of all the days of the period June to September for the year are not available) have been utilised for the study. A day (0830 to 0830 IST of following day) receiving at least 10 cents of precipitation has been considered as a wet day, otherwise as a dry day. To ascertain whether a spell belongs to a particular month or not, the following conventions have been followed.

(a) A wet spell is included in a month if any day of this particular spell falls within that month, no matter whether the spell ends or does not end in that month.

(b) A dry spell is included in a month if the immediately following wet spell is included in that particular month as mentioned in (a).

(c) A wet-dry cycle (a cycle beginning with wet spell and ending with dry spell) or dry-wet cycle (a cycle beginning with dry and ending with wet spell) has been included in a month if any part of the wet spell falls in that month.

The conditional probabilities p_1 and p_0 for each month have been estimated by the corresponding relative frequencies as the relative frequencies are the Maximum Likelihood Estimate of the Probabilities (Anderson& Goodman, 1957). The estimated values of p_1 and p_0 for the different months are given in Table 1.

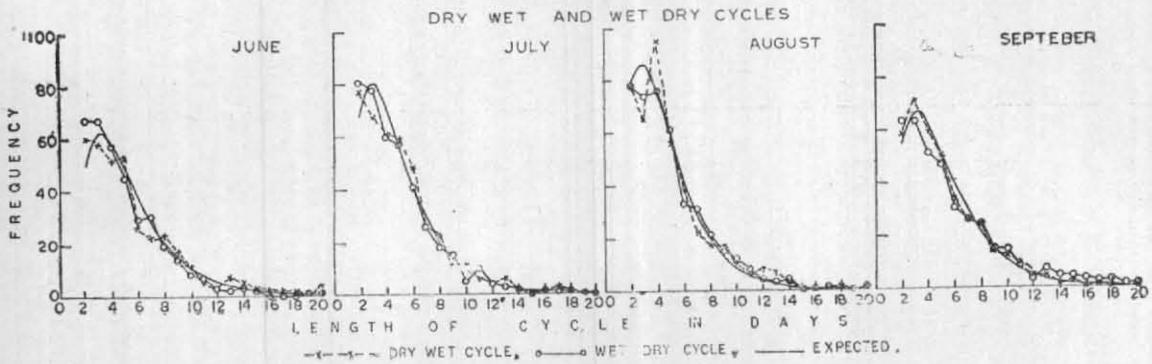

Fig. 1 Fig. 2 Fig. 3 Fig. 4

Figs. 1-4. Dry-wet and wet-dry Cycles

TABLE 2

Expected and observed frequency of dry and wet spells and their tests of significance

Length of the spell (day)	Wet spell								Dry spell							
	June		July		August		September		June		July		August		September	
	Obs	Exp	Obs	Exp	Obs	Exp	Obs	Exp	Obs	Exp	Obs	Exp	Obs	Exp	Obs	Exp
1	182	154	154	154	152	157	161	155	127	114	188	177	206	203	141	128
2	68	87	82	95	87	99	80	90	79	78	109	99	100	106	87	84
3	38	49	65	59	66	62	50	52	44	53	26	56	56	55	45	54
4	30	28	42	36	43	39	29	30	20	36	40	31	30	28	46	36
5	13	16	21	23	26	25	19	18	20	24	13	18	13	15	12	23
6	9	9	13	14	18	15	9	10	14	16	7	10	7	8	15	15
7	6	5	8	9	8	10	12	6	11	11	6	5	7	4	6	10
8	5	3	4	5	9	6	1	3	11	8	5	3	2	2	3	6
9	2	2	5	3	6	4	5	2	6	5	6	2	1	1	3	
10	0	1	3	2	3	2	1	1	6	3	0	1	1	1	4	
11	1	1	3	1	2	2	1	1	5	2	0	1	0	0	1	
12	0	0	3	1	2	1	0	0	3	2	0	0	0	0	2	
13	0	0	0	1	0	1	0	0	4	1	0	0	0	0	2	
14	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	1
15	0	0	0	0	1	0	0	0	1	1	0	0	0	0	1	0
16	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
17	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
18	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
Total	355	355	403	403	423	423	368	368	355	355	403	403	423	423	368	368
χ^2	13·1865		5·6686		6·4221		5·3784		23·6166		22·0991		2·0617		12·7883	
D.F.	6		6		7		5		6		6		5		6	
P	·02–·05		·30–·50		·30–·50		·30–·50		—		·01–·001		·80–·90		·02–·05	

TABLE 3

Expected and observed frequency of dry-wet and wet-dry cycles and their tests of significance

