Indian J. Met. Hydrol. Geophys. (1975), **26**, 2, 221-226

551.506.3 : 551.513.1

Markov-dependent geometric models for weather spells and weather cycles — A study

N. SUNDARARAJ and S. RAMACHANDRA

Department of Statistics, University of Agricultural Sciences, Bangalore

(*Received 16 August 1973*)

ABSTRACT. The repetitive behaviour of seasonal weather has been studied earlier by many using a variety of mathematical and probability models. The present paper proposes to investigate the empirical validity of Markov-dependent geometric models for wet and dry spells, and weather cycles. The daily rainfall of 50 years' duration serves as the data-base for the study.

The necessary formulae based on the geometric probability models using Markovian probabilities for dry and wet days, are given for lengths of dry spells, wet spells and wet- dry weather cycles.

The data, as judged by χ^2-test, seems to fit better for wet-spells than for dry spells.

1. Introduction

The repetitive behaviour of seasonal weather has always fascinated meteorologists and statisticians alike : meteorologists seeking physical explanations for such phenomena and statisticians in exploring possibilities of model building to explain the observed phenomena. Such models serve the important function of providing an orderly basis and permit further use of the deductive power of mathematics to reach conclusions that could not be reached otherwise and may even provide clues to physical understanding of the complex phenomena. Among the earlier studies may be mentioned the work of Cochran (1938) who proposed a probability model, based on the 'theory of runs' to study the 'persistancy' behaviour of rainy days while others like Gabriel and Neumann (1957) suggested a geometric distribution as a suitable model for wet and dry weather spells : yet others like Gabriel and Neumann (1962), Ramabhadran (1954) and Basu (1971) considered empirical evidence to find out the suitability or otherwise of some of these propositions.

The present study is intended to seek empirical evidence to study the suitability or otherwise of a Markov-based geometric model for representing daily rainfall occurrence at Hebbal (a village near Bangalore), where a major research station under the University of Agricultural Sciences, Bangalore is located.

2. Method used

For the purposes of the present study, a **wet spell** of length w days is defined as a sequence of w wet days followed and preceded by dry days and a dry spell of d days as a sequence of d dry-days followed and preceded by wet days. An occurrence of a wet-spell and an adjacent dry-spell (or *vice versa*) together constitutes a weather cycle and their total length defining the length of the weather cycle itself.

The probabilities of obtaining these three events— a wet-spell of length w days, a dry-spell of length d days and a weather-cycle of length n days can— be constructed from the Markovian-dependent geometric models. The probability generating expression for wets-pell length (x), dry-spell length (y) and weather cycle length (z) are :

$$P_{(x=w)} = [\pi_{(w/w)}]^{w-1} [1-\pi_{(w/w)}] \tag{1}$$

$$P_{(y=d)} = [\pi_{(D/D)}]^{d-1} [1-\pi_{(D/D)}] \tag{2}$$

$$P_{(z=n)} = [\pi_{(D/w)}] [\pi_{(w/D)}] \times \left\{ \frac{[\pi_{(w/w)}]^{n-1} - [\pi_{(D/D)}]^{n-1}}{[\pi_{(w/w)}] - [\pi_{(D/D)}]} \right\} \tag{3}$$

Here,

$P_{(x=w)}$ = Probability of a wet-spell of length w days

$P_{(y=d)}$ = Probability of a dry-spell of length d days.

$P_{(z=n)}$ = Probability of weather cycle of length n days.

$\pi_{(w/w)}$ = The conditional probability that a day is a wet (rainy) day given that the immediate preceding day was a wet day.

$\pi_{(D/D)}$ = The conditional probability that a day is a dry day given that the immediate, preceding day was a dry day.

$$\pi_{(D/w)} = 1 - \pi_{(w/w)}$$

$$\pi_{(w/D)} = 1 - \pi_{(D/D)} \tag{4}$$

Expressions (1) through (4) are employed later to obtain the expected frequencies for wet spells, dry spells, wet-dry weather cycles of varied lengths for Hebbal data.

