Determination of the Order of a Markov Chain for Daily Rainfall Data of North East India: "Application AIC criterion"

Surobhi Deka

Cotton University, Panbazar-781001, Guwahati (Assam), INDIA
E-mail: surobhi1965@gmail.com

Abstract: The paper aims at demonstrating the application of the Akaike information criterion to determine the order of two state Markov chain for studying the pattern of occurrence of wet and dry days during the rainy season (April to September) in North-East India. For each station, each day is classified as dry day if the amount of rainfall is less than 3 mm and wet day if the amount of rainfall is greater than or equal to 3 mm. We apply Markov chain of order up to three to the sequences of wet and dry days observed at seven distantly located stations in North East region of India. The Markov chain model of appropriate order for analyzing wet and dry days is determined. This is done using the Akaike Information Criterion (AIC) by checking the minimum of AIC estimate. Markov chain of order one is found to be superior to the majority of the stations in comparison to the other order Markov chains. More precisely, first order Markov chain model is an adequate model for the stations North Bank, Tocklai, Silcoorie, Mohanbari and Guwahati. Further, it is observed that second order and third order Markov chains are competing with first order in the stations Cherrapunji and Imphal, respectively. A fore-knowledge of rainfall pattern is of immense help not only to farmers, but also to the authorities concerned with planning of irrigation schemes. The outcomes are useful for taking decisions well in advance for transplanting of rice as well as for other input management and farm activities during different stages of the crop growing season.

Keywords: Akaike Information Criterion, Markov chain, North East India, Rainfall

INTRODUCTION

Amount of daily rainfall is an important factor that impacts agriculture system. It governs the crop yields and determines the choice of the crops that can be grown. In India majority of the people depends on agriculture for their livelihood. Around 60–70% of Indian population (directly or indirectly) depends upon agriculture sector and currently it contributes to 16–17% of the GDP. Rice, being the highest produced cereal in India, is the most important staple food for the major proportion of the population in the country. Only 45% of gross cropped area of the country, by the year 2010–2011, was irrigated, out of which the contribution from rice was only 13.5% (Ghosh *et al.*, 2015), which restricted cultivation of rice as a rain-fed crop mostly to the kharif.

North East India, located at east of $80°$E and North of $21°$N, is one of the major disaster prone region of India because of their unique geographical locations and physical features, witnessing the fury of monsoon. Geographically, two-thirds of the area of NE-Indian region is a hilly terrain interspersed with valleys and plains. Apparently from the geographical features, the rain producing mechanism of NE rainfall is a complex phenomenon and is largely dominated by an elevated

orography of the Eastern Himalayas and large forests. In the North East India, the rainfall distribution is not even. While the excess rainfall in the monsoon months of June-September causes natural disaster, in the longer dry spell during November to March crop goes down in spite of having sufficient rainfall in the monsoons. In general, rainfall distribution pattern over different regions of India is inhomogeneous due to influence of several local and remote factors. For instance, over the central Indian plain region, monsoon trough and the Himalayas along with local and external factors play a dominant role in its interannual rainfall variability. Recently, summer monsoon rainfall variability over NE- region has been studied by Prabhu *et. al.* (2016). The average annual rainfall in North East India ranges from 2000-4000 mm with a maximum of 11000 mm in Cherrapunjee (Deka *et al.,* 2011, Prabhu *et al.*, 2016). The summer monsoon influence this region from June to September contributing more than 80% of the annual rainfall (see, Deka (2013). During this season major floods occur that often lead to disaster (ASDMA, 2015)

According to Fisher (1925) crop yield during a season mainly influenced by the distribution of rainfall rather than season total amount rainfall. Thus, rainfall forecasts at different temporal scales, at the beginning of

the season and within the season, with possible impact on the crops in terms of productivity, may help the farming community to take appropriate decisions on input management and farm activities well in advance (Ghosh et. al., 2015). Extreme precipitation events (heavy rain storm, cloud burst) may have their own impacts on the fragile geomorphology of the Himalayan part of the Brahmaputra basin causing more widespread landslides and soil erosion. The response of hydrologic systems, erosion processes and sedimentation in the Himalayan river basins could alter significantly due to climate change. Two extremely intense cloud bursts of unprecedented intensity- one in the western Meghalaya hills and western Arunachal Pradesh in 2004 produced two devastating flash floods in the Goalpara and Sonitpur districts of Assam bordering Meghalaya and Arunachal respectively causing hundreds of deaths. The most recent examples of such flash floods originating from extreme rainfall are two events that occurred in the north bank of the Brahmaputra river and caused significant damage to human life and property. The first of the two events occurred during the monsoon season on June 14th, 2008 due to heavy rainfall on the hills of Arunachal Pradesh north of Lakhimpur District causing flash floods in the rivers of Ranganadi, Singara, Dikrong and Kakoi that killed at least 20 people and inundated more than 50 villages leading to displacement of more than 10,000 people. The other that occurred in the post monsoon season on October 26 affected a long strip of area of northern Assam valley adjoining foothills of Bhutan and Arunachal Pradesh causing flash flooding in four major rivers (all are tributaries of the river Brahmaputra) and a number of smaller rivers. This episode of flash floods caused by heavy downpour originated from the Tropical Depression 'Rashmi' (a depression over the West Central Bay of Bengal adjoining Andhra coast). For details, we refer to Deka *et al.* (2011). Hence, stochastic modeling of rainfall pattern is of immense help not only to farmers, but also to the authorities concerned with planning of irrigation schemes and disaster management.