Length of the cycle	June Wet-dry (obs)	June Expected	June Dry-wet (obs)	July Wet-dry (obs)	July Expected	July Dry-wet (obs)	August Wet-dry (obs)	August Expected	August Dry-wet (obs)	September Wet-dry (obs)	September Expected	September Dry-wet (obs)
2	67	49	60	81	67	77	78	75	78	63	54	58
3	67	62	58	78	80	67	74	86	65	63	66	71
4	57	58	51	60	70	61	76	75	95	51	61	59
5	45	48	53	57	56	59	61	58	56	47	51	50
6	29	38	26	41	41	48	33	42	38	30	39	33
7	30	28	22	26	29	26	31	29	22	26	29	26
8	19	21	23	18	20	22	21	20	17	24	21	24
9	14	15	17	15	14	11	15	13	17	14	15	13
10	8	11	12	5	9	11	12	9	10	14	10	12
11	7	8	5	10	6	6	8	6	6	8	7	9
12	3	5	3	4	4	4	4	4	8	3	5	4
13	2	4	7	3	3	6	3	2	7	7	3	4
14	3	3	5	2	2	1	4	2	2	4	2	1
15	2	2	3	0	1	0	0	1	1	4	2	2
16	0	1	2	1	1	1	0	1	0	3	1	0
17	0	1	2	1	0	2	1	0	0	2	1	1
18	0	1	2	1	0	1	1	0	1	3	1	1
19	1	0	0	0	0	0	0	0	0	1	0	0
20	1	0	4	0	0	0	1	0	0	1	0	0
Total	355	355	355	403	403	403	423	423	423	368	368	368
χ^2	12·0099		12·2764	5·0944		10·4386	7·6371		19·5752	19·4814		3·9271
D.F.	7		7	7		7	7		7	8		8
P	·10–·20		·05–·10	·50–·70		·10–·20	·30–·50		·001–·01	·01–·02		·80–·90

The expected frequencies of the lengths of the dry and wet spells have been calculated using the equations (8) and (9) respectively. The observed and expected frequencies together with chi-square test for goodness of fit are given in Table 2. The observed and theoretical distribution of the lengths of the wet and dry spells are shown for the four months in Figs. 5–8. Also, the expected frequencies of the different lengths of the cycles for different months have been calculated by equation (10).

It may be mentioned here that the expected frequency of the length of the wet-dry or of dry-wet cycle should be same. The observed frequency of the length of the wet-dry and dry-wet cycle with the expected frequency are given in Table 3 (Figs. 1-4) and also chi-square test for goodness of fit are shown in the same table.

Chi–square test has been applied for each month to see whether the proportions of wet days are

