Journal of Crop and Weed, 15(2): 52-57 (2019)
ISSN- O : 2349 9400 ; P : 0974 6315
http://cwssbckv.org
www.cropandweed.com

An information theory approach for active monsoon period of coastal zone of Karnataka

S. BHAVYASHREE AND B. BHATTACHARYYA

Department of Agricultural Statistics,
Bidhan Chandra Krishi Viswavidyalaya, Mohanpur, Nadia, West Bengal

Received : 06.02.2019 ; Revised : 21.06.2019 ; Accepted : 06.07.2019

ABSTRACT

A study of monthly rainfall for active monsoon period from June to September was carried out for the period of 1980-2013 (34 years) for coastal zone of Karnataka i.e, Dakshina Kannada and Udapi districts. 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 Markov chain is one of the best tool for describing the meteorological persistence among the tools which are available. The studies of uncertainty about the nature of day's rainfall are rarely available. Every probability distribution has some uncertainty associated with it. The concept of Information theory approach (entropy) is introduced here to provide a quantitative measure of this uncertainty, and to test the stochastic Markovian dependence. The redundancy of state of occurrences is ranging from 35 to 46 per cent. A test equivalent to the likelihood ratio test has shown better result against, the Markov dependence on the same observations. Using likelihood ratio test, Markovain dependence is best fitted during the study period.

***Keywords*:** Entropy, information theory, likelihood ratio test, Markovian dependence, redundancy test and uncertainty.

Karnataka witnesses three types of climate. The state has a dynamic and erratic weather that changes from place to place within its territory. Due to erratic geo-physiographic features, State faces climatic variations that range from dry to semi dry in upland areas, sub-humid to humid dry in the Western Ghats and humid tropical monsoon in coastal regions. More than 75 per cent of Karnataka's total area, comprising interior Karnataka, observes dry or semi-dry climate. Karnataka has about 15 per cent of the total semi-dry area (3 per cent of the total dry area) marked in India. Hence, Karnataka is divided into 3 meteorological divisions (with 10 Agro climatic zones) based on prevailing climate, Coastal Karnataka, North Interior Karnataka and South Interior Karnataka. The *Coastal Karnataka* comprises of Udupi, Uttara Kannada and Dakshina Kannada districts. Entire belt and the surrounding areas have sultry monsoon with heavy showers. Average rainfall of this region is about 3456 mm per annum which is much higher than the rainfall received in the other parts of the state.

The knowledge of weather in particular of rainfall is much needed for performing the pre-cultivation operations and sound crop planning. Mostly the work in this direction is confined to the total rainfall, maximum and minimum amount of rainfall and its ranges *etc*. However, not only the inadequacy of rainfall but also excess or deficiency is harmful to a crop growth. It is also natural to imagine that for the total agricultural production the total amount of rainfall is not of much importance but the pattern of its occurrence such as spell of rainy and dry days expected number of dry days between two rainy days and their repetitions are of much use for the rainfed crops in this region.

Some researches focused on measure of uncertainty on nature of day's rainfall, Satish and Banjul Bhattacharyya (2017) studied the concept of Information theory approach (entropy) to provide a quantitative measure of this uncertainty, and to test the stochastic Markovian dependence. The study revealed that redundancy values (R) in Hill sub region of West Bengal *i.e.*, Darjeeling, during monsoon months are ranging from 0.116 to 0.536, these values are indicating that Markovian system tends to minimum favorable conditions *i.e.*, almost uncertain to follow Markovian dependence. But, the likelihood ratio test is the most powerful test to check the Markovian dependence. Except in September, Markovian dependence was best fitted during the study period using the likelihood ratio test. Basu (1988) have been studied monsoon daily rainfall at Maithan by Markovian model and he employed information theory to study the uncertainty of daily rainfall. He evaluated the nature of rainfall with the help of information theory. It was revealed that the weighted entropy values over the transition probabilities during the beginning and closing months of monsoon at Maithan were less in comparison with the other monsoon months. The weighted entropy of the active monsoon months was higher. It was also revealed that the difference in uncertainty between the Markovian model and the random model were small almost negligible. He observed that during the monsoon months the degree of uncertainty of heavy rainfall and very heavy rainfall were more than the light, moderate and non-rainy rainfalls.

Email : bhavyashree5591@gmail.com

He used the redundancy test to determine the favourableness and unfavourableness of Markovian system. Kawachi et al. (2001) employed the informational entropy to measure the degree of uncertainty of rainfall occurrence in time or the over-a-year temporal rainfall apportionment based on the probability density function of rainfall, randomly apportioned over fragmented times. Pechlivanidiset al. (2010) used Entropy for Model diagnostic in rainfall modeling.