It is possible further to predict in advance the *a priori* probabilities of a rainy day (or a dry day) k days after a rainy (or a dry) day. These probabilities and their complementaries are given in expressions (5) through (8). These expressions are :

$$\pi_{k(w/w)} = \left[\frac{1-\pi_{(w/w)}}{2-\pi_{(w/w)}-\pi_{(D/D)}}\right] \times \times [\pi_{(w/w)} + \pi_{(D/D)} - 1]^k + + \left[\frac{1-\pi_{(D/D)}}{2-\pi_{(w/w)}-\pi_{(D/D)}}\right] \tag{5}$$

$$\pi_{k(D/D)} = \left[\frac{1-\pi_{(D/D)}}{2-\pi_{(w/w)}-\pi_{(D/D)}}\right] \times \times [\pi_{(w/w)} + \pi_{(D/D)} - 1]^k + + \left[\frac{1-\pi_{(w/w)}}{2-\pi_{(w/w)}-\pi_{(D/D)}}\right] \tag{6}$$

$$\pi_{k(D/w)} = 1 - \pi_{k(w/w)} \tag{7}$$

$$\pi_{k(w/D)} = 1 - \pi_{k(D/D)} \tag{8}$$

where,

$\pi_{k(w/w)}$ = the conditional probability that a day, k days after a wet day, would be a rainy day.

$\pi_{k(D/D)}$ = the conditional probability that a day, k days after a dry day, would be a dry day.

$\pi_{k(D/w)}$ = the conditional probability that a day, after a dry day would be a wet day.

$\pi_{k(w/D)}$ = the conditional probability that a day, k days after a wet day, would be a rainy day.

In the present study, these formulae are employed only for three specific values of k, *viz.*, $k=2$, 3 and 4 to test if the basic assumptions of Markovian independence are satisfied by the observed data for preceding days other than the immediate one of the actual day.

The data used in the present study is the daily rainfall records for the months, July, August and September* for fifty years covering a period from 1920 to 1969. The following definitions have been employed.

(1) A day receiving 0.05″ (or 5 cents) of rainfall between 08 30 A.M. to 08 30 A.M. of the following day is defined as a rainy day or a wet day. Otherwise the day is counted as a dry day.

(2) A wet spell (or a dry spell) which overlaps between two adjacent months is assigned to that month which shares longer than half its total length. In case of equality, it is arbitrarily assigned to the previous month.

(3) In case of weather cycles, the whole length of the cycle which overlaps between two adjacent months is assigned to that month which shares more than half its length. In case of equality, it is arbitrarily assigned to the previous month. The above rules for the assignments of spells and cycles to different months, we believe are likely to introduce little or no bias in the long-run, as a result of the random characteristics associated with the mid-points of the spells and cycle lengths.

The conditional probabilities $\pi_{(w/w)}$, $\pi_{(D/D)}$, $\pi_{(D/w)}$ and $\pi_{(w/D)}$ which serve as the basic elements in all the foregoing formulae are estimated separately for each month, July, August, September, utilizing the corresponding observed frequencies of these events based on 50 years of daily rainfall data for Hebbal region, Bangalore. These estimates along with the observed frequencies are reported in Table 1. For instance the conditional probability $\pi_{(w/w)}$ that a day would be a wet day given that the preceding day was a wet day is given by the ratio 262/(262+265) =0.4972. Similarly $\pi_{(D/D)}$ is estimated by the ratio 755/(268+755)= 0.7380 etc. Table 1 also reports the unconditional binomial probability of a wet (or dry) day regardless of the condition whether the preceding day was a wet day or a dry day. For instance, the probability of a wet day for July is given by the ratio 530/(530+1020)= 0.3419. From Table 1, it may be noted that the corresponding probabilities for July, August and September are close to each other in their values;

* The month of June which was originally included was later discarded since the actual estimate differed significantly from the rest.

