Int.J.Curr.Microbiol.App.Sci (2018) *7*(5): 3762-3767

International Journal of Current Microbiology and Applied Sciences
ISSN: 2319-7706 **Volume 7 Number 05 (2018)**
Journal homepage: http://www.ijcmas.com

Original Research Article https://doi.org/10.20546/ijcmas.2018.705.436

Probabilistic Analysis of Monsoon Daily Rainfall at Hisar Using Information Theory and Markovian Model Approach

Gurjeet Kaur[1*], Safeer Alam[2] and Jasvinder Kaur[3]

[1]MGCGVV, Chitrakoot, Satna, M.P., India
[2]Directorate of Extension, SKUAST-Kashmir, Shalimar, Srinagar, J&K, India
[3]Extension Education Institute, Nilokheri, CCSHAU, Hisar, India
**Corresponding author*

Keywords

Markov chain, Maximum likelihood principle, Measure of entropy, Redundancy test, Redundancy, Shannon's information theory, Transitional probabilities

Article Info

Accepted:
30 April 2018
Available Online:
10 May 2018

A B S T R A C T

A study of south-west monsoon daily rainfall at Hisar was carried out for the period of 1925-2003 (79 years). In semi-arid tropics the weather conditions have a tendency to cluster together to a certain extent which means that the occurrence of rainfall on a particular day depends on the weather conditions of the previous day. This reality of meteorological persistence can be best described by Markov Chain of order one. In this paper we have obtained the probability matrices of transition from one state of occurrence to other state by Markov chain model for different categories of rainfall. The disorderliness of the transitional system of South-West monsoon rainfall by maximum likelihood estimation method and the redundancy of the discrete information source with memory has been found out. The entropy of the Markovian system has been obtained and favorableness is tested by Redundancy test. At Hisar during South-West monsoon season, the Markovian system has been found to be favorable by redundancy test.

Introduction

The agricultural scenario of India is closely linked with the distribution of rainfall throughout the year. Several researchers have used statistical technique for studying the behavior of rainfall and related aspects. Fisher (1924) was the first to make a systematic study in this field by studying the influence of rainfall on the yield of wheat at Rothamsted Agricultural Experiment Station and showed that distribution of rainfall during a season is more effective force than its total amount in the yield of a crop. Several authors like Gabriel and Neumann (1962) and Medhi (1976) have studied the probability of occurrences of dry and wet days through Markovian Model. Robertson *et al.,* (2008) used a hidden Markov Model to investigate the seasonal predictability of rainfall over a small rice-growing district of Java, Indonesia in terms of daily rainfall characteristics like rainfall frequency, mean daily intensity, median length of dry spells as well as the onset date of rainy season. Basak (2014) used Markov Chain Models of various orders to analyzed the probability distribution of pattern of rainfall during monsoon season (June-September) over different regions of West Bengal based on data of 25 years (1971-1995) for ten meteorological stations and find out that the first order Markov Chain model is the best for rainfall forecasting. Tan *et al.,* (2016) carried out a study on the northeast rainfall monsoon data of 40 stations in Peninsular Malaysia for a period of 1975 to 2008 using a homogeneous Hidden Markov Model (HMM) and they found that model is able to assess the behavior of rainfall characteristics. Basu (1988) has studied the probability of occurrences of different rainfall characteristics through one step transitional probability matrices during monsoon period at Maithon by markovian model. Domenico (1972) has suggested the test of uncertainty of transitional probabilities of occurrence for determining favorable or unfavourableness of the system through redundancy method.

Present study is on the rainfall data of Hisar which is in the Haryana state of India. Haryana state receives rainfall in the range of 300mm-1200mm in different agro-climatic zones. Hisar is situated at latitude $29^0 10^/$N longitude $75^0 76^/$ E. Its height is 215.2 mt above mean sea level. Hisar receives an average of 318mm rainfall in S-W monsoon which is about 81% of the average annual rainfall. In this study the rainfall at Hisar was categorized into five classes according to definition such as non-rainy days, light rains, moderate rains, heavy rains, very heavy rains. The main objective of this study is to test whether the individual state of occurrence depends on previous state of occurrence by using Markovian model and to estimate the transitional probabilities of all states of occurrence. Data of daily rainfall at Hisar during monsoon period for 79 years (1925-2003) have been used here.

Materials and Methods

Markov process

If $\{X(t), t \in T\}$ is a stochastic process such that given the value X(s), the value of X(t), t>s do not depend on the value of X(u), u<s then the process $\{X(t), t \in T\}$ is called a Markov process.