Stochastic modeling of rainfall data has become a frontier research area over the years. As far our knowledge is concerned, the statistical modeling of rainfall data started with the work of Gabriel and Neumann (1962, 1972). They applied a first order chain to Tel Aviv precipitation data (November-June, 1973-1975) on the basis of multiple hypotheses testing procedures and it is observed from their study that two state Markov chain give a good description of the occurrences of wet and dry days during the rainy period at Tel Aviv. Bhargava *et al.* (1972) studied the occurrence of rainfall with the help of Markov chain model of order one in Raipur District India. The number of years for which the data were available varied from station to station, the variation being from 39 to 60 years. A sequence of

wet and dry days for each center and for each year during the period 1st June to 30th September was obtained. Markov chain model was fitted for each center and the expected number of dry days varies from 70 to 78 whereas the number of wet days varies from 44 to 52 days. Latter, Gates and Tong (1976) reanalyzed the same Tel Aviv data applying the AIC procedure and suggested that a Markov chain of order not lower than 2 should be fitted, instead of the previously fitted first order. For the application of higher order Markov chain, we refer to Katz (1981) and Zhao *et al.* (2001). Although there is a disagreement on the appropriate order for the Tel Aviv model but one must agree that Markov chains are obvious candidate to model the occurrence of rainfall. For instance, we refer to Raheem et. al. (2015) and references therein. A three-state Markov chain was employed to examine the pattern and distribution of daily rainfall in Uyo metropolis of Nigeria using 15 years (1995-2009) rainfall data. However, multi state higher order Markov chain modeling of rainfall data is an important task in its own right and will be considered elsewhere. For other scientific works on Markov chain modeling to daily rainfall data in eastern part of India, we refer to Dash (2012) and Basak (2014). Recently, Singh *et al.* (2014) has carried out a study to analyze the rainfall data for storage and irrigation planning under humid southeastern region of Rajasthan using a time series record for 32-year (1980-2011) periods. Monthly estimation of rainfall data for assessment of normal years, abnormal years, drought years, normal months, abnormal months, and drought months has been made which is expected to be useful from the view point of cropping pattern, irrigation planning and management. More recently, Ghosh *et al.* (2015) has developed an experimental extended range forecast system (ERFS) in crop models is investigated for improving the rice grain yield prediction skill for the ensuing monsoon season in the experimental station at Bhubaneswar, India. Such methods will provide valuable information well in advance, to overcome uncertainties of monsoon rainfall in respect of onset of monsoon, long dry spells causing soil moisture stress or drought and extremely heavy rainfall leading to floods.

A good number of literatures are available describing the Markov chain model for daily precipitation round the globe, but, no rigorous work barring the works by Medhi (1976) pursued in North East region of India. The present study is an effort to demonstrate the application of higher order two state Markov chain. More precisely, an analysis regarding the fitting of Markov chain of appropriate order has been made in this article. We apply Markov chain of order up to three to the sequences of wet and dry days observed at seven distantly located stations in North East region of India. For the majority of the stations Markov chain of order one is identified as the most appropriate model, fol-

lowed by order two, for describing the daily precipitations occurrences over North East India during Indian summer monsoon season. If the best fitting stochastic model is known for a particular station, it can be used in the future for an estimation of the rainfall pattern in the area under study. Rainfall information forms the basis for designing water related structure in agriculture planning, in water management and also in monitoring climate changes.

MATERIALS AND METHODS

In this exposition, series of daily rainfall data of seven stations in North East India viz. Imphal (2001-2005), Mohanbari (1993-2006), Guwahati (2001-2005), Cherrapunji (2001-2005), Silcoorie (1986-2005), North Bank (1986-2005), Tocklai (1986-2005) have been selected. The locations of these seven stations of North East India are shown in Fig. 1 (Source: Deka *et al.* (2011)). The series of daily rainfall data are taken from Regional Meteorological Centre, Guwahati and Tocklai Experimental Station, Jorhat involving the aforesaid seven stations for the summer season (April to September) in each year.