TABLE 1

Relative frequencies for obtaining conditional probabilities

Month		Frequency (Actual Day)		Conditional probabilities	
		Wet	Dry	Wet	Dry
Jul (preceding day)	Wet	262	265	·4972	·5028
	Dry	268	755	·2620	·7380
	Total	530	1020	·3419	·6581*
Aug	Wet	273	279	·4946	·5054
	Dry	280	718	·2806	·7194
	Total	553	997	·3567	·6433*
Sep	Wet	259	235	·5243	·4757
	Dry	240	766	·2386	·7614
	Total	499	1001	·3326	·6674*
Jul to Sep	Wet	794	779	·5048	·4952
	Dry	788	2239	·2603	·7397
	Total	1582	3018	·3439	·6561*

* These are the unconditional binomial probabilities for wet and dry days respectively

as such pooled-estimates for each of the four probabilities $\pi_{(w/w)}$ etc, are also computed by pooling the corresponding frequencies for July, August and September. These pooled estimates are also reported in Table 1.

Incidentally it may be noted that the differences between the conditional probabilities and the corresponding unconditional binomial probabilities are very pronounced.

By employing these probabilities in (1) and (2) one can generate the expected relative frequencies for the lengths of wet spells and dry spells respectively and which, when multiplied by the appropriate sample sizes will result in the expected frequencies for lengths of wet and dry spells. These expected frequencies along with the observed ones are reported in Tables 2 (a) and 2 (b). From the these tables, the following observations may be made regarding wet and dry spells.

(*i*) *Wet-spells*

From Table 2 (a), one would note a general agreement between the observed and the predicted frequencies for wet-spells of varied lengths for each of the three months July, August and September, although the degree of agreement is better in the case of July and August than in the case of September. The observed χ^2 values in all the three cases are, however, non-significant at 5 per cent level ($P > .5$ for July, $P > .3$ for August and $P > .05$ for September) and suggest that the discrepancies between the observed and the predicted frequencies could arise due to chance factors. Thus, in general, a Markov-based geometric model seems to fit well for the prediction of wet spells.

(*ii*) *Dry-spells*

Unlike the wet spells, the dry spells do not seem to conform to the geometric model as solely judged from the final χ^2 values for each of the three months and for the consolidated period, although for the month of July, χ^2 is not significant. However, a noteworthy feature in these cases is that the observed frequencies are much in excess of the expected frequencies (based on geometric model) only at the end points, (*viz.*, the short and long dry spells) which together, contribute nearly 50 per cent to the total observed χ^2 value ; about 7.89 out of 15.13 in July, 14.90 out of 28.52 in August, 18.39 out of 33.01 in September and 53.49 out of 95.59 for the consolidated period, July to September.

Although it is hard to disregard this evidence against the assumed geometric distribution model for dry spells, we should also bear in mind, the possible implications of the arbitrariness in the definition of a dry day which includes all days with a rain of less than 0.05″. But the model itself does not theoretically presume this dependence. This

TABLE 2 (a)

Observed and expected frequencies of wet spells

Wet	Jul O	Jul E	Aug O	Aug E	Sep O	Sep E	Consolidated (Jul-Sep) O	Consolidated (Jul-Sep) E
1	136	136·3	144	139·5	130	123·5	410	388·5
2	72	67·8	66	69·0	47	59·4	185	196·2
3	33	33·7	35	34·1	22	28·6	90	97·2
4	14	16·7	11	16·9	19	13·8	44	48·1
5	5	8·3	7	8·4	10	6·6	22	23·1
6	11*	8·2	13*	8·1	10*	6·1	19	11·8
7							6	5·9
8							9†	14·2
Total	271	271	276	276	238	238	785	785
χ^2	4·12		4·38		10·03		8·55	
d.f.	5		5		5		7	
P	0·50-0·70		0·30-0·50		0·05-0·10		0·20-0·30	

* Spells ⩾ 6 days † Spells ⩾ 8 days
O-Observed E-Expected

TABLE 2 (b)