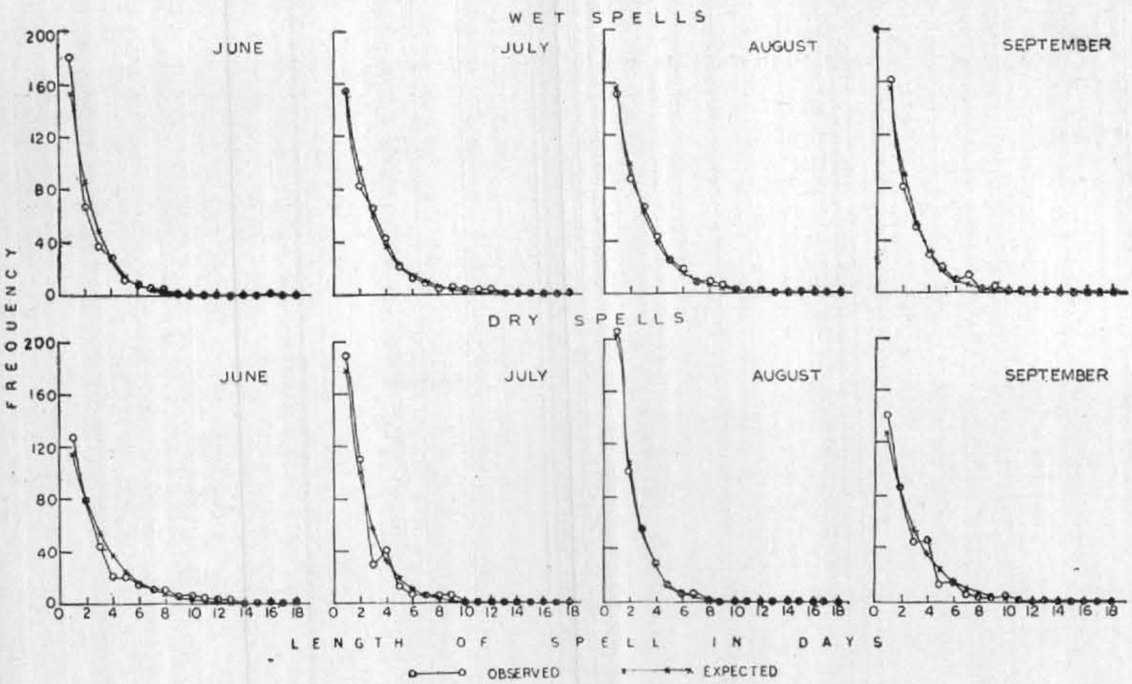

Fig. 5 Fig. 6 Fig. 7 Fig. 8

Figs. 5-8. Wet and dry spells

independent of the weather of two or more preceding days. The data have been entered in Table 4.

The distribution of the number of wet days in a week has also been computed for each month by Eq. (15). The actual and computed values are shown in Table 5.

The mean standard deviation and mean error of mean of the number of wet days have been computed for four months by Eq. (16) and shown in Table 6.

4. Results

The main results of the above analysis are given below—

(a) *Wet and dry spells* — It is seen from Table 2 that the length of the wet spell in each of the three months July, August and September fits the geometric distribution well. The goodness of fit is shown by the χ^2-test of significance. However in June, the expected frequencies come close to the observed values except the one-day and two-day spells where wide differences are noted.

But in the case of dry spell the monthly fit is not good except for August. In July, the expected frequency of three-day spell is very much higher than the corresponding observed frequency which is unusually low. It may be mentioned that dry spells of longer length (say$>$6) are more frequent in June in comparison with those of other monsoon months.

(b) *Wet-dry and dry-wet cycles* — Table 3 shows that the expected frequencies of the wet dry cycle calculated on the basis of geometric distribution agree well with those of the observed in all the months except September, as is shown by the χ^2- test of significance. In September the frequencies of the cycles of longer lengths (say $>$10) are more in comparison to those in other months. In case of dry-wet cycle, the observed distribution compares well with the theoretical distribution in all these months except for August, where the observed frequencies of the length of three and four-day cycles are somewhat unusual in comparison to those observed in the other months.

Gabriel-Neumann (1957) have shown from theoretical consideration that the model length of the dry-wet or wet-dry cycle cannot be less than three days. Here, the model length of the theoretical distribution is three days in all the four months. The model length of the observed distributions of September (both wet-dry and dry-wet) and June (wet-dry only) coincide with those of the respective theoretical distribution. In other cases the observed model lengths are 2, 2, 4 and 2 days for respective cycles noted in the brackets in June (dry-wet), July (wet-dry and dry-wet), August (dry-wet) and August (wet-dry) respectively.

TABLE 4

Conditional relative frequencies for different months

	Preceding days 3rd	2nd	1st	Actual day Wet	Dry	Total	RFW	GFT	Actual day Wet	Dry	Total	RFW	GFT
				June					**July**				
(a)	T	T	T	788	1,072	1,860	·42366		10·30	892	1922	·53590	
(b)	T	T	Wet	438	335	773	·56662		640	394	1034	·61896	
			Dry	350	737	1087	·32199		390	498	888	·43919	
(c)	T	Wet	Wet	269	155	424	·63443	—	399	240	639	·62441	·10–·20
		Dry	Wet	169	180	349	·48424		241	154	395	·61013	
		Wet	Dry	127	210	337	·37685		184	204	388	·47423	
		Dry	Dry	223	527	750	·29733		206	294	500	·41200	
(d)	Wet	Wet	Wet	165	93	258	·63953	—	243	156	399	·60902	·001–.01
	Dry	Wet	Wet	104	62	166	·62651		156	84	240	·65000	
	Wet	Dry	Wet	66	60	126	·52381		106	77	183	·57923	
	Dry	Dry	Wet	103	120	223	·46188		135	77	212	·63679	
	Wet	Wet	Dry	62	93	155	·40000		108	127	235	·45957	
	Dry	Wet	Dry	65	117	182	·35714		76	77	153	·49673	
	Wet	Dry	Dry	77	130	207	·37198		102	102	204	·50000	
	Dry	Dry	Dry	146	397	543	·26888		104	192	296	·35135	
				August					**September**				
(a)	T	T	T	1085	837	1922	·56452		846	1014	1860	·45484	
(b)	T	T	Wet	682	401	1083	·62973		496	360	856	·57944	
			Dry	403	436	839	·48033		350	654	1004	·34861	
(c)	T	Wet	Wet	434	248	682	·63636	·80–·90	300	206	506	·59289	·30–·50
		Dry	Wet	248	153	401	·61845		196	154	350	·56000	
		Wet	Dry	193	209	402	·48010		131	229	360	·36389	
		Dry	Dry	210	227	437	·48055		219	425	644	·34006	
(d)	Wet	Wet	Wet	271	166	437	·62014	·30–·50	179	128	307	·58306	·10–·20
	Dry	Wet	Wet	163	82	245	·66531		121	78	199	·60804	
	Wet	Dry	Wet	115	80	195	·58974		75	55	130	·57692	
	Dry	Dry	Wet	133	73	206	·64563		121	99	220	·55000	
	Wet	Wet	Dry	114	134	248	·45968		66	143	209	·31579	
	Dry	Wet	Dry	79	75	154	·51299		65	86	151	·43046	
	Wet	Dry	Dry	94	114	208	·45192		88	140	228	·38596	
	Dry	Dry	Dry	116	113	229	·50655		131	285	416	·31490	