The Akaike information criterion was introduced by Akaike (1972) as an extension to final prediction error. It *is* a way to compare different models on a given outcome. AIC has been used successfully in various fields of statistics (Akaike, 1972, 1976), Otsu *et al.* (1976), engineering (Otomo *et al.*, 1972), hydrology (Salas *et al.*, 1980) and numerical analysis (Sakamoto *et al.* (1978)). For recent application of Akaike information criterion, we refer to Snipes and Taylor (2014). The procedure for the determination of the order of a Markov chain by Akaike Information Criterion (AIC) was developed by Tong (1975). In the present work, application of the Akaike information criterion is demonstrated to determine the order of two state Markov chain for studying the pattern of occurrence of wet and dry days during the rainy seasons in North-East India.

In statistical inference situations, Akaike (1972) proposed the use of the entropy B $[\bullet]$ given by:

$$B[f;z] = \int \log\left\{\frac{g(z;x)}{f(z)}\right\} f(z)dz . \quad (1)$$

Here, x is the vector of observations, and $f(z)$ and $g(z;x)$ are the probability density functions of the true and fitted models, respectively. According to the entropy maximization principle, Akaike (1972), the objects of statistical inference are to estimate $f(z)$ from the data x and to try to find $g(z;x)$ which maximizes the expected entropy:
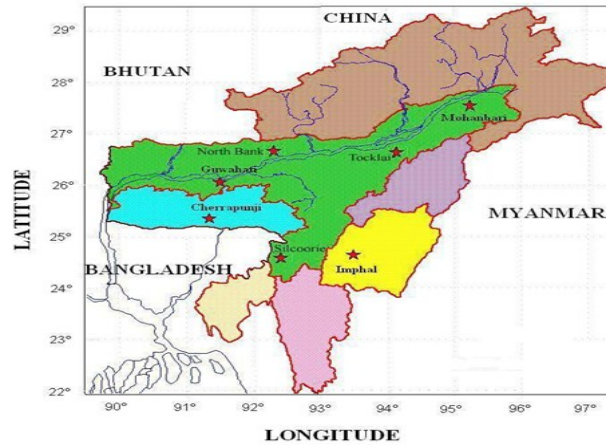
$$E\{B[f;g]\} = \int B[f;g]f(x)dx \quad (2)$$

Fig. 1. *Geographical locations of the stations considered for this study.*

Where E denotes the expectation operator and x is the vector of observations. Akaike 1972) showed that for the number of observations $n \geq 30$:

$$-2nE\{B[f;g]\} \approx \eta + 2k - L \quad (3)$$

η is a log-likelihood ratio test function given by:

$$\eta = -2\sum_{i=1}^{n} \log\left\{\frac{g(x_i /_k \hat{\theta})}{g(x_i /_L \hat{\theta})}\right\} \quad (4)$$

with

L = number of parameters (dimension) of the true model,

k = number of parameters (dimension) of the fitted model,

$_k\hat{\theta}$ and $_L\hat{\theta}$ are the estimated parameters of the fitted and true models respectively.

Thus, from equation (3) and by ignoring the constant terms, Akaike derived a criterion which is now called the Akaike Information Criterion (AIC) given by:

$$AIC = -2\sum_{i=1}^{n} \log\{g(x_i /_k \hat{\theta})\} + 2k. \quad (5)$$

The above relation can also be written as:

AIC= (-2) $\times$ log (Maximum-likelihood for model) + 2 $\times$ (Number of independent parameters in the model). (6)

This statistics was introduced as a measure of the deviation of the fitting model from the true structure. The first term on the right hand side of the above equation (6) is a measure of the lack-of-fit of the chosen model, while the second term measures the increased unreliability of the chosen model due to the increased number of model parameters. The best approximating model is the one which achieves the minimum AIC in the class of the competing models. The procedure which, given

several models, adopts the model that minimizes the AIC is called the minimum AIC estimate (MAICE). It is important to note that, since the AIC test is based on the maximum likelihood function, which is asymptotically effective and unbiased, the test yields fairly accurate results for $n \geq$ 30, where n is the number of observations. However, the test has been used with considerable success for $n \geq$ 20 (Kitagawa, 1979).

