IOSR Journal of Engineering (IOSRJEN) www.iosrjen.org
ISSN (e): 2250-3021, ISSN (p): 2278-8719
Vol. 10, Issue 2, February 2020, ||Series -II|| PP 07-16

Prediction of Daily and Yearly Rainfall of Kolar Region of Karnataka State By Markov Chain and Regression Model

Mohammed A Rasiq[1], Konanna Gari Jaswanth Reddy[2], Dr. K M Roopa[3], V. K. Srinivasu[4]

[1, 2]Students, Department of Civil Engineering, Bangalore Institute of Technology, Bangalore, India.
[3] Professor and Head, Department of Mathematics, Bangalore Institute of Technology, Bangalore, India.
[4] Scientific Officer, Karnataka State and Technology Academy, Bangalore, India.
Received 06 February 2020; Accepted 20 February 2020

Abstract: This paper explains the prediction of probabilities of rainfall by stochastic process and magnitude of rainfall using Linear Regression Models. Using Markov chain techniques, a two state and four state first order transition probability matrices were developed for the daily and yearly analysis respectively. For daily rainfall analysis, the transition states classified as rainy and non-rainy days. In yearly rainfall analysis, the transition sequences were established based on rainfall intensities. For long term rainfall analysis, limiting probabilities were calculated with steady state transition matrix and initial state probability matrix. Linear regression models are used for prediction of the magnitude of annual rainfall for 5 years, with the help of previous 15 years rainfall data and the predicted rainfall is compared to actual rainfall data. The average error is estimated by calculating the difference in the actual and predicted rainfall values. All the models were developed using the previous 30 years (1989-2018) rainfall of Kolar region.

Key Words: Markov chain, stochastic process, Transition Probabilities, Initial State Probabilities, Linear Regression, Coefficient of correlation and Coefficient of determination.

I. INTRODUCTION

Water is essential for every form of life, for all aspects of socio-economic development such as agricultural, industrial development and for the maintenance of healthy ecosystems. Approximately 70% of the freshwater used by humans goes to agriculture. The primary source of water in every area is “RAIN”. Agriculture in India is majorly dependent on rainfall. Even though water to the agriculture can be supplied from irrigation, it is necessary to store the water for irrigation supply. Ultimately, we need proper rainfall to serve the purpose. Thereby the yield of the crops depends on rainfall analysis. So it is significant for us to predict the rainfall patterns and its occurrence with the help of previous records of data available. Rainfall is a meteorological phenomenon that has most impact on humans and most important environmental factor limiting the development of semi-arid regions. Understanding rainfall variability is important to optically manage the scarce water resources that are under continuous stress due to increasing water demands, increase in population and economic development. There are many aspects of water resource management including the optimal water allocation, quality assessment and preservation and prediction of future water demands to strategize water utilization, planning and decision making. Hence, it is very much important to use the established methods to carry out this assessment to analyze the rainfall is through probabilistic process using Markov Chains [1] [3] and linear Regression model.

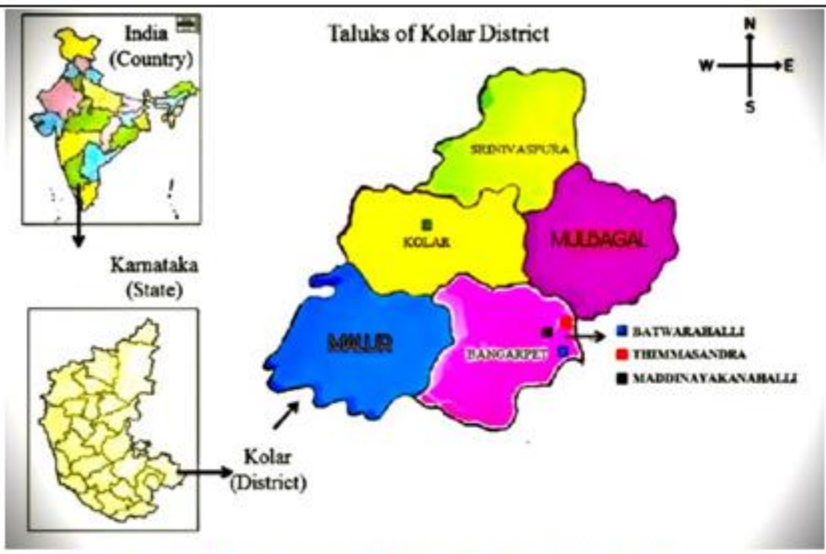

Fig –1: Location and details of Kolar district.