Markov chain

A Markov process is called a Markov chain if the parameter space is discrete. A stochastic process $\{X_n, n = 0, 1, 2\ldots\}$ is called a Markov chain of order one if for i, $i_1 \ldots i_{n-1}, \in N$.

$$P (X_n = j/X_{n-1} = i, X_{n-2} = i_1 \ldots X_0 = i_{n-1}) \\ = P (X_n = j / X_{n-1} = i)$$

The outcomes E_i (or the values of i) are called the states of Markov chain. If X_n has outcome E_i (i.e. $X_n = i$) the process is said to be in the i^{th} state at n^{th} trial.

Order of Markov chain

A Markov chain $\{X_n\}$ is said to be of order s (s=1, 2…) if for all n,

$$Pr (X_n = j/X_{n-1} = i, X_{n-2} = i_1, \ldots X_{n-s} = i_{s-1}, \ldots) \\ = Pr (X_n = j/X_{n-1} = i, X_{n-2} = i_1, \ldots X_{n-s} = j_{s-1})$$

A Markov chain $\{X_n\}$ is said to be of order one if

$$Pr (X_n = j/X_{n-1} = i, X_{n-2} = i_1, \ldots) \\ = Pr (X_n = j / X_{n-1} = i) \\ = p_{ij}$$

A Markov chain is said to be of order zero if $p_{ij} = p_j$ for all i. This implies independence of X_n on X_{n-1} (Medhi, 1982).

Estimation of transitional probabilities

South-West monsoon daily rainfall amount of Hisar have been classified into five different classes, mentioned above. A one step 5X5 transitional probability matrix from one transitional state to another has been found for each of the rainy month from June to September from the frequency of occurrence of daily rainfall amount at Hisar. The Maximum Likelihood method of estimation of transitional probabilities daily rainfall has been used. The disorderness of such transitional probability is tested from the redundancy of the system. Domenico (1972) has suggested the test of uncertainty of transitional probabilities of all states of occurrence for determining the favourable or unfavourable of the system through redundancy method.

One step 5X5 transitional probability matrices from one transitional state to another have been computed for each monsoon month. The transitional probability matrix is given by

$$P_{ij} = \begin{pmatrix} P_{00} \ldots\ldots\ldots\ldots P_{04} \\ P_{10} \ldots\ldots\ldots\ldots P_{14} \\ \vdots \qquad\qquad \vdots \\ \vdots \qquad\qquad \vdots \\ P_{40} \ldots\ldots\ldots\ldots P_{44} \end{pmatrix}$$

Where P_{ij} (i, j = 0, 1,--------4) is the transitional probability from the j^{th} state of occurrence to the i^{th} state and $\sum P_{ij} = 1$ for each i. Here '0' indicates non rainy days; '1' for light rains, '2' for moderate rain; '3' for heavy rains; '4' for very heavy rains.

Information theory

Information theory is a branch of the mathematical theory of probability and mathematical statistics that quantifies the concept of information. Shannon (1948) is generally considered to be the founder of the information theory, he associate information with uncertainty using the concept of probability.

Average amount of information

Let X be a discrete random variable taking a finite number of possible value $x_1, x_2 \ldots x_n$ with probabilities $p_1, p_2 \ldots. p_n$ respectively such that $p_i \geq 0$, i = 1, 2 ….. n and $\sum_{i=1}^{n} p_i = 1$.

The average amount of information or amount of uncertainty associated with the event X = x_i, i = 1, 2, ….n is given by

$$H(X) = -\sum p_i \log p_i$$

When logarithm to base 10 is used, the unit of information H, is called "Hartley" (Tuller, 1950) and when natural logarithm is used the unit of information H, is called "nit" (MacDonald, 1952).

H(x) is also called the Shannon's measure of entropy or average uncertainty associated with the event $\{X = x_i\}$, i = 1, ….n.

In case of two outcomes with probabilities p_1 = p and q = 1-p, we find the measure of information as

$$H = -p \log p - (1-p) \log (1-p)$$

Measure of entropy

A mathematical model of Shanon serves as a measure of uncertainty (entropy) of the transitional probability (P_{ij}) of the above system and is:

$$H_i = \sum_{i=0}^{4} P_{ij} \log P_{ij} \text{ for each } i$$

Where H_i denotes the entropy of the ith state of occurrence (i= 0,1,-----,4).

The entropy H of the Markovian system of transitional probability matrix P_{ij} obtained from the probability of individual state of occurrence P_i and weighted entropy H_i and is given by

$$H = -\sum_{i=0}^{4} P_i H_i$$

Where P_i is the probability of the i^{th} state of occurrence.