Denote the transition probability for a r-order chain by $p_{ij...kl}, i = 1,2,...s,$ s being the finite number of states of the chain and the suffix contain r+1 characters. Then the maximum likelihood estimates of $p_{ij...kl}$ is given by

$$\hat{p}_{ij...kl} = \frac{n_{ij...kl}}{n_{ij...k\bullet}} . \tag{7}$$

Where $n_{ij...kl}$ is the number of transition from the state i to the state l through the states j…k and

$$n_{ij...k\bullet} = \sum_l n_{ij...kl} .$$

The hypothesis tested is

$$H_{r-1} : p_{ij...kl} = p_{j...kl}, i = 1,2,...s$$

(that the chain (r-1) dependent against $H_r :$ that the chain is r dependent). The statistics constructed is

$$_{r-1}\eta_r = -2\log\lambda_{r-1,r} = 2\sum_{i,j,...k,l} n_{ij...kl}\left(\ln\frac{n_{ij...kl}}{n_{ij...k\bullet}} - \ln\frac{n_{\bullet j...kl}}{n_{\bullet j...k\bullet}}\right)$$

It is a χ^2 variable with $s^{r-1}(s-1)^2$ degrees of freedom under H_{r-1}. The hypothesis H_k that the chain is k dependent implies the hypothesis H_r that the chain is r dependent, whenever k<r. A r-order Markov chain means the dependence of the weather of a day on the weather conditions of the immediately preceding r days (see, Gates and Tong (1976)). Hence, the hypothesis H_k is a subset of the hypothesis H_r. Denote by $\lambda_{k,r}$, the ratio of the maximum likelihood given H_k to that given H_r, then

$$\lambda_{k,r} = \lambda_{k,k+1}\lambda_{k+1,k+2}...\lambda_{r-1,r}$$

and so

$$_k\eta_r = -2\log\lambda_{k,k+1} - 2\log\lambda_{k+1,k+2} - ... - 2\log\lambda_{r-1,r}, \quad k < r$$

Again, under H_k, Good (1953) has shown that $-2\log\lambda_{k,r}$, i.e., $_k\eta_r$ has a χ^2 variable with degrees of freedom

$$\nabla s^{r+1} - \nabla s^{k+1} \qquad k \geq 0,$$
$$\nabla s^{r+1}, \qquad k = -1.$$

Here, ∇ denotes the standard backward operator given by $\nabla s^r = s^r - s^{r-1}$.

If the statistical identification procedure is considered as a decision procedure, the most basic problem is the appropriate choice of the risk (expected loss) function. The loss functions considered in classical theory of hypothesis testing are defined by the probabilities of accepting the incorrect hypothesis or rejecting the correct hypothesis. Interested readers are referred to Medhi (1999) for more details elaboration on loss function and hypothesis testing.

Tong (1975) proposes the choice of the loss function, based on AIC approach as

$$R(k)= {}_k\eta_M - 2(\nabla S^{M+1} - \nabla S^{k+1}). \tag{8}$$

Where M is the highest order model to be considered and k is the order of the fitting model. The minimum AIC estimate (MAICE) of the order of the Markov chain is that value of k which gives the minimum of R (k) over all orders considered. Raising the order of Markov chain does not necessarily do away the imperfections of the model. On the other hand, the number of parameters to estimate increases with 2^k for two state, k order Markov chain which may rapidly enhance the uncertainty of the estimation. Therefore the present study is confined to the Markov chain of order up to three.

RESULTS AND DISCUSSION

Nature is a very complicated system. Despite this, it is often necessary to simulate the behavior of natural systems. To attempt this by modeling all the physical processes deterministically is a very difficult task. Instead stochastic models are used. Stochastic models are designed to reproduce the important patterns evident in the observations based on the current knowledge of the physical processes. Realistic sequences of meteorological variables such as precipitation are key inputs in many hydrologic, ecologic and agricultural models. Simulation models are needed to model stochastic behavior of climate system when historical records are of insufficient duration or inadequate spatial and /or temporal coverage. It is well known fact that Markov Chain model can be fitted to daily rainfall occurrence and several authors have used Markov chain model to estimate the wet and dry days in past, see Bhargava *et al.* (1972), Gabriel and Neumann (1962, 1972), Gore

Table 1. Likelihood statistic for North Bank.

