

MARKOV CHAIN MODELING OF DAILY RAINFALL OCCURRENCE IN THE MAHANADI DELTA OF INDIA

M. K. SUKLA¹, A. K. MANGARAJ² & L. N. SAHOO³

¹Department of Statistics, S.V. College, University of Delhi, New Delhi, India

²Department of Statistics, Rajendra Junior College, Bolangir, India

³Department of Statistics, Utkal University, Bhubaneswar, India

ABSTRACT

A two-state Markov chain probability model has been used to investigate the pattern of occurrence of daily precipitation during the rainy season over the Mahanadi delta of Odisha state of India. The study is based on the daily rainfall data for a period of 28 years for four meteorological stations in the region. Under the assumption of the dependence of daily precipitation on that of the previous day, attention has been focused on the analysis of certain aspects of the random structure of the precipitation phenomenon in conformity with the Markov chain properties.

KEYWORDS: Chi-square Test, Dry and Wet Spells, Goodness of Fit, Markov Chain Model, Steady State Probability, Transition Probability, Weather Cycle

1. INTRODUCTION

It is natural to expect that the total crop production and crop yield in any region depend not only on the total rainfall amount but also on the pattern of occurrence of rainfall such as spells of wet/rainy and dry days, the number of dry days between two rainy days, length of weather cycle etc. A model-based scientific study of the pattern of occurrence of daily rainfall at a regional level is therefore crucial to assess the crop failure due to deficiency or excess of rainfall for the rain fed agriculture under the local climatic conditions. Once the rainfall process is adequately and appropriately modeled, the model can be used to provide prior knowledge of the structural characteristics of varying rainfall systems which are very much essential for crop planning and management, and water management decisions. As the distribution of rainfall varies over space and time, it is required to analyze the data covering long periods recorded at various locations to obtain reliable information.

The Markov chain model has already been shown in many instances to be an appropriate model for studying sequences of wet and dry days [see, for example Gabriel and Neumann (1962)]. Because, other models for constant probabilities are not able to describe the daily persistence of wet and dry conditions. Weiss (1964), Basu (1971), Bhargava *et al.* (1973), Sundararaj and Ramachandra (1975), Aneja and Srivastava (1986), Rahman (1999a, 1999b), Ravindranan and Dani (1993), Akhter and Hossian (1998), Rahman *et al.* (2002), Banik *et al.* (2002), Zhao and Chu (2006), Spooof and Pryor (2008), Dastidar *et al.* (2010) among others analyzed situations that apply the Markov chain process.

The Mahanadi delta, which is situated on the eastern coast of India, gets rainfall from the south-west monsoon with an average annual rainfall 1572 mm and the total number of rainy days in a year ranging from 55 to 80 days. The most pre-dominant crop in this region is paddy covering about 95% of the total area under cultivation. As sufficient

supplementary irrigation facilities are not available in most parts, people mainly depend on winter and autumn paddy which are grown during monsoon season (June-September) and harvested during post-monsoon season (October and November). During monsoon season a large variety of vegetables are also grown here. Although the quantum of rainfall received by this river basin is fairly good, it's irregular distribution and variation in time and space leads to heavy downpour or very low precipitation in some areas. Variability in rainfall is therefore a cause of great stress to the farming activities, crop production and crop yield as the agriculture is mostly rain fed. Hence, there is need for in-depth study and understanding of the effects of various rainfall characteristics in this region for planning response measures.

This paper presents a statistical analysis in order to identify various characteristics of the pattern of occurrence of rainfall by applying a 2-state Markov chain probability model to the data on daily rainfall amount for 28 years during rainy season (June-October). The research is confined to only the rainy season because during this season, Mahanadi delta receives more than 85% of its total annual rainfall and agricultural activity depends on the amount of received rainfall.

2. DATA AND METHODOLOGY

Source and Nature of Data

The present investigation utilizes daily rainfall data of the four meteorological stations – Bhubaneswar, Cuttack, Paradip and Puri of the Mahanadi delta region for 28 years (1982-2009). The relevant data were collected from the Meteorological Centre, Bhubaneswar, Odisha. As the rainy season in the deltaic region is generally confined to five months *i.e.*, from June to October, the period considered for the study was taken from 1st June to 31st October of each year which also coincides with the growth season of the paddy crop, the major cash crop in the tract.