Observed and expected frequencies of dry spells

Dry	Jul O	Jul E	Aug O	Aug E	Sep O	Sep E	Consolidated (Jul-Sep) O	Consolidated (Jul-Sep) E
1	91	67·9	105	78·0	86	57·5	282	202·5
2	43	50·1	58	56·1	45	43·8	146	149·8
3	33	37·0	29	40·4	21	33·3	83	110·8
4	20	27·3	21	29·0	19	25·4	60	82·0
5	17	20·1	11	20·9	14	19·3	42	60·6
6	14	14·9	11	15·0	8	14·7	33	44·8
7	9	11·0	7	10·8	6	11·2	22	33·2
8	9	8·1	9	7·8	9	8·5	27	24·5
9	3	6·0	4	5·6	4	6·5	11	18·2
10	7	4·4	23‡	14·4	4	4·9	15	13·4
11	13†	12·2			24†	14·9	10	9·9
12							11	7·3
13							12	5·4
14							34*	15·6
Total	259	259	278	278	240	240	778	778
χ^2	15·31		28·52		33·01		95·59	
d.f.	10		9		10		13	
P	0·10-0·20		0·001		0·001		0.001	

† Spells ⩾ 11 ‡ Spells ⩾ 10 * Spells ⩾ 14

TABLE 3

Probabilities for a wet and dry days computed using (9)

	Jul	Aug	Sep	Jul-Sep
P (Wet day)	·4898	·5025	·5169	·5025
P (Dry day)	·7290	·7191	·7561	·7355

needs further investigation. However, the closer agreement between the observed and the expected frequencies in all other dry spell lengths is also a worthy point to note in the present situation.

It may be recalled here that the Markovian dependence of basic probabilities is not a precondition for the geometric distribution *per se*. In other words, we can estimate the probability of a wet day (or a dry day) directly from the observed frequency distribution for wet (or dry) spells *via* the assumption of a geometric distribution. In fact, if U has a geometric distribution, *viz.*,

$$P_{(U\,=\,k)} = p^{k-1}(1-p) \qquad k = 1, 2 \ldots\ldots$$

Then

$$E_{(U)} = \frac{1}{1-p}, \quad \text{Var }(U) = \frac{p}{(1-p)^2} \qquad (9)$$

Then for wet-spells p is the probability of a wet day and for dry-spells, the same represents the probability of a dry day. The maximum likelihood estimates of these probabilities may be obtained from the reciprocal values of the sample means of the respective observed frequency distributions. These estimates are computed for the present data (Table 2) for different months and are reported in Table 3.

Surprisingly we note that these estimates are closer to the conditional probabilities obtained in Table 1 under Markovian assumption and at the same time considerably different from the unconditional binomial probabilities reported in the same table, this being true for the case of every month as well as for the consolidated period. This is a good finding indeed and is a good pointer for the possible validity of the basic Markovian assumption regarding weather spells.

(*iii*) *Weather cycles*

As mentioned earlier, weather cycles are defined as the occurrence of wet and dry (or dry and wet) spells in succession. Table 4 reports the observed and the expected frequencies for wet-dry and dry-wet cycles. To obtain the expected frequencies the estimates of the basic probabilities reported in Table 1 are employed in (5)&(6) and the resulting probabilities multiplied by the observed

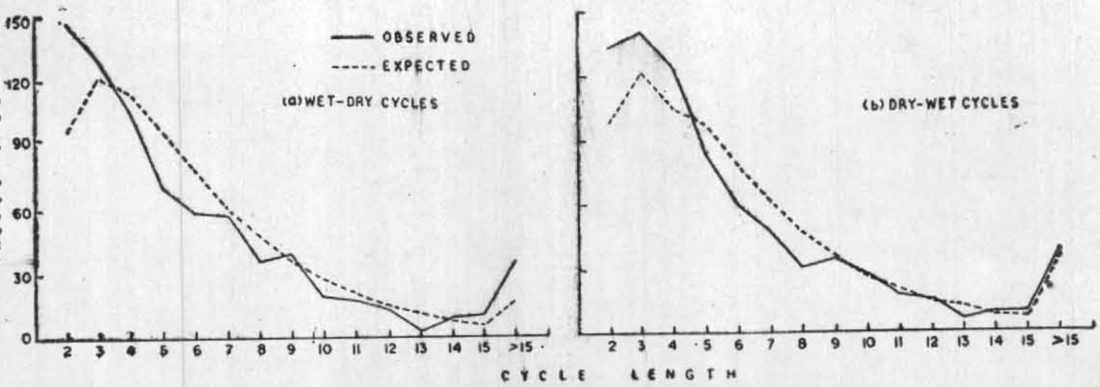

Fig. 1. Observed and expected frequencies of (a) wet-dry and (b) dry-wet cycles for the months of July, August and September