MATERIALS AND METHODS

For the study, daily rainfall data for active monsoon period (May31st to September 31st) of Dakshina Kannnada and Udapi districts (which covers costal zone of Karnataka) was collected for the period of 34 years (1980-2013) from AICRP on Agro Meteorology, University of Agricultural Sciences, GKVK, Bengaluru and India Meteorological Department, Pune.

Stochastic matrices

A unit step 5x5 Transition Probability Matrix from one transition state to another has been formed for each month, June to September and the season from the frequency of the occurrence of daily rainfall at Dakshina Kannada and Udupi districts separately.

In the present study, the choice of minimum threshold values for determining the states of Markovian model on daily rainfall are of much implications whenever this information of a region are extremely valuable in planning agricultural operational adjustments in rainfed farming and also critical to plants as well as various human activities. Bearing it in mind, the intensity of daily rainfall have been assigned to five categories and grouped as follows:

Categories	Nature of daily rainfall	Intensity of daily rainfall (in mm)
H_1	Non rainy day	0.00-2.40
H_2	Light rainy day	2.50-10.00
H_3	Moderate rainy day	10.10-50.00
H_4	Heavy rainy day	50.10-125.00
H_5	Very heavy rainy day	125.1 and above

The mode of classification yields a sequence of non rainy, light rainy, moderate rainy, heavy rainy day, very heavy rainy days which can be regarded as a five states Markov chain. The classification is successfully used by Basu (1988). There is twenty five possibilities in which each daily rainfall with its transition may be classified as one of them. Thus for each year, the nature of rainfall on 1st June is classified as one of them which also depends on the nature of rainfall on 31st May for transition property. Repeating this procedure for each year under the experimental period, the cell frequencies for twenty five possibilities are obtained accordingly. The nature of the intensity of rainfall on 31st May has been regarded as the initial state of the chain under study. Let the cell frequency of unit step transition $i \rightarrow j$. i.e. the process can go from i^{th} to j^{th} state of occurrence and is denoted by n_{ij}, where i and j belongs to s; s = 1, 2,….m (5) and also s is a closed set of five states. The set of m^2 (25) transition frequencies have sufficient statistics for the transition matrix. The transition probabilities have been estimated by applying the maximum likelihood method.

The estimate of the conditional probability is given by

$p_{ij} = p[\, x_n = j / x_{n-1} \,]$

where x_{n+1} be the realization and i and

$j \in S: S = 1 \ldots\ldots\ldots m.$

The initial probability

$p_u^{\,0} = P[x_0 = u] \quad : \quad u \in S.$

But it has been ignored from the chain. The transition probabilities are calculated by dividing the transition frequencies by the total frequencies of that particular state and is given by

$$p_{ij} = \frac{n_{ij}}{n_i}$$

where $n_i = \sum_j n_{ij}$. Subject to the restrictions:

$$\sum_j p_{ij} = 1 \quad \text{for each i and} \quad P_{ij \geq 0} \quad \text{for all i and j}$$

Each row constitutes the probability vector and it is convenient to give the set of all such vectors as a matrix. This matrix is known as transition probability matrix or stochastic matrix. This is a square matrix (5x5) with non negative elements and unit row sum.

The same method of estimation has been also employed to estimate the stationary probability of each state and is given by

$$\pi_i = \frac{n_i}{n}$$

where $n = \sum_i n_i$ subject to condition.

$$\sum_i \pi_i = 1 \qquad i \in S.$$

The method of estimation of transition probability and stationary probability has been discussed by Bhattacharyya and Waymire (1990).

Redundancy test

The mathematical model of information theoretic entropy given by Shannon (1948) have been employed to measure the uncertainty of the transition probability and is given by

$$H_i = - \sum p_{ij} \log p_{ij} \text{ for each i}$$

where H_i denotes the entropy of i^{th} state. The entropy for each state of transition matrix is to be measured separately for the month of June, July, August and September as well as for the monsoon season at all the stations.

The entropy of stationary distribution i.e. the entropy of the individual state of occurrence is given by

$$H_{11} = -\sum_i \pi_i \log \pi_i$$

where π_i is the unconditional probability of the i^{th} state of occurrence. The entropy has an important use in measuring the uncertainty as well as in testing the hypothesis of Markov dependence. The calculation of the second entropy is based upon the weighted average of the entropies for each transition state. The weighted entropy is given by

$$H = -\sum_i \pi_i H_i$$

where, π_i is the probability of individual state of occurrence. i.e. the i^{th} state and H_i is the entropy of the i^{th} state.