Markov Chain Model and Estimation

Let us identify a day as a rainy or wet day (a dry day) if it receives more than or equal to 2.5 mm (less than 2.5 mm) of rainfall according to the definition proposed by the Indian Meteorological Department [*cf.*, Basu (1971), Reddy *et al.* (1986)]. Further, under the assumption that the occurrence of a wet or a dry day is influenced only by the weather condition of the previous day, the process of occurrence of wet and dry days can be described by a 2 – state Markov chain with wet and dry days as the two states. The transition probability matrix, which describes the 2 – state Markov chain model is

$$P = \begin{bmatrix} P_{00} & P_{01} \\ P_{10} & P_{11} \end{bmatrix}, \quad (2.1)$$

with $P_{00} + P_{01} = 1$ and $P_{10} + P_{11} = 1$, where P_{00} , P_{01} , P_{10} and P_{11} are the transition probabilities *i.e.*, they are respectively the probabilities of occurrence of the following conditional events :

H_{00} : A day is a dry day given that the preceding day was a dry day

H_{01} : A day is a wet day given that the preceding day was a dry day

H_{10} : A day is a dry day given that the preceding day was a wet day

H_{11} : A day is a wet day given that the preceding day was a wet day

Suppose that each day from 1st June to 31st October of each year is classified according to the occurrence of the four events H_{00} , H_{01} , H_{10} and H_{11} such that 1st June is classified on the consideration of weather condition (wet or dry) of 31st May. Let a , b , c and d respectively be the observed frequencies of the occurrences of H_{00} , H_{01} , H_{10} and H_{11} . Then the

maximum likelihood estimates of the unknown transition probabilities (model parameters) P_{01} and P_{11} are obtained as

$$\hat{P}_{01} = p_{01} = \frac{b}{a+b} = \frac{b}{n_0} \quad \text{and} \quad \hat{P}_{11} = p_{11} = \frac{d}{c+d} = \frac{d}{n_1},$$

where $a + b = n_0$ and $c + d = n_1$. Estimated variances of p_{01} and p_{11} are respectively given by

$$v(p_{01}) = \frac{p_{01}(1-p_{01})}{n_0} \tag{2.2}$$

and

$$v(p_{11}) = \frac{p_{11}(1-p_{11})}{n_1} \tag{2.3}$$

[cf., Bhargava *et al.* (1973)].

The transition probabilities are conditional probabilities. But, the probability of a dry day (P_0) and the probability of a wet day (P_1) are also estimated from the observed frequencies of the conditional events as follows:

$$\hat{P}_0 = p_0 = \frac{a+c}{n_0+n_1} \quad \text{and} \quad \hat{P}_1 = p_1 = \frac{b+d}{n_0+n_1}.$$

In order to test that the occurrence of a wet or dry day is influenced by the immediately preceding day's weather, so that the Markov chain model works reasonably well, a normal test (assuming large n) can be employed by computing the usual normal deviate test statistic

$$Z = \frac{p_{01} - p_{11}}{\text{est. S.E. } (p_{01} - p_{11})} \tag{2.4}$$

[cf., Bhargava *et al.* (1973)].

To have good geographical coverage and good quality of information, the transition probabilities P_{01} and P_{11} are estimated by $\bar{p}_{01} = \frac{\sum b_i}{\sum n_{0i}}$ and $\bar{p}_{11} = \frac{\sum d_i}{\sum n_{1i}}$ respectively for the whole study region where $a_i, b_i, c_i, d_i, n_{0i}$ and n_{1i} are the respective values of a, b, c, d, n_0 and n_1 for the i^{th} station. Taking these estimates as the expected probabilities, we can apply two chi-square tests for each station, to test the discrepancies between the observed and the expected values of p_{01} and p_{11} . For the i^{th} station, the chi-square statistics, each with 1 degree of freedom (df), are defined by

$$\chi^2(p_{01}) = \frac{a_i^2}{n_{0i}(1-\bar{p}_{01})} + \frac{b_i^2}{n_{0i}\bar{p}_{01}} - n_{0i}, \tag{2.5}$$

and

$$\chi^2(p_{11}) = \frac{c_i^2}{n_{1i}(1-\bar{p}_{11})} + \frac{d_i^2}{n_{1i}\bar{p}_{11}} - n_{1i}, \quad i = 1, 2, 3, 4, \tag{2.6}$$

[cf., Rohatgi and Saleh (2000)].