Year	$_0\eta_1$	$_0\eta_2$	$_0\eta_3$	$_1\eta_2$	$_1\eta_3$	$_2\eta_3$	$_3\eta_3$
1986	16.7276	16.8175	19.8498	0.0899	3.1222	3.0323	0
1987	17.7543	22.8454	28.7556	5.0911	11.0013	5.9102	0
1988	41.1159	42.1816	50.9963	1.0657	9.8804	8.8147	0
1989	19.1132	24.3546	32.6416	5.2414	13.5284	8.287	0
1990	21.1622	21.2464	27.8429	0.0842	6.6807	6.5965	0
1991	6.2446	12.6413	21.1247	6.3967	14.8801	8.4834	0
1992	4.6952	9.8015	11.2007	5.1063	6.5055	1.3992	0
1993	18.9827	23.6082	25.9841	4.6255	7.0014	2.3759	0
1994	20.5584	23.2673	26.4903	2.7089	5.9319	3.223	0
1995	28.3186	36.5763	44.1681	8.2577	15.8495	7.5918	0
1996	23.5193	24.5876	27.8354	1.0683	4.3161	3.2478	0
1997	17.4515	19.047	26.7143	1.5955	9.2628	7.6673	0
1998	42.0208	45.0825	50.9531	3.0617	8.9323	5.8706	0
1999	12.1565	16.5651	20.6124	4.4086	8.4559	4.0473	0
2000	4.8409	8.9134	15.9362	4.0725	11.0953	7.0228	0
2001	6.8661	7.3759	9.4704	0.5098	2.6043	2.0945	0
2002	22.3583	22.542	26.9619	0.1837	4.6036	4.4199	0
2003	17.034	19.485	23.7815	2.451	6.7475	4.2965	0
2004	25.857	29.9857	30.2572	4.1287	4.4002	0.2715	0
2005	12.9343	16.5079	20.4342	3.5736	7.4999	3.9263	0
df	1	3	7	2	6	4	0

Table 2. AIC values for the station Cherrapunji.

Year	R(0)	R(1)	R(2)	R(3)	Min R(i)	order
2001	25.296	-5.5744	-4.7736	0	-5.5744	1
2002	16.5769	1.8441	1.4442	0	0	3
2003	16.7548	-0.0637	-1.2467	0	-1.2467	2
2004	0.7354	-4.4915	-0.9573	0	-4.4915	1
2005	21.1522	-2.5002	-3.3747	0	-3.3747	2

Table 3. AIC values for the station Guwahati.

Year	R(0)	R(1)	R(2)	R(3)	Min R(i)	order
2001	-1.0056	-5.0543	-4.2653	0	-5.0543	1
2002	10.4165	-4.4754	-1.8223	0	-4.4754	1
2003	17.846	11.4287	13.3948	0	0	3
2004	-1.36577	-5.60417	-1.8506	0	-5.60417	1
2005	18.7865	7.188	2.8948	0	0	3

Table 4. AIC values for the station Imphal.

Year	R(0)	R(1)	R(2)	R(3)	Min R(i)	order
2001	4.2847	-6.3741	-2.5225	0	-6.3741	1
2002	32.9413	8.7132	-5.0412	0	-5.0412	2
2003	16.9987	2.6758	0.1431	0	0	3
2004	9.5471	1.3503	2.0861	0	0	3
2005	12.7122	-4.5871	-2.0123	0	-4.5871	1

Table 5. AIC values for the station Mohanbari.

Year	R(0)	R(1)	R(2)	R(3)	Min R(i)	order
1993	6.3891	-3.1913	-3.0882	0	-3.1913	1
1994	5.4926	-6.3	-5.2347	0	-6.3	1
1995	7.322	-1.9796	-0.0287	0	-1.9796	1
1996	13.724	-7.2275	-6.3125	0	-7.2275	1
1997	7.6405	-5.1164	-1.9863	0	-5.1164	1
1999	5.5773	2.508	-4.5951	0	-4.5951	2
2001	-2.6165	-10.0504	-7.9391	0	-10.0504	1
2002	7.6799	-10.6054	-6.9666	0	-10.6054	1
2003	8.2287	-0.0316	-5.7523	0	-5.7523	2
2004	15.5802	0.3969	1.9405	0	0	3
2005	7.4747	1.6295	4.336	0	0	3
2006	17.0646	-0.4816	-0.8071	0	-0.8071	2

and Thapliyal (2000) for first order Markov chain model and Gates and Tong (1976), Katz (1981), Zhao *et al.* (2001), Sørup (2011), Dash (2012), Basak (2014) etc. for higher order Markov chain model.

In this study we concentrate on higher order two states Markov chain model over a series of daily rainfall data of seven stations in North East India. The determination of the proper order that best describes the precipitation over the region is carried out using Akaike Information Criteria. Table 1 illustrates the estimates of likelihood ratio statistic for station North Bank. It is interesting to see that the columns corresponding to

$_0\eta_1$, $_0\eta_2$, $_0\eta_3$ are significant at 5% level of significance except for the years 1992 and 2001. Therefore, we may note that the chain is at least of order one. For the simplicity of the exposition other tables concerning likelihood ratio statistic for other stations are not included. Then R(k) values for each station over each year are calculated. The calculated

values are displayed in Tables 2-8. Then according to MAICE procedure we adopt as the order that value of k which gives minimum R(k) and those values for k are illustrated in column 7 of the Tables 2-8. Finally, the performance of the best fitting order of the Markov chain is given in Table 9. The analysis clearly indicates that Markov chain of order one is the best one for rainfall forecasting to the majority of the stations in comparison to the other order Markov chains. More precisely, first order Markov chain model is an adequate model for the stations North Bank, Tocklai, Silcoorie, Mohanbari and Guwahati. This observation agrees with the findings of Dash (2012) and Basak

Table 6. AIC values for the station Northbank.