Testing favorable or unfavorableness of the system

As the probability $P_{i,\ of}$ the individual state of occurrence of categorized rainfall is taken as random, its distribution is given by using Shannon's formula for information (in bits) of the system in the for $I(P_i) = -\log_{10} P_i$

The favorable and unfavorableness of the system is tested by redundancy R. The redundancy R of the state of occurrence is obtained as the difference from 1 of the ratio of the weighted entropy value H to the maximum possible entropy, is given by

$$R = 1 - H / H_{max}$$

Where $H_{max} = \log 5$ ($\log 5 = 0.6990$)

This redundancy value is used to determine the favorable or unfavorableness of the system. When R tends to 1, the system tends to maximum favorable condition i.e. almost certain.

Computational procedure

The rainfall at Hisar has been categorized into five states according to amount.

In Table 1, the conditional transitional probabilities P_{ij} are calculated by dividing the frequencies of one state of occurrence for the day followed from another state of occurrence of the previous day f_{ij} by the frequencies of the state f_i, i.e. $P_{ij} = f_{ij} / f_i$ and are arranged in the form of a matrix for each monsoon month. In Table 2, the probability and corresponding information (in bits) for different states of occurrence for each of the monsoon month at Hisar has been calculated. In Table 3, entropy and redundancy has been calculated for south-west monsoon period. The weighted entropy values for each of the months are calculated as a sum of the entropy values of all the different categories with the random probability of the corresponding state.

Results and Discussion

Estimation of transitional probability

If the previous day is dry then the results show that the probability of next day to be dry is maximum (i.e. 0.94) in the month of June and September while in the month of July and August it is around 0.88.

The probability that of a light rain (6-15mm) if the last day is dry is obtained as 0.06, 0.05 and 0.03 in the month of July, August and September respectively while the probability is negligible of heavy and very heavy rain in July, August and September if the last day is dry.

If there has been a light rain on any day then the probability lies between 0.74 and 0.79 in all the four months for next day to be dry. But a combination of W_1W_1 has a maximum probability in the month of June and September as 0.12 and 0.10 respectively while in July and August it is around 0.08.

The probability of the consecutive days having moderate rain (15-35mm) also lies between 0.05 and 0.10 in the whole south-west monsoon season from June to September (the maximum 0.10 in the month of August). The probability of combination of heavy rain (35-75mm) and moderate rain (15-35mm) is maximum in the month of September i.e. 0.20 and is equal to 0.18 in the month of July while in the month of June and August it is 0.08 and 0.12 respectively.

Table.1 Transitional probability matrixes for each monsoon month (June- September) at Hisar

June

$$P_{ij}=\begin{pmatrix} .94 & .04 & .02 & .00 & .00 \\ .75 & .12 & .12 & .01 & .00 \\ .81 & .12 & .05 & .02 & .00 \\ .75 & .08 & .08 & .08 & .00 \\ 1 & .00 & .00 & .00 & .00 \end{pmatrix}$$

July

$$P_{ij}=\begin{pmatrix} .88 & .06 & .04 & .02 & .00 \\ .74 & .09 & .09 & .07 & .01 \\ .71 & .11 & .09 & .08 & .01 \\ .56 & .16 & .18 & .08 & .01 \\ .85 & .15 & .00 & .00 & .00 \end{pmatrix}$$

August

$$P_{ij}=\begin{pmatrix} .89 & .05 & .04 & .02 & .00 \\ .79 & .08 & .12 & .01 & .00 \\ .75 & .08 & .10 & .06 & .00 \\ .68 & .09 & .12 & .11 & .00 \\ .43 & .29 & .14 & .14 & .00 \end{pmatrix}$$

September

$$P_{ij}=\begin{pmatrix} .94 & .03 & .02 & .01 & .00 \\ .79 & .10 & .05 & .04 & .02 \\ .73 & .15 & .05 & .07 & .00 \\ .63 & .06 & .20 & .05 & .06 \\ .33 & .00 & .17 & .50 & .00 \end{pmatrix}$$

Table.2 Probability and corresponding information for different individual state of occurrence for each monsoon month (June to September) at Hisar