As a follow-up to some earlier works on the 2-state Markov chains, the system of occurrence of the sequences of wet and dry days, after a sufficiently long period of time, is expected to settle down to a condition of statistical equilibrium with steady state or equilibrium probabilities which are independent of the initial conditions. These probabilities corresponding to dry and wet days are given by

$$\pi_0 = \frac{1-P_{11}}{1+P_{01}-P_{11}} \quad \text{and} \quad \pi_1 = \frac{P_{01}}{1+P_{01}-P_{11}},$$

respectively. The number of days after which the state of equilibrium *i.e.*, the original state is attained is equal to the number of steps or the power of the P –matrix so that its diagonal elements with equal to π_0 and π_1 [*cf.*, Cox and Miller (1967)].

Expected Lengths of Wet and Dry Spells

The wet and dry spell (or run) lengths are very important statistical descriptors of wet and dry periods in a geographical area. Assuming that the lengths of wet and dry spells (denoted by W and D respectively) follow a geometric distribution [*cf.*, Bhargava *et al.* (1973), Sundararaj and Ramachandra (1975), Ravindran and Dani (1993)], the probability of a wet spell of length x is given by

$$P(W = x) = (1 - P_{11})P_{11}^{x-1}, \quad x = 1, 2, \dots$$

and therefore, the expected length of the wet spell is obtained as

$$E(W) = \sum_{x=1}^{\infty} x (1 - P_{11})P_{11}^{x-1} = \frac{1}{1-P_{11}}. \quad (2.7)$$

On the other hand, the probability of a dry spell of length y is

$$P(D = y) = P_{01}(1 - P_{01})^{y-1}, \quad y = 1, 2, \dots$$

and the expected length of the dry spell is given as

$$E(D) = \frac{1}{P_{01}}. \quad (2.8)$$

Hence, the expected length of weather cycle C *i.e.*, a dry spell followed by a wet spell or vice-versa is given by

$$E(C) = E(W) + E(D) = \frac{1}{1-P_{11}} + \frac{1}{P_{01}}. \quad (2.9)$$

In order to test the strength of fitting of the geometric distribution for describing the distributions of dry and wet spell lengths under the Markovian preconditions of dependence, a chi-square goodness of fit test can be performed using the test statistic

$$\chi_g^2 = \sum_k \frac{(\text{observed frequency} - \text{expected frequency})^2}{\text{expected frequency}}, \quad (2.10)$$

which is asymptotically distributed as chi-square with $k - 1$ df, where k = number of spells.

As discussed in Cox and Miller (1967), the occurrence of the wet and dry days can be easily treated as dependent Bernoullian trials so that the expected values of the number of wet and dry days in a n - day period, denoted by W_n and D_n respectively, are given by

$$E(W_n) = n\pi_1 \quad \text{and} \quad E(D_n) = n\pi_0, \quad (2.11)$$

[*cf.*, Reddy *et al.* (1986)]. Assuming n to be large, the asymptotic variance of the number of wet (or dry) days in a n - day period is given by

$$V_n \sim \frac{nP_{01}(1-P_{11})(1+P_{11}-P_{01})}{(1-P_{11}+P_{01})^3} \quad (2.12)$$

[*cf.*, Bhargava *et al.* (1973)]

The maximum likelihood estimates of $\pi_0, \pi_1, E(W), E(D), E(C), E(W_n), E(D_n)$ and V_n are obtained in the usual way replacing P_{01} and P_{11} by their estimates p_{01} and p_{11} respectively. These estimates are denoted by $\hat{\pi}_0, \hat{\pi}_1, \hat{E}(W), \hat{E}(D), \hat{E}(C), \hat{E}(W_n), \hat{E}(D_n)$ and \hat{V}_n .