A sequence of uncertainty, M, of the stationary model is obtained from the individual states of occurrence over the Markovian model in the system and is given by

$$M = H_{11} - H = (-\sum_i \pi_i \log \pi_i) - \sum_i \pi_i H_i$$

The redundancy of the state of occurrence, R, is obtained as the difference from one of the ratio of the weighted entropy value H to the maximum possible entropy (H_{max}).

Here, H_{max} = log 5, as we have only five state of occurrence.

So,

$$R = 1 - \frac{H}{H_{max}}$$

This redundancy value is used to determine the favorableness unfavorableness of the Markovian system. As the redundancy value, R, tends to 1, the Markovian system tends to maximum favorable condition i.e. almost certain. Now on the light of this argument, we may examine the Markovian dependency on the monsoon months and season.

Likelihood ratio test by entropy

In this section we use the informational measure to test the hypothesis of Markovian dependence. The mathematical model of Shannon is used to obtain the measure of entropy in individual state of occurrence denoted by H_{11}. The average conditional uncertainty can be measured by

$$H_{21} = - \sum_i \pi_i P_{ij} \log P_{ij.}$$

This is the same as the weighted entropy.

The hypothesis testing, involving Markov chain has been considered by several ways. Mainly the Chi-square and the likelihood ratio criterion have been used for testing the hypothesis of independence of the random variable. Here we introduce a test criterion which involves the entropy but this is equivalent test of the likelihood ratio criterion. The test statistic is

$$T_1 = 2n (H_{11} - H_{21})$$
$$= 2 \sum n_{ij} \log \frac{n.n_{ij}}{n_i n_j}$$

That is T_1 is the same as the likelihood ratio criterion. The test statistic has a limiting Chi- square distribution with $(m-1)^2$ degrees of freedom and the large values of the statistic correspond to rejection of the hypothesis.

RESULTS AND DISCUSSION

From the table 1 and 2, it is found that for both the study meteorological stations during the monsoon months and over the season transition probability matrix is irreducible. All the states of the stochastic matrices for all the meteorological stations during monsoon months and seasons are observed to be transient in nature. It is also observed that the return of the state itself is uncertain as their conditional probabilities are always less than one. The transition probability moderate rainy day followed by moderate rainy days probabilities are the highest in both Dakshina Kannada and Udupi districts (expect in the June month of Udupi district, heavy rainy day followed by very heavy rainy days probability is highest among all other).

In case of Dakshina Kannada, non rainy day followed by non-rainy day highest in the month of September (0.6656), however over the season it is absent. Light rainy day followed by light rainy day is also highest in the month of September (0.4138), over the season it is 0.0000. Moderate rainy day followed by moderate rainy day highest in August (0.7011) and over the season it is 0.3824. Heavy rainy day followed by heavy rainy day is highest in the month of June (0.5000) and over the season it is 0.7052. Very heavy rainy day followed by very heavy rainy day also highest in the month of June (0.2143) and over the season 0.5382.

Table 1 : Transition probability matrix for Dakshina Kannada during monsoon months and season

JUNE						JULY					
	A	B	C	D	E		A	B	C	D	E
A	0.4904	0.2308	0.2692	0.0096	0.0000	A	0.2963	0.3333	0.3704	0.0000	0.0000
B	0.1761	0.3396	0.4340	0.0503	0.0000	B	0.0876	0.3723	0.4818	0.0584	0.0000
C	0.0382	0.1364	0.6709	0.1455	0.0091	C	0.0111	0.1171	0.6899	0.1804	0.0016
D	0.0156	0.0365	0.4167	0.5000	0.0313	D	0.0000	0.0123	0.4877	0.4631	0.0369
E	0.0714	0.0000	0.2143	0.5000	0.2143	E	0.0000	0.0000	0.1538	0.6923	0.1538
AUGUST						SEPTEMBER					
	A	B	C	D	E		A	B	C	D	E
A	0.4737	0.3158	0.1842	0.0263	0.0000	A	0.6656	0.2366	0.0978	0.0000	0.0000
B	0.1637	0.2807	0.5380	0.0175	0.0000	B	0.2226	0.4138	0.3417	0.0219	0.0000
C	0.0203	0.1502	0.7011	0.1236	0.0047	C	0.0983	0.3006	0.5618	0.0393	0.0000
D	0.0000	0.0186	0.5155	0.4534	0.0124	D	0.0000	0.2222	0.5556	0.2222	0.0000
E	0.0000	0.0000	0.1667	0.6667	0.1667	E	0.0000	0.0000	0.0000	0.0000	0.0000