Year	R(0)	R(1)	R(2)	R(3)	Min R(i)	order
1986	5.8498	-8.8778	-4.9677	0	-8.8778	1
1987	14.7556	-0.9987	-2.0898	0	-2.0898	2
1988	36.9963	-2.1196	0.8147	0	-2.1196	1
1989	18.6416	1.5284	0.287	0	0	3
1990	13.8429	-5.3193	-1.4035	0	-5.3193	1
1991	7.1247	2.8801	0.4834	0	0	3
1992	-2.7993	-5.4945	-6.6008	0	-6.6008	2
1993	11.9841	-4.9986	-5.6241	0	-5.6241	2
1994	12.4903	-6.0681	-4.777	0	-6.0681	1
1995	30.1681	3.8495	-0.4082	0	-0.4082	2
1996	13.8354	-7.6839	-4.7522	0	-7.6839	1
1997	12.7143	-2.7372	-0.3327	0	-2.7372	1
1998	36.9531	-3.0677	-2.1294	0	-3.0677	1
1999	6.6124	-3.5441	-3.9527	0	-3.9527	2
2000	1.9362	-0.9047	-0.9772	0	-0.9772	2
2001	-4.5296	-9.3957	-5.9055	0	-9.3957	1
2002	12.9619	-7.3964	-3.5801	0	-7.3964	1
2003	9.7815	-5.2525	-3.7035	0	-5.2525	1
2004	16.2572	-7.5998	-7.7285	0	-7.7285	1
2005	6.4342	-4.5001	-4.0737	0	-4.5001	1

Table 7. AIC values for the station Silcoorie.

Year	R(0)	R(1)	R(2)	R(3)	Min R(i)	order
1986	6.1958	-5.4116	-5.222	0	-5.4116	1
1987	12.6532	-4.9303	-5.5398	0	-5.5398	2
1988	3.3324	-7.0767	-4.8481	0	-7.0767	1
1989	14.4774	-5.8251	-5.4369	0	-5.8251	1
1990	5.8544	-8.6228	-4.7136	0	-8.6228	1
1991	0.4625	-9.8248	-7.3887	0	-9.8248	2
1992	13.3544	0.621	-7.6684	0	-7.6684	2
1993	13.737	0.3297	-3.6292	0	-3.6292	2
1994	14.1871	-4.5688	-2.558	0	-4.5688	1
1995	-1.9577	-9.4869	-6.0686	0	-9.4869	1
1996	27.6852	-1.9304	-5.0619	0	-5.0619	2
1997	5.6338	-5.7394	-4.8588	0	-5.7394	1
1999	24.1983	-3.3469	-2.7311	0	-3.3469	1
2001	7.5898	-0.8134	-4.293	0	-4.293	2
2002	11.6622	-5.0198	-7.2194	0	-7.2194	2
2003	16.485	-0.6414	-3.548	0	-3.548	2
2004	3.6647	-0.6573	1.2379	0	-0.6573	1
2005	44.7494	6.1562	9.6934	0	0	3

(2014). In the scientific work of Dash (2012), a stochastic daily precipitation generation model was adapted for the state of Odisha. The model simulates the sequence of precipitation occurrence using the method of transitional probability matrices, while daily precipitation amount was generated using a two parameter Gamma distribution. Daily average precipitation data from Odisha from the year 2001 to 2010

Table 8. AIC values for the station Tocklai.