II. AREA OF STUDY

Kolar district is the eastern gateway to Karnataka. It is land locked district and hard rock terrain of Karnataka in the maiden (plain) region and covers an area of 8223 sq.km. Kolar district lies between North latitude 12˚ 46’ to 13˚ 58’ and East longitude 77˚ 21’ to 78˚ 35’.The Kolar subdivision consists of Kolar, Bangarpet, Malur, Mulbagal and Srinivaspur taluks.The area irrigated by wells constitutes 99% of the total irrigated area. Dug well irrigation practice is largely replaced by bore-well irrigation. Irrigation is being practiced both in the valley as well as in upland areas.

Kolar district falls in the Eastern dry agro climatic Zone. It experiences a semi-arid climate, characterized by typical monsoon tropical weather with hot summers and mild winters. The year is normally divided into four seasons, dry season during Jan-Feb, Pre-monsoon season during Mar-May, Southwest Monsoon season during Jun-Sep and Northeast monsoon season during Oct-Dec. Based on rainfall data pertaining to there are 11 rain gauge stations in each of the 11 taluks. Data from these stations for the period from 1989 to 2018 is analyzed. The southwest monsoon contributes around 55 percent of the annual rainfall. The other monsoon (NE) yields around 30 percent. The balance of around 15 percent results from the pre-monsoon. On annual basis, the variability coefficient is less than 30 percent indicating consistent rainfall. The lowest annual rainfall recorded in the district is around 300mm while the highest is over 1300mm. There is one meteorological observatory at KGF, which has long term records. The one at Kolar is of recent origin. Normally April and May are hottest months with temperatures as high as 40° C. They are generally lowest during December being as low as 10°C. Potential evapotranspiration is around 1550mm annually ranging from 170mm in Apr-May period to less than 100mm during Nov–Dec period. Being a semi-arid area the district is drought prone. In the recent years, 2002 and 2003 are deficient in rainfall. On an average year 2004 is a normal year and 2005 is a rainfall excess year amounting to nearly 50 per cent excess over the normal.

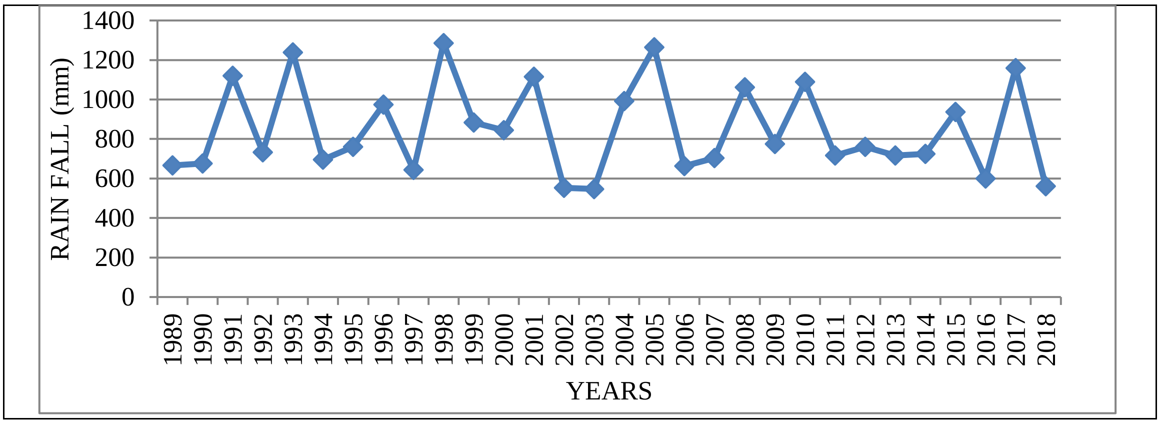

Fig - 2: Annual Rainfall Pattern of Kolar region.