MONSOON					
	A	B	C	D	E
A	0.0000	0.0000	0.0000	0.0000	0.0000
B	0.0000	0.0000	1.0000	0.0000	0.0000
C	0.0000	0.0196	0.3824	0.5392	0.0588
D	0.0000	0.0000	0.0910	0.7052	0.2037
E	0.0000	0.0000	0.0100	0.4518	0.5382

In Udupi, non rainy day followed by non-rainy day highest in the month of September (0.7352), however over the season it is absent. Light rainy day followed by light rainy day highest in the month of July (0.4588), over the season it is absent. Moderate rainy day followed by moderate rainy day highest in August (0.7328) and over the season it is 0.3607. Heavy rainy day followed by heavy rainy day is highest in the month of june (0.5451) and over the season it is 0.6599. Very heavy rainy day followed by very heavy rainy day is also highest in the month of June (0.2174) and over the season 0.6674.

In both Dakshina Kannada and Udupi, the unconditional probability of moderate rainy day is highest followed by heavy rainy day during the June and July. For the month of August, also it is same as above for Udupi district but, for Dakshina Kannada, moderate rainy day is highest followed by light rainy day. And in both districts, for the month of September the highest is moderate rainy day followed by light rainy day. For season heavy rainy day is highest followed by very heavy rainy day (Table 1 and 2).

From the table 3 and 4, *i.e.*, in Dakshina Kannada and Udapi districts, it is found that, the unconditional probability of moderate rainy day is highest followed by heavy rainy day during the monsoon months. And for the season, heavy rainy day is highest followed by very heavy rainy day.

During the monsoon season (June – September), In both districts, the degree of uncertainty is highest in Moderate rainy day than the other state. Considering the values of the entropy of the stationary distribution of the study districts, it is necessary to regard the occurrence of rainfall at Udapi district is lowest uncertain when compared with Dakshina Kannada. The weighted entropy values of the conditional probabilities during the monsoon period is also lowest uncertain at Udupi (Table 5 and 6).

From the table 7 it is found that, the calculated value of likelihood ratio test by entropy (T_1) is significant at both 5% and 1% level of significance. Thus, it indicates rejection of null hypothesis of independence against Markovian dependence. Therefore, at the statistical point of view, Markov Chain models have been fitted significantly on the occurrence of daily rainfall with its intensity may depend on the state of occurrence of rainfall on the previous day only.

Considering the monsoon months in all the meteorological stations separately it found that in both

Table 2 : Transition probability matrix for Udupi during monsoon months and season

	JUNE					JULY				
	A	**B**	**C**	**D**	**E**	**A**	**B**	**C**	**D**	**E**
A	0.5270	0.3378	0.1081	0.0270	0.0000	0.5500	0.1000	0.3500	0.0000	0.0000
B	0.1216	0.3108	0.5000	0.0608	0.0068	0.0471	0.4588	0.4706	0.0235	0.0000
C	0.0241	0.1509	0.6117	0.2012	0.0121	0.0064	0.0668	0.7234	0.1971	0.0064
D	0.0144	0.0072	0.3935	0.5451	0.0397	0.0034	0.0069	0.4241	0.4931	0.0724
E	0.0435	0.0435	0.0435	0.6522	0.2174	0.0000	0.0000	0.1724	0.6897	0.1379

	AUGUST					SEPTEMBER				
	A	**B**	**C**	**D**	**E**	**A**	**B**	**C**	**D**	**E**
A	0.5490	0.2941	0.1176	0.0392	0.0000	0.7352	0.1994	0.0592	0.0062	0.0000
B	0.1250	0.3529	0.4706	0.0441	0.0074	0.2518	0.4137	0.3129	0.0216	0.0000
C	0.0110	0.1102	0.7323	0.1433	0.0031	0.0401	0.2620	0.6471	0.0508	0.0000
D	0.0000	0.0136	0.4455	0.5136	0.0273	0.0000	0.0217	0.5652	0.4130	0.0000
E	0.0000	0.0000	0.0909	0.7273	0.1818	0.0000	0.0000	0.0000	0.0000	0.0000

	MONSOON				
	A	**B**	**C**	**D**	**E**
A	0.0000	0.0000	0.0000	0.0000	0.0000
B	0.0000	0.0000	0.0000	0.0000	0.0000
C	0.0000	0.0000	0.3607	0.5738	0.0656
D	0.0000	0.0000	0.0688	0.6599	0.2714
E	0.0000	0.0000	0.0066	0.3260	0.6674

Table 3: Stationary probability of monsoon months and season for Dakshina Kannada