sample sizes. However it may be noted that, due to the symmetry of the expression (3) in $\pi_{(w/w)}$ $\pi_{(D/D)}$ the probabilities for different lengths of cycles would remain the same for wet-dry and dry-wet cycles, although the actual expected frequencies could be different due to the marginal differences in sample sizes. Fig. 1 shows the graph of the expected and the observed frequencies for weather cycles for the consolidated period, July to September.

As judged solely from the χ^2 values, the wet-dry and the dry-wet cycles do not appear to conform to the model (3) (χ^2-values not reported here). A finding similar to the case of dry spells seem to operate here also. Again, as observed elsewhere in the case of dry spell, the contributions, to χ^2–values from very short and very long cycles are substantial in this case also. Otherwise the agreement between the observed and the expected frequencies seems reasonable.

In any of the foregoing analyses, we have not tested directly for the fundamental Markovian assumption, *viz.*, the dependence of the occurrence of an event only on the outcome of the immediate preceding day but not on further preceding days. If this assumption is true, one should expect reasonable correspondence between observed frequencies and predicted frequencies based on the Markovian probabilities of future events given by the expressions (5) through (8).

The observed frequencies and the predicted frequencies of wet (or dry) days k days after a wet (or dry) day are also worked out for three specific values of $k = 2, 3$ and 4. The values, together with the χ^2 values are reported in Table 4.

As is evident from the χ^2 values, the goodness of fit between the observed and the expected frequencies is very poor in the case of wet days and reverse is true in the case of dry days. The breakdown in the case of wet days may likely to raise some doubts as to the validity of a Markovian assumption, if not against the geometric model itself. However for a large mass of data of the present type, accumulated over a long period of fifty years, the influence of many extraneous factors affecting the reliability and uniformity of data cannot be ruled out. Further the arbitrariness in the definition of a wet day, athough not implied by the conditions of the model nevertheless affect the outcome of the results, thus affecting the judgement about the validity of the models and their basic assumptions. Putting all the factors together and bearing in mind some of the inadequacies in the data-base itself, a geometric distribution model for weather spells and a Markovian type dependence of weather occurrence is not too unreasonable,

TABLE 4

The observed and the expected frequencies (to the nearest whole No.) and the χ^2 value for each row

	Observed		Expected		χ^2 (1d.f.) for each row
	Wet	Dry	Wet	Dry	
		$k=2$			
Wet	332	343	259	416	33·3
Dry	357	708	345	720	0·62
		$k=3$			
Wet	323	373	247	449	36·7
Dry	368	693	360	701	0·25
		$k=4$			
Wet	309	401	246	464	24·4
Dry	395	670	366	699	3·59

Acknowledgements

The authors are grateful to Shri Suryanarayana Rao, Assistant Meteorologist and Shri B.V. Ramana Rao, Assistant Professor of Physics, University of Agricultural Sciences, Bangalore for making the necessary data available for this study and also to Mr. Gurmurthy, Assistant Professor of Statistics, University of Agricultural Sciences for the preparation of the graphs in Fig. 1 and to the referee for his comments in improving the original version of this paper.

REFERENCES

Basu, A. N. 1971 *Indian J. Met. Geophys.*, **22**, pp. 67-73.

Cochran, W. G. 1938 *Quart. J.R. met. Soc.*, **64**, p. 631.

Gabriel, K. R. and Neumann, J. 1957 *Ibid.*, **83**, pp. 375-380.

1962 *Ibid.*, **88**, pp. 90-95.

Ramabhadran, V. K. 1954 *Indian J. Met. Geophys.*, **5**, pp. 48-55.