T—Total GFT—Goodness of fit χ^2-test RFW—Relative frequency of wet days

TABLE 5

Distribution of the number of wet days in a week — Theoretical and Observed

No. of wet days	June			July			August			September		
	Theoretical distribution		Observed freq.	Theoretical distribution		Observed freq.	Theoretical distribution		Observed freq.	Theoretical distribution		Observed freq.
	Probability	Expected freq.		Probability	Expected freq.		Probability	Expected freq.		Probability	Expected freq.	
0	0·05570	14	27	0·01446	4	4	0·00858	2	2	0·04176	10	13
1	0·13750	34	39	0·06110	15	14	0·04410	11	8	0·11689	29	22
2	0·20842	52	43	0·14051	35	41	0·11796	29	34	0·19482	48	40
3	0·22737	56	38	0·21769	54	44	0·20634	51	50	0·23005	57	72
4	0·18764	47	39	0·24144	60	54	0·25215	63	54	0·20311	51	50
5	0·11697	29	43	0·19189	48	55	0·21524	53	60	0·13385	33	31
6	0·05228	13	16	0·10281	25	30	0·12039	30	32	0·06237	16	14
7	0·01412	3	3	0·03010	7	6	0·03524	9	8	0·01715	4	6
Total	1·00000	248	248	1·00000	248	248	1·00000	248	248	1·00000	248	248
χ^2		28·83294			5·69674			4·02867			8·01118	
D.F.		4			4			4			4	
P		—			·20–·30			·30–·50			·05–·10	

TABLE 6

Mean and Standard Deviation of the number of wet days for different months

	Mean	Standard deviation σ	$\sigma/\sqrt{n}$
June	12·83	3·48	·44
July	16·60	3·33	·42
August	17·51	3·21	·41
September	13·60	3·45	·44

(c) *Fitting of the model* — The data for June to September are shown in Table 4. The two conditional probabilities have been calculated and noted for each month. It has also been tested whether these probabilities (proportion of wet days) are independent of the weather of two or three preceding days. From this test, it is seen that Markov model fits well for the months of August and September as shown by the P-values of the χ^2-test. In July χ^2-test is not significant in case of second preceding day but it is significant in case of the third preceding day. However in case of June the χ^2 test is significant in both the cases.

The probabilities of wet day i and days after a wet or dry day may be computed from equations (4) and (5) for all the months. Both these probabilities converge to the absolute probability (*i.e.*, probability of wet day) $P = {\cdot}42759$, ·53545, ·56470 ·45323 in June, July, August and September respectively.

The actual and theoretical distributions, based on Markov's model, of the number of wet days in a week have been shown in Table 5. The χ^2-test of significance for each month (consisting of 4 weeks) shows that the theoretical distribution does not deviate from the observed distribution, except for the month of June.

5. Conclusion

A Markov chain model has been fitted to the rainfall data of Calcutta. Based on this model, a theoretical distribution has been fitted to the distribution of the number of wet days in a week. It has been found that the 'wet spell' and 'weather cycles' obey geometrical distribution.

Acknowledgement — The author is grateful to Shri M. Gangopadhyaya, Director of Agricultural Meteorology Division, Poona for his valuable suggestions and encouragement, and thankful to Sarvashri J. R. Banerjee and G. Rama Rao for their help in computational work. The author is also thankful to Shri A. C. Chatterjee for going through the paper.

REFERENCES

Anderson, T. W. and Goodman, L. A.	1957	*Ann. Math. Stat.* **28**, pp. 89-110.
Gabriel, K. R. and Neumann, J.	1957	*Quart. J.R. met. Soc.*, **83**, pp. 375-380.
	1962	*Ibid.*, **88**, pp. 90-95.