III. METHODOLOGY

A. Markov Chain Model:

Markov chain matrices were developed by Andrei Markov a Russian Mathematician at the beginning of 20th century. A stochastic process with dependent variables in discrete state and time is defined as Markov Chain. A Markov chain is a process that consists of a finite number of states and some know probabilities describing a sequence of possible events in which the probability of each event depends only on the states attained in the previous events. These models have gained importance in recent years due to its easy statistical calculation and less errors. The model is said to be satisfied the Markovian property if the upcoming future condition depends upon the present condition and not the past condition. In other words the present condition is the cumulative of all the existing past conditions.

Generally the number of states at any given time is considered as m and number of transitions between two successive timings were assumed as $m \times m$. Then the transition matrix is generated with the defined values. From these transition values probabilities were calculated from state i to j at any given time instant. From these concept the transition probability matrix can be developed as [7].

$$P = \begin{bmatrix} P11 & P12 & . & . & . & P1m \\ P21 & P22 & . & . & . & P2m \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ Pm1 & Pm2 & . & . & . & Pmm \end{bmatrix} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(1)}$$

These probabilities P_{ij} can be calculated from the equation

$$P_{ij} = \frac{F_{ij}}{\sum_{j=1}^{m} F_{ij}} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(2)}$$

Where, P_{ij}– Probability density from state i to j.

F_{ij} – Transition frequency from state i to j.

m – Maximum number of states.

P_{ij}, always $0 \le P_{ij} \le 1$where i, j = 1, 2, 3….m and$\sum_{j=1}^{m} P_{ij} = 1$.

Once getting the probability matrix P it is further extended to calculate the long term behavior of the rainfall. $P_i^{(n)}$ denotes the probability that the observation is in i^{th} state and at time n.
Considering $P^{(1)}$is the initial probability state vector of the Markov chain.
For long term analysis it can be written as

$$P^{(n+1)} = P^{(n)} \times P^{(1)} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(3)}$$

Where, $P^{(n)}$ is the probability matrix of previous transition values.
As n value in the equation (3) increases, the matrix reaches a steady state with a large period of time. From the steady state matrix and initial probability vector, the limiting state probabilities can be calculated with the equation (4) [1].

$$P^{(s)} = P^{(0)} \times P^{(n+1)} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\text{(4)}$$

Where, $P^{(s)}$ is the limiting state probabilities
$P^{(0)}$ is the initial state probability matrix.

In this study, $P^{(0)}$ = [0.5, 0.5] for daily analysis and [0.25, 0.25, 0.25, 0.25] for yearly analysis.

B. Linear Regression Model:

Linear Regression is a statistical approach of modeling the relation between a dependent variable and one or more independent variables. Most commonly regression analysis is done using the method of least squares which gives the best possible results with least errors. The best fit in the least squares sense minimum sum of the squared residuals. Residuals are defined as the difference between the observed values and predicted values. If the point lies exactly on the line then the algebraic sum of residual errors is zero [5], [6].
The linear regression equation is given by:

$$y = a + bx \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(5)}$$

Where, y is the dependent variable, in this case annual rainfall.
x is the independent variable, in this case year of the rainfall.
a is the intercept of the line on the vertical axis.
b is the slope of the line.
From the equation (5), intercept *'a'* can be calculated as:

$$a = \bar{y} - b\bar{x} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(6)}$$

Where, $\bar{y}$ - is the mean of the observed dependent variables.

$\bar{x}$ - is the mean of the observed independent variables.

$$b = \frac{N\sum X_i Y_i - \sum Y_i \sum X_i}{N\sum X_i^2 - ((\sum X_i)^2)} \text{.. (7)}$$

The strength of the relation is given by the coefficient of correlation (r). It takes the value with in the range of [-1, 1]. If the value is close to unity then it is said that the relation existed is perfect, if the correlation coefficient is close or equal to zero no relation exists between the variables. If the correlation coefficient is positive then there will be positive relation between the variables, line moves upward. If the correlation coefficient is negative it indicates a negative relationship between the variables, the line downwards. The value of *r* is given by:

$$r = \frac{N\sum X_i Y_i - \sum X_i \sum Y_i}{\sqrt{[N\sum X_i^2 - (\sum X_i)^2] \times [N\sum Y_i^2 - (\sum Y_i)^2]}} \text{....................... (8)}$$

Coefficient of determination (r^2) is the statistical measure of how close the data are fitted to regression line equation. It gives the percentage of the response variable variation of the linear model and it can be expressed as the ratio of the expected variation to the total variation.