Year	R(0)	R(1)	R(2)	R(3)	Min R(i)	order
1986	-2.6125	-5.3556	-4.1269	0	-5.3556	1
1987	1.5623	-6.9284	-6.0647	0	-6.9284	1
1988	6.4774	-7.2628	-4.5987	0	-7.2628	1
1989	0.8393	-10.9265	-7.5985	0	-10.9265	1
1990	10.4431	6.0319	3.7292	0	0	3
1991	-2.322	-3.1392	0.179	0	-3.1392	1
1992	-4.6171	-5.4782	-3.4837	0	-5.4782	1
1993	-0.3524	-2.3872	-4.0922	0	-4.0922	2
1994	2.2017	-2.4974	-3.6537	0	-3.6537	2
1995	1.1907	-6.3817	-5.8844	0	-6.3817	1
1996	8.1645	1.0601	-4.7042	0	-4.7042	2
1997	1.08526	-5.95314	-3.40004	0	-5.95314	1
1998	4.8008	3.4529	0.8996	0	0	3
1999	5.4214	2.0093	-3.656	0	-3.656	2
2000	2.1448	-0.7117	1.128	0	-0.7117	1
2001	-5.6387	-5.6057	-6.1419	0	-6.1419	2
2002	3.7916	-4.6991	-1.9707	0	-4.6991	1
2003	-0.3613	-7.6114	-4.3962	0	-7.6114	1
2004	-5.4475	-6.7144	-7.4141	0	-7.4141	2
2005	0.2872	-2.3047	-5.778	0	-5.778	2

were used for this model. The study reveals that first order Markov chains can adequately represent the precipitation occurrences in all the months. Markov chain models for monsoonal rainfall occurrence in different zones of West Bengal is considered in Basak (2014) and first order Markov chain model is found to be the best for rainfall forecasting. Further, it is observed that second order and third order Markov chains are competing with first order in the stations Cherrapunji and Imphal, respectively. For higher order best fitting Markov chain model, we refer to Sørup (2011) and, Gates and Tong (1076).

A fore-knowledge of rainfall pattern is of immense help not only to farmers, but also to the authorities concerned with planning of irrigation schemes. The outcomes are useful for taking decisions well in advance for transplanting of rice as well as for other input management and farm activities during different stages of the crop growing season.

Conclusion

Precipitation is the unique source of fresh water which is necessary for living life. Thus, changes that take place precipitation pattern affect all ecosystems. In order to escape damages, with minimum loss, resulting from precipitation changes, making prudential plans in the light of several estimates is seen as the most rationalist way. In this context, demands for studies on precipitation model building increase at the present time.

Table 9. Percentages of the best fitting orders of Markov Chain.

Order of MC	North Bank	Tocklai	Silcoorie	Mohanbari	Cherrapunji	Guwahati	Imphal
0	0	0	0	0	0	0	0
1	60	55	50	58	40	60	40
2	30	35	44	25	40	0	20
3	10	10	6	17	20	40	40

These studies have great important not only for data production purpose but also for water resources management, hydrology and agricultural sector. Because information about the probability of occurrence of the precipitation for future can be used to make decisions relating to agricultural production planning and management and water management, it can decrease risks originating weather condition uncertainty.The Markov chain models have few advantages: firstly, forecasts are available immediately after the observations because of the use of predictors only on the local information on the weather. Secondly, the chain needs minimal computation after the climatologically data have been processed. It may also be revealed that if the record length is short, lower order chain represents the appropriate fit. Markov chains specify the state of each day as wet and dry and develop a relation between the state of the current day and the states of the preceding days. The order of the Markov chain is the number of preceding days taken into account. When literature is investigated, the common observations of these studies suggest that the occurrence of weather state is best described by first order Markov chain. Consequently, in the present study, attempts have been made to study the weather state over the region of North East India during the rainy season. The appropriate order Markov model proposed and can be used to obtain forecasts of the weather states (such as dry or wet day) at some future time using information given by current state. The present study leads to the conlusions: i) Markov chain model can be fitted to daily rainfall occurrence of North East regions of India. ii) The first order Markov chain model that has been used extensively, is an adequate model for most of the stations of North East regions of India to determine the daily precipitation. In the future, the outcomes of this work will be extended in the direction of Ghosh *et. al.* (2015) and Singh *et al.* (2014).

ACKNOWLEDGEMENTS

REFERENCES

Akaike, H. (1972). Information Theory and an Extension of the Maximum Likelihood Principle, Second International Symposium on Information Theory, Ed. B. N. Petrov & F. Csoki, Budapest: Akademiai Kiado, 267-281.

Akaike, H. (1976). Canonical Correlation Analysis of time Series and the use of an Information Criterion, Systems Identification: Advances and Case Studies, Ed. R. K. Mehra & D. G. Lainiotie, New York Academic Press, USA, 27-96.

ASDMR (2015)). Assam State Disaster Management Authority report (ASDMR) in collaboration with North Eastern Application Centre. Flood early warning sys- tem, http://www.asdma.gov.in /pdf/publication FLEWS.pdf

Basak, P. (2014). On the Markov chain models for monsoonal rainfall occurrence in different zones of West Bengal, Indian Journal of Radio & Space Physics, 43:349-354.

Bhargava, P. N., Narain, P., Aneja, K. G. and Asha, P. (1972). A study of the occurrence of rainfall in Raipur District with the help of Markov chain model, *Journal of Indian Society of Agricultural statistics*, 24:197-204.