	June	July	Aug	Sep	Season
Pi1	0.102	0.026	0.074	0.311	0.000
Pi2	0.157	0.130	0.163	0.314	0.002
Pi3	0.538	0.602	0.605	0.348	0.098
Pi4	0.188	0.231	0.153	0.026	0.616
Pi5	0.014	0.011	0.006	0.000	0.284

Table 4: Stationary probability of monsoon months and season for Udupi

	June	July	Aug	Sep	Season
Pi1	0.073	0.019	0.051	0.315	0.000
Pi2	0.146	0.081	0.130	0.273	0.000
Pi3	0.487	0.599	0.601	0.367	0.059
Pi4	0.272	0.274	0.208	0.045	0.512
Pi5	0.023	0.027	0.010	0.000	0.429

Stations across the monsoon months likelihood ratio test is significant at 5 % and 1% level of significance. Thus, the likelihood ratio test suggests that the weather of a day is influenced by the immediately preceding day's weather for all the districts for monsoon months and season.

In contexts of redundancy value, Markovian favorableness is established up to the value of 0.46, which is very low compared to the value of 1.0. Therefore, it is observed that the redundancy test is not so powerful test of Markov dependence. However, a test equivalent to the likelihood ratio test has shown better result against, the Markov dependence on the same observations. But considering the empirical result of Medhi (1976), it would have been suggested that the likelihood ratio test is more powerful as well as more appropriate test against the Markov dependence.

Markov chain model stands out as one of the best tool among the many available tools to study the distribution of rainfall and its characteristics,. Also, it is important to study the uncertainty of rainfall. Using redundancy test to estimate entropy values for monsoon rainfall in costal zone of Karnataka, uncertainty is calculated. From results, redundancy test used to estimate

Table 5: Measure of entropy for Dakshina Kannada during monsoon month and season

	June	July	Aug	Sep	Season
H1	0.4715	0.4753	0.4887	0.3645	0.0000
H2	0.5148	0.4772	0.4592	0.4996	0.0000
H3	0.4288	0.3806	0.3894	0.4519	0.4101
H4	0.4366	0.3833	0.3600	0.4321	0.3425
H5	0.5191	0.3607	0.3768	0.0000	0.3206
H11	0.5346	0.4576	0.4822	0.5165	0.3891
H22	0.4490	0.3960	0.4039	0.4387	0.3422
M	0.0857	0.0616	0.0783	0.0777	0.0469
R	0.3577	0.4335	0.4221	0.3723	0.5104

Table 6: Measure of entropy for Udupi during monsoon month and season

	June	July	Aug	Sep	Season
H1	0.4527	0.4024	0.4638	0.3243	0.0000
H2	0.5081	0.4101	0.5021	0.5032	0.0000
H3	0.4568	0.3472	0.3550	0.3965	0.3758
H4	0.4007	0.4154	0.3732	0.3348	0.3528
H5	0.4428	0.3616	0.3299	0.0000	0.2903
H11	0.5486	0.4508	0.4758	0.5323	0.3790
H22	0.4489	0.3724	0.3832	0.4001	0.3273
M	0.0997	0.0785	0.0926	0.1322	0.0517
R	0.3578	0.4672	0.4517	0.4276	0.5317

Notes: *Hi : Information theoretic entropy, H11: Entropy of stationary distribution or individual state of occurrence, M: Sequence of uncertainty of stationary model (difference in uncertainty of Markovian and random model) = H11-H, R: Redundancy value, H22= Weighted entropy. (To test the hypothesis of Markovian dependence)*

entropy values, which ranged from 0.35 to 0.46 shows that monsoon rainfall in coastal zone of Karnataka, is highly uncertain. Likelihood ratio test is more powerful as well as more appropriate in testing the Markov dependence. The results during this study period from June to September showed that, likelihood ratio test are more significant, depicting that system to follow Markovain dependence.

REFERENCES

Basu, G.C. 1988. A study of monsoon daily rainfall at maithan by markovian model and information theory. *Mausam,* **39**: 83- 86.

Bhattacharyay, R.N. and Waymire, E.C. 1990. *Stochastic Process with Application*. John Wiley and Sons, New York.

Kawachi, T., Maruyama, T.and Singh, V.P., 2001. Rainfall entropy for delineation of water resources zones in Japan. *J. Hydrol*., **246**: 36–44.

Sathish, G. and Bhattacharyya, B. 2017. Active monsoon and daily rainfall in the hill sub region of West Bengal – an information theory approach. *Int. J. Agri Sci. Res.,* **7**: 497-504.

Shannon, C.E. 1948. A note on the concept of entropy. *Bell System Tech. J*., **27**: 379-423.