Fitting the Linear Regression lines were executed using Microsoft Excel.

IV. RESULTS and DISCUSSION:

A. Analysis of Daily Rainfall:

For the prediction of daily rainfall, the 30 years rainfall data during a period of 1989 to 2018 is segregated into number of rainy days and non-rainy days in a year and tabulated in table1. The occurrence of rainfall in any given day is compared to its successor and tabulated for each year. Based on these values transition matrix is developed.

Table1: Number of rainy days and non-rainy days in a year from 1989 to 2018.

Year	Rainy days	Non rainy days	Total days	Year	Rainy days	Non rainy days	Total days
1989	56	309	365	2004	86	280	366
1990	66	299	365	2005	104	261	365
1991	94	271	365	2006	63	302	365
1992	65	301	366	2007	87	278	365
1993	70	295	365	2008	78	288	366
1994	75	290	365	2009	139	226	365
1995	54	311	365	2010	172	193	365
1996	74	292	366	2011	80	285	365
1997	57	308	365	2012	52	314	366
1998	76	289	365	2013	79	286	365
1999	56	309	365	2014	62	303	365
2000	80	286	366	2015	101	264	365
2001	79	286	365	2016	62	304	366
2002	55	310	365	2017	96	269	365
2003	69	296	365	2018	85	280	365

Consider the year 1989 for the analysis, assuming that it represents the whole 30 year data.

The number of Rainy days and Non Rainy days in year 1989 is tabulated below:

YEAR 1989	
Number of non- rainy days	309
Number of rainy days	56
Total number of days	365

The following table shows the comparison of first day to its next day with respect to rain.

Second day / First day	Non rainy day	Rainy day	Total
Non rainy day	278	30	308
Rainy day	30	26	56

The transition matrix is given by from equation (1) & (2).

Second day / First day	Non rainy day	Rainy day
Non rainy day	278/308	30/308
Rainy day	30/56	26/56

Second day / First day	Non rainy day	Rainy day
Non rainy day	0.9026	0.0974
Rainy day	0.5357	0.4643

This says that on any given day if it is non-rainy days then there will be chance of 90.26 % that the following day will also be non-rainy day and 9.74% chances of being rainy day. In the same way on any rainy day there is a chance of 46.43% that the next day will be a rainy day and 53.57% will be non-rainy day.
Long term analysis by considering 1989 year from equation (3), the transition matrix is:

$P = \begin{bmatrix} 0.9026 & 0.0974 \\ 0.5357 & 0.4643 \end{bmatrix}$; $P^{(2)} = \begin{bmatrix} 0.8669 & 0.1331 \\ 0.7322 & 0.2677 \end{bmatrix}$; $P^{(5)} = \begin{bmatrix} 0.8461 & 0.1538 \\ 0.8405 & 0.1594 \end{bmatrix}$

$P^{(10)} = \begin{bmatrix} 0.8461 & 0.1538 \\ 0.8461 & 0.1538 \end{bmatrix}$; $P^{(20)} = \begin{bmatrix} 0.8461 & 0.1538 \\ 0.8461 & 0.1538 \end{bmatrix}$

$P^{(S)} = [0.5 \quad 0.5] \begin{bmatrix} 0.8461 & 0.1538 \\ 0.8461 & 0.1538 \end{bmatrix} = [0.8461 \quad 0.1538]$

From these above limiting probabilities, on any given day the probability percentage of non-rainy days is 84.61% and probability percentage of rainy days15.38%.
Rainfall prediction for 30 years:

Number of rainy days = $\left(\frac{15.38}{100}\right) * 10957 = 1687$

Number of non-rainy days = $\left(\frac{84.61}{100}\right) * 10957 = 9284$

Percentage of error in actual and predicted values:

% of error in rainy days = $\left(\frac{2372-1687}{2372}\right) * 100 = 28.88\%$

% of error in non-rainy days = $\left(\frac{9284-8585}{8585}\right) * 100 = 8.14\%$

Where, the error in the 1989 model found to be very high values, the conclusion from this model is questionable. Due to the huge errors in the 1989 model, it is advisable to try with another year. The average number of rainy days of entire 30 years is 79 days when compared with the year 1989 which has huge difference. Therefore, it is better to consider a year with almost same number of rainy days and non-rainy days compared to the average values.
Considering for the year 2001, the number of Rainy days and Non Rainy days in year 2001 is tabulated below:

YEAR 2001	
Number of non-rainy days	286
Number of rainy days	79
Total number of days	365

The following table shows the comparison of first day to its next day with respect to rain:

Second day / First day	Non rainy day	Rainy day	Total
Non rainy day	251	34	285
Rainy day	34	45	79

The transition matrix is given by from equation (1) & (2):

Second day / First day	Non rainy day	Rainy day
Non rainy day	251/285	34/285
Rainy day	34/79	45/79

Second day / First day	Non rainy day	Rainy day
Non rainy day	0.8807	0.1193
Rainy day	0.4304	0.5696

The above transition matrix says that on any given day if it is non-rainy day then there will be chance of 88.07 % of not having rainfall and the following day will also be non-rainy day and 11.93% chances of being rainy day. In

the same way on any rainy day there is a chance of 56.96% that the next day will be a rainy day and 43.04% will be non-rainy day.
Long term analysis by considering 2001 year from equation (3):

$P = \begin{bmatrix}0.8807 & 0.1193\\0.4304 & 0.5696\end{bmatrix}$; $P^{(2)} = \begin{bmatrix}0.8270 & 0.1730\\0.6242 & 0.3758\end{bmatrix}$; $P^{(5)} = \begin{bmatrix}0.7869 & 0.2130\\0.7684 & 0.2315\end{bmatrix}$

$P^{(10)} = \begin{bmatrix}0.7830 & 0.2169\\0.7829 & 0.2172\end{bmatrix}$; $P^{(20)} = \begin{bmatrix}0.7829 & 0.2170\\0.7829 & 0.2170\end{bmatrix}$

$$P^{(S)} = [0.5 \quad 0.5]\begin{bmatrix}0.7829 & 0.2170\\0.7829 & 0.2170\end{bmatrix} = [0.7829 \;\; 0.2170]$$

From these above limiting probabilities, on any given day the probability percentage of non-rainy days is78.29% and probability percentage of rainy days is 21.7%.
Rainfall prediction for 30 years:

Number of rainy days = $\left(\frac{21.7}{100}\right) * 10957 = 2378$ days

Number of non-rainy days = $\left(\frac{78.29}{100}\right) * 10957 = 8587$ days

Percentage of error in actual and predicted values

% of error in rainy days = $\left(\frac{2378-2372}{2372}\right) * 100 = 0.25\%$

% of error in non-rainy days = $\left(\frac{8585-8578}{8585}\right) * 100 = 0.08\%$

As the error from the year 2001 is less, this year can be used for long term analysis. From this, we can say on any given day the chance of occurrence of rain is 21.7% and78.29% chances that it will not be rain.

Table-2: State rating for different rainfall intensity of Kolar region

Intervals	No. of Years	Category
0 to 700 mm	9	1
700 to 900 mm	10	2
900 to 1150 mm	7	3
More than 1150mm	4	4

B. Analysis of annual rainfall:
In this study, the yearly rainfall status of the Kolar region is divided into 4 categories on basis of the depth of rainfall in that year as Category-1 to Category- 4. It is tabulated in table - 2:

Development of Transition Sequence of Each Category:

Table-3: State transition of Kolar region for entire 30 years.