Deka, S, Borah, M. and Kakaty, S. C. (2011). Statistical analysis of annual maximum rainfall in North-East India: an application of LH-moments, *Theoretical and Applied Climatology,* 104, 111-122.

Deka, R. L. (2013). Climate change in the Brahmaputra valley and impact on rice and tea productivity, Ph.D thesis submitted to Indian Institute of Technology Guwahati, *http://gyan.iitg.ernet.in/handle/123456789/416*.

Dash, P. R. (2012). A Markov chain modelling of daily precipitation occurrences of Odisha, Int. Jour. Adv. Comp. Math. Sci., 3:482-486.

Fisher, F.A. (1925). The influence of rainfall on the yield of wheat at Rothamsted, *Phil. Trans. Roy. Soc., London, B,* 213:89-142.

Gabriel, K. R. and Neumann, J. (1962). A Markov chain model for daily occurrence at Tel Aviv, *Quart. J. R. met. Soc.* 88:90-95.

Gabriel, K. R. and Neumann J. (1972). On a distribution of weather cycles by length, *Quart. J. R. Met. Soc.* 83:375-380.

Gates, P. and Tong, H. (1976). On Markov chain modeling to some weather data, *J. App. Meteorol.* 15:1145-1151.

Ghosh, K., Singh, A., Mohanty, U. C., Acharya, N., Pal, R. K., Singh, K. K. and Pasupalak, S. (2015). Development of a rice yield prediction system over Bhubaneswar, India: combination of extended range forecast and CERES-rice model, *Meteorol. Appl.*, 22:525–533.

Good I. J. (1953). The serial test for sampling numbers and other tests for randomness, *Proc. Camb. Phil. Soc.* 49:276-284.

Gore, P.G. and Thapliyal, V. (2000). Occurrence of dry and wet weeks over Maharastra, *Mausam*, 51:24-38

Katz, R. W. (1981). On some criteria for estimating the order of a Markov chain, *Technometrics,* 23:243-249

Kitagawa, G. (1979). On the use of AIC for the detection of outliers, *Technometrics,* 21:193-199

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

Medhi, J. (1999). Stochastic Processes, 2nd edition, New Age International Publishers, New Delhi.

Otomo, T. Nakagawa, T. and Akaike, H. (1972). Statistical approach to computer control of cement rotary kiln, *Automatica,* 8:35-48.

Otsu, K., Horigome, M. and Kitagawa, G. (1976). On the prediction and stochastic control of ship's motion, *Proc. 2nd. IFAC/IFIP Sympos.,* Ed. M. Pitkin, J. J. Roche & T. J. Williams, Washington, DC, USA, 69-76.

Prabhu, A., Oh. J., Kim, I, Kripalani, R. H., Mitra, A. K. and Pandithurai, G. (2016). Summer monsoon rainfall variability over North East regions of India and its association with Eurasian snow, Atlantic Sea Surface temperature and Arctic Oscillation, *Climate Dynamics*, DOI 10.1007/s00382-016-3445-4

Raheem, M. A., Yahya, W. B. and Obisesan, K. O. (2015). A Markov chain approach on pattern of rainfall distribution, *Journal of Environmental Statistics*, **7**:1-13.

Sakamoto, Y. and Akaike, H. (1978). Analysis of cross classified data by AIC, *Ann. Inst. Statist. Math., B* 30:85-197.

Salas, J. P., Delluer, J. W., Yevjevich, V. and Lane, W. L. (1980). Applied Modelling of Hydrologie Time Series, Water Resources Publication, Fort Collins, Colorado, USA.

Singh, B., Arya, C. K., Singh, J. and Mourya, K. K. (2014). Analysis of rainfall data for storage and irrigation planning in humid south-eastern plain of Rajasthan in India, *Journal of Applied and Natural Sciences*, 6:14-219.

Snipes, M. and Taylor, D. C. (2014). Model selection and Akaike Information Criteria: An example from wine ratings and prices, *Wine Economics and Policy*, 3:3-9.

Sørup, H. J. D., Madsen, H. and Nielsen, K. A. (2011). Markov chain modeling of precipitation time series: Modeling waiting times between tipping bucket rain gauge tips, paper presented at 12th International Conference on Urban Drainage, Porto Alegre/Brazil, 11-16.

Tong, H. (1975). Determination of the order of a Markov chain by Akaike's Information Criterion, *Journal of Applied Probability* 12:488-497.

Zhao, L. C., Dorea, C. and Goncalves, C. R. (2001). On determination of the order of a Markov chain, *Statist. Inferen. Stoch. Proc*. 4:273-282.