3. RESULTS AND DISCUSSIONS

Estimation of Model Parameters

The collected raw data on the daily rainfall are classified into four classes according to the conditional events H_{00}, H_{01}, H_{10} and H_{11} . From the actual frequencies of these classes, the corresponding relative frequencies are computed in order to obtain the maximum likelihood estimates of the transition (conditional) probabilities P_{00}, P_{01}, P_{10} and P_{11} along with the unconditional probabilities P_0 and P_1 for the four meteorological stations.

From the calculated value of the Z – statistic defined in (2.4), it is found that $|Z| > 3$ for all stations. This high significant value shows that the weather of a day is influenced by the weather of the previous day. Hence, the occurrences of wet and dry days in our study domain can be rightly modeled by a 2 – state Markov chain.

In order to test for differences in p_{01} and p_{11} from station to station, χ^2 -tests for the homogeneity of rainfall between the stations were run by using formulae (2.5) and (2.6). It is found that the calculated χ^2 values for all four stations are insignificant in respect of both the parameters P_{01} and P_{11} at 5 % as well as 1 % levels of significance. Therefore, the patterns of the occurrence of rainfall at these four precipitation stations are regarded as similar. Hence, their daily rainfall amounts are grouped together (pooled) in the usual manner in order to obtain a single estimate of the daily rainfall amount, to compose common estimates of the model parameters and to study various rainfall characteristics for the whole Mahanadi delta climatic situation.

Estimated values of the various unconditional probabilities and conditional probabilities associated with our 2-state Markov chain model, considered separately for each month and for their combination, are displayed in Table 3.1. Entries of this table clearly indicate that the probabilities of the rainfall are maximum and minimum in August and October respectively. The conditional probabilities for October show how rapidly and markedly the dry conditions establish themselves. The probabilities of wet conditions for this month are low, indicating dry conditions. The unconditional probabilities are pronouncedly different from those of the conditional ones.

Table 3.1: Estimates of Conditional and Unconditional Probabilities

Months	Conditional Probabilities				Unconditional Probabilities	
	P_{00}	P_{01}	P_{10}	P_{11}	P_0	P_1
June	0.7599	0.2401	0.4797	0.5203	0.6714	0.3286
July	0.6635	0.3365	0.4545	0.5455	0.5734	0.4266
August	0.6057	0.3943	0.4118	0.5882	0.5104	0.4896
September	0.7274	0.2726	0.4430	0.5570	0.6190	0.3810
October	0.8576	0.1424	0.4969	0.5031	0.7748	0.2252
June to October	0.7350	0.2650	0.4502	0.5498	0.6296	0.3704

Estimation of Expected Dry and Wet Days (With Spell Lengths)

Various statistical descriptors of the Markov chain model viz., estimated values of expected number of dry and wet days ($\hat{E}(D_n)$ and $\hat{E}(W_n)$) and their spell lengths ($\hat{E}(D)$ and $\hat{E}(W)$), length of weather cycle ($\hat{E}(C)$), S.D. of the estimated

number of wet or dry days, steady state probabilities ($\hat{\pi}_0$ and $\hat{\pi}_1$) and number of days required for equilibrium, as explained in the preceding section, are computed are compiled in Table 3.2.

Table 3.2: Statistical Descriptors the Markov Chain Probability Model

Statistical Descriptors	Months					
	June	July	August	September	October	June to October
$\hat{\pi}_0$	0.6664	0.5746	0.5109	0.6190	0.7772	0.6295
$\hat{\pi}_1$	0.3336	0.4254	0.4891	0.3810	0.2228	0.3705
$\hat{E}(D_n)$	19.9934 \cong 20	17.8122 \cong 18	15.8365 \cong 16	18.5714 \cong 19	24.0934 \cong 24	96.3090 \cong 96
$\hat{E}(W_n)$	10.0066 \cong 10	13.1878 \cong 13	15.1635 \cong 15	11.4286 \cong 11	6.9066 \cong 07	56.6910 \cong 57
$\hat{E}(D)$	4.1649	2.9714	2.5339	3.6684	7.0210	3.7732
$\hat{E}(W)$	2.0845	2.2000	2.4282	2.2575	2.0126	2.2210
$\hat{E}(C)$	6.2494	5.1714	4.9641	5.9259	9.0336	5.9942
S. D. of dry or wet days	3.4438	3.4029	3.3869	3.5636	3.3801	8.0060
No. of days to equilibrium	09	08	07	09	13	10