Prob. of Occurrences P_i	Information in bits $I(P_i)= \log P_i$			
	June	July	August	September
Non-rainy day $P_0/I(P_0)$	.93/.03	.84/.08	.86/.07	.92/.04
Light rain $P_1/I(P_1)$	.04/1.39	.07/1.15	.06/1.22	.04/1.39
Moderate rain $P_2/I(P_2)$	.02/1.69	.05/1.30	.05/1.30	.03/1.52
Heavy rain $P_3/I(P_3)$	.01/2.0	.03/1.52	.03/1.52	.01/2.0
Very heavy rain $P_4/I(P_4)$	.00/-	.01/2.0	.00/-	.00/-
$-\sum P_i \log P_i$	0.14	0.28	0.24	.15

Table.3 Measure of entropy and redundancy at Hisar

	June	July	August	September
Non-rainy days (H_0)	0.12	0.21	0.24	0.12
Light rain (H_1)	0.33	0.39	0.30	0.35
Moderate rain (H_2)	0.28	0.41	0.37	0.37
Heavy rain (H_3)	0.36	0.51	0.42	0.48
Very Heavy rain (H_4)	0.00	0.18	0.55	0.44
$H= -\sum_{i=0}^{4} P_i H_i$	0.13	0.24	0.21	0.14
Redundancy, R (%)	81	66	70	80
Measure of entropy $M= (-\Sigma P_i \log P_i)- H$	.01	0.04	0.03	0.02

Computational procedure

Categories	Daily rainfall amount (mm)
0. Non-rainy days (D)	0-6
1. Light rain (W_1)	6-15
2. Moderate rain (W_2)	15-35
3. Heavy rain (W_3)	35-75
4. Very heavy rain (W_4)	≥75

Measure of entropy and redundancy test

The value of weighted entropies of the different categorized states of rainfall amounts and also the weighted entropies for each of the monsoon month, June to September at Hisar are given in Table 3. It is found that the value of the weighted entropy is less during the beginning month June and ending month September as compare to July and august at Hisar as shown in Table 3. The redundancy values of the system during monsoon period at Hisar are obtained as shown in Table 3. The maximum value is 81% during the month of June and minimum value is 66% during July at Hisar. This indicates that favorableness of the Markovian system is maximum in July at Hisar during the monsoon period.

The transitional probability at all states of occurrence for each of the monsoon month at Hisar is estimated by M.L.P. The entropy (disorderness) of such probability for each of the categorized classes of rainfall amount at Hisar during monsoon period with their weighted entropy values have been determined and the monthly redundancy of the system for each of the monsoon month have been find out on the basis of weighted entropy and H_{max}. The result in Table 3 shows that the redundancy value is more during the month of June and September as compare to its value in July and August. Since this redundancy value is used to determine the favorableness or unfavorablness of the system, as the redundancy value R tends to 1, the system tends to maximum favorable conditions i.e. almost certain. Hence, at Hisar during S-W monsoon season, the Markovian system has been found to be favorable by redundancy test.

References

Basak, P. 2014. On the Markov Chain Model for monsoon rainfall occurrence in different zones of West-Bengal. Indian Journal of Radio and Space Physics. 43: 349-354.

Basu, G.C. 1988. A study of monsoon daily rainfall at Maithon by Markovian model and information theory. Mausam. 39(1): 83-86

Domenico, P. A.1972. Concept and Models in Groundwater Hydrology. pp. 2-57.

Fisher, R. A. 1924. Influence of rainfall on the yield of wheat at Rothamsted. Phel. Trans. Roy. Stat. London. B 213: 89-142.

Gabriel, K. R. and Neumann, J. 1962. A Markov chain model for daily rainfall in Tel-Aviv. Quart. J. R. Met. Soc. 88: 90-95.

Medhi, J. 1976. A Markov chain model for the occurrence of dry and wet days. Ind. J. Met. Hyd. and Geophys. 27(4): 431-435.

Robertson, A.W., Moron, V. and Swarinoto, Y. 2008. Seasonal predictability of daily rainfall statistics over Indramayu district Indonesia. Int. J. Climatol. Published online (www.interscience.wiley.com) with doi: 10.1002/joc.1816.

Tan, W.L., Yusof, F. and Yusop, Z. 2016. Sub seasonal to multidecadal variability of northeast monsoon daily rainfall over Peninsular Malaysia using a hidden Markov Model. Theor. Appl. Climatol. Published online with doi: 10.1007/s00704-016-1795-9.

How to cite this article:

Gurjeet Kaur, Safeer Alam and Jasvinder Kaur. 2018. Probabilistic Analysis of Monsoon Daily Rainfall at Hisar Using Information Theory and Markovian Model Approach. *Int.J.Curr.Microbiol.App.Sci.* 7(05): 3762-3767. doi: https://doi.org/10.20546/ijcmas.2018.705.436