Table 3.3: Observed and Expected Frequencies of Wet and Dry Spells (June-October)

Spell Length (Days)	Wet Spell		Dry Spell	
	Observed Frequency	Expected Frequency	Observed Frequency	Expected Frequency
1	1369	1316	903	824
2	655	724	642	606
3	412	398	440	445
4	242	219	309	327
5	114	120	207	241
6	66	66	157	177
7	28	36	112	130
8	16	20	87	96
9	10	11	65	70
10	12*	14*	43	52
11	-	-	35	38
12	-	-	26	28
13	-	-	19	21
14	-	-	15	15
15	-	-	11	11
16	-	-	10	8
17	-	-	9	6
18	-	-	21**	16**
Calculated Value of χ^2 -Statistic (χ_g^2)	14.872		27.199	
Degrees of Freedom	09		17	
1% Critical Value of χ^2 ($\chi_{0.01}^2$)	21.666		33.409	
5% Critical Value of χ^2 ($\chi_{0.05}^2$)	16.919		27.587	

*Frequencies corresponding to spell length ≥ 10

** Frequencies corresponding to spell length ≥ 18

From Table 3.2, it can be seen that the expected length of a dry spell varies from 2.5339 to 7.0210 days whereas that of wet spell varies from 2.0126 to 2.4282 days. This means that after every 3 to 7 consecutive dry days, a wet day is likely to occur and after every 2 consecutive wet days, a dry day is likely to occur. However, computed overall expected values of the spell lengths indicate that after every 4 consecutive dry days, a wet day is expected and after every 2 consecutive wet days, a dry day is expected during the rainy season. Hence, for this period the expected length of weather cycle is about 6 days.

It is also evident from the Table 3.2 that the months August possess the highest number of expected rainy days *i.e.*, 15 days and the lowest number of expected dry days *i.e.*, 16 days. Assuming that the variables W_n and D_n follow normal distribution, we have computed 95% confidence intervals for $E(W_n)$ and $E(D_n)$. From these confidence intervals we may conclude that the rainy days (dry days) are expected to lie between 41 to 72 days (81 to 112 days) during the period of 153 days of the rainy season.

For August, $\hat{\pi}_0$ and $\hat{\pi}_1$ values are respectively smaller and larger than other months and for the consolidated period from June to October these values are 0.6295 and 0.3705 respectively. As the number of days to equilibrium for the months varies from 7 to 13 days, this proves that after 7 to 13 days, during the rainy season, the probability of the day being wet or being dry is independent of the initial weather conditions.

As a check of the adequacy of the Markov based geometric distribution that is fitted to the lengths of dry and wet spells, the χ^2 -test for goodness of fit at 1% and 5% levels of significance has been applied using the test statistic defined in (2.10). The test results for the different months are more or less similar and they provide evidence for quite good fit in each case. We do not present the results in respect of goodness of fit test for the individual months but results for the consolidated months (June – October) *i.e.*, for the rainy season in Table 3.3. The insignificant values of the calculated χ^2 statistic prove that the observed distributions of the lengths of wet and dry spells are very well fitted by the Markov based geometric distribution. The graphical representations of the observed and expected frequency distributions for the wet and dry spell lengths are shown in Figures 3.1 and 3.2 respectively to illustrate the closeness of such fits visually.

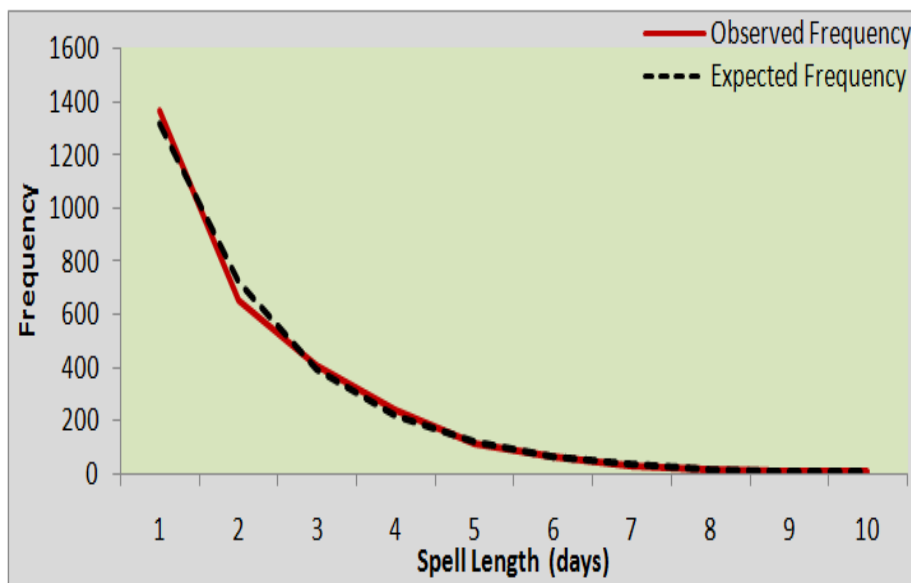


Figure 3.1: Observed and Expected Frequencies of Wet Spells

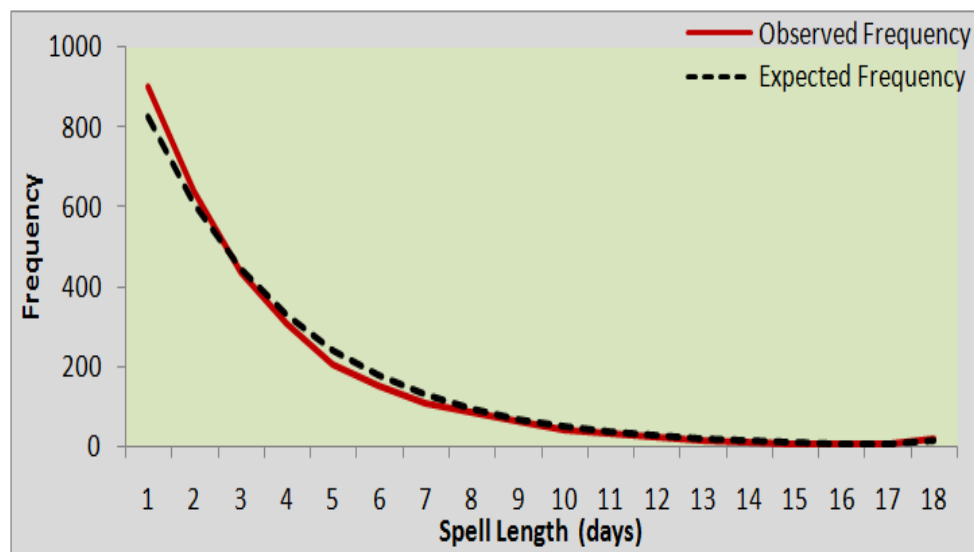


Figure 3.2: Observed and Expected Frequencies of Dry Spells

4. CONCLUSIONS

Our study on the different aspects of the pattern of daily rainfall occurrence at Mahanadi delta leads to the following tentative conclusions:

- The Markov chain probability model appears to have a good approximation for describing the occurrence of the sequence of wet and dry days.
- The rainfall distributions of the four meteorological stations of the study domain exhibits more or less similar pattern.
- On the whole, as judged by the χ^2 - test of goodness of fit, the geometric distribution under the assumption of Markovian dependence of weather occurrence seems to be satisfactory for describing the distributions of wet and dry spell lengths.
- During the rainy season, the expected lengths of dry spell and wet spell are about 4 days and 2 days respectively, and the system settles down after about 7 to 13 days to a condition of statistical equilibrium in which the occupation probabilities are independent of the initial conditions. The estimated ranges for the expected numbers of dry days and rainy days during the period of 153 days are 81–112 and 41–72 days respectively.

Although the Markov chain model provides a satisfactory fit to our daily rainfall data for evaluating probability of occurrence of the sequence of wet or dry days, we stress that more detailed and exhaustive investigations may be made with the help of other models and with new definitions of wet and dry days as well as other goodness of fit tests.

REFERENCES

1. Akhter, S. S. and Hossain, M. F. (1998). A study on rainfall occurrences using probability approach. *Journal of Statistical studies*, **18**, 37-50.
2. Aneja, D. R. and Srivastava, O. P. (1986). A study technique for rainfall pattern. *The Aligarh Journal of Statistics*, **6**, 26-31.

3. Banik, P., Mandal, A. and Rahman, M. S. (2002). Markov chain analysis of weekly rainfall data in determining drought proneness. *Discrete Dynamics in Nature and Society*, **7**, 231-239.
4. Basu, A. N. (1971). Fitting of a Markov chain model for daily rainfall data at Calcutta. *Indian Journal of Meteorological Geophysics*, **22**, 67-74.
5. Bhargava, P. N., Narain, P., Areja, K. G. and Pradhan, Asha (1973). A study of the occurrence of rainfall in Raipur District with the help of Markov chain model. *Journal of the Indian Society of Agricultural Statistics*, **25**, 197-204.
6. Cox, D. R. and Miller, H. D. (1967). *The Theory of Stochastic Process*. New York: John Wiley and Sons.
7. Dastidar, A.G., Ghosh, D., Dasgupta, S. and De, U.K. (2010). Higher order Markov chain models for monsoon rainfall over West Bengal, India. *Indian Journal of Radio and Space Physics*, **39**, 39-44.
8. Gabriel, K. R. and Neumann, J. (1962). A Markov chain model for daily rainfall occurrence at Tel Aviv. *Quarterly Journal of the Royal Meteorological Society*, **88**, 90-95.
9. Rahman, M. S. (1999a). A stochastic simulated first order Markov chain model for daily rainfall at Barind, Bangladesh. *Journal of Interdisciplinary Mathematics*, **2**, 7-32.
10. Rahman, M. S. (1999b). Logistic regression estimation of a simulated Markov chain model for daily rainfall in Bangladesh. *Journal of Interdisciplinary Mathematics*, **2**, 33-40.
11. Rahman, M. S. and Basher Mian, M.A. (2002). Stochastic study of asymptotic behavior of rainfall distribution in Rajshahi. *International Journal of Statistical Sciences*, **1**, 36-48.
12. Ravindranan, C. D. and Dani, R. G. (1993). Markov based models for weather spells – A case study. *Journal of the Indian Society of Agricultural Statistics*, **45**, 285-297.
13. Reddey, J. V. L. N., Kulkarni, V. S. and Nageshwar Rao, G. (1986). Rainfall distribution of Hyderabad: A qualitative and quantitative analysis. *Journal of Research APAU*, **11**, 190-194.
14. Rohatgi, V. K. and Saleh, S. (2000). *An Introduction to Probability Theory and Mathematical Statistics*. New Delhi: Wiley Eastern Limited.
15. Spooft, J.T. and Pryor, S.C. (2008). On the proper order of Markov chain model for daily precipitation occurrence in the contiguous United States. *Journal of Applied Metrology*, **47**, 2477-2486.
16. Sundararaj, N. and Ramachandra, S. (1975). Markov dependent geometric models for weather spells and weather cycles – A study. *Indian Journal of Meteorological Hydrology and Geophysics*, **26**, 221-226.
17. Weiss, L. L. (1964). Sequences of wet and dry days described by a Markov chain probability model. *Monthly Weather Review*, **92**, 169-176.
18. Zhao, X. and Chu Pao-Shin (2006). Bayesian multiple change point analysis of Hurricane activity in the eastern North Pacific: A Markov chain Monte Carlo approach. *Journal of Climate*, **19**, 564-578.