Year	Rain fall	Category	Year	Rainfall	Category
1989	665.4	1	2004	992.1	3
1990	675.8	1	2005	1264.8	4
1991	1120.0	3	2006	662.2	1
1992	731.8	2	2007	702.7	2
1993	1238.3	4	2008	1061.8	3
1994	695.8	1	2009	774.0	2
1995	760.1	2	2010	1088.5	3
1996	973.6	3	2011	717.0	2
1997	644.2	1	2012	759.5	2
1998	1285.2	4	2013	716.5	2
1999	883.5	2	2014	725.0	2
2000	844.5	2	2015	936.0	3
2001	1115.0	3	2016	600.0	1
2002	552.1	1	2017	1158.5	4
2003	547.0	1	2018	561.0	1

The 4 rainfall intensity categories have tabulated in table 2. It describes the 4 different possible conditions that were classified with the help of existing data. With this information, every year is classified into the above mentioned 4 categories and was tabulated in table – 3; [4].

Development of Transition Matrix from Available Data for Each Category:
The data matrix is generated for each category of the rainfall intensity. This describes the transition from the one category to other category. The following table compares the state of the present year with respect to state of the successor year. The probability matrix obtained from the transition matrix is given by from equation (1) & (2) and tabulated below.

To Category / From Category	1	2	3	4
1	2	2	2	2
2	0	4	5	1
3	3	3	0	1
4	3	1	0	0

To Category / From Category	1	2	3	4
1	0.25	0.25	0.25	0.25
2	0	0.4	0.5	0.1
3	0.43	0.43	0	0.14
4	0.75	0.25	0	0

$$P = \begin{bmatrix} 0.25 & 0.25 & 0.25 & 0.25 \\ 0 & 0.4 & 0.5 & 0.1 \\ 0.43 & 0.43 & 0 & 0.14 \\ 0.75 & 0.25 & 0 & 0 \end{bmatrix}$$

From the above transition probability matrix, it can be stated that if the rainfall in any year is in Category 1 (0 to 700 mm) and in its successive year is 25% of chance that it will remain as same, 25% of chances that it will be in Category 2 (700 to 900 mm), 25% of chances it may similar to Category 3 (900 to 1150 mm) and 25% of chances it may similar to Category 4 (more than 1150 mm). If this year's rainfall is between 700 to 900 mm i.e., in Category 2, then in the successive year there is 40% of chance that it will remain same, 50% chances that it will be in Category 3(900 to 1150 mm) and 10% of chances that it will be in Category 4 (more than 1150 mm).
For long term analysis from equation (3):

$$P^{(2)} = \begin{bmatrix} 0.358 & 0.333 & 0.188 & 0.123 \\ 0.290 & 0.400 & 0.200 & 0.110 \\ 0.213 & 0.315 & 0.323 & 0.151 \\ 0.188 & 0.288 & 0.313 & 0.213 \end{bmatrix}; \quad P^{(5)} = \begin{bmatrix} 0.273 & 0.344 & 0.245 & 0.139 \\ 0.271 & 0.345 & 0.246 & 0.138 \\ 0.281 & 0.347 & 0.237 & 0.136 \\ 0.286 & 0.347 & 0.233 & 0.134 \end{bmatrix};$$

$$P^{(10)} = \begin{bmatrix} 0.276 & 0.345 & 0.242 & 0.137 \\ 0.276 & 0.345 & 0.242 & 0.137 \\ 0.276 & 0.345 & 0.242 & 0.137 \\ 0.276 & 0.345 & 0.242 & 0.137 \end{bmatrix}; \quad P^{(20)} = \begin{bmatrix} 0.276 & 0.345 & 0.242 & 0.137 \\ 0.276 & 0.345 & 0.242 & 0.137 \\ 0.276 & 0.345 & 0.242 & 0.137 \\ 0.276 & 0.345 & 0.242 & 0.137 \end{bmatrix};$$

Limiting probabilities from the equation (4):

$$P^{(0)} * P^{(20)} = \begin{bmatrix} 0.25 & 0.25 & 0.25 & 0.25 \end{bmatrix} \begin{bmatrix} 0.276 & 0.345 & 0.242 & 0.137 \\ 0.276 & 0.345 & 0.242 & 0.137 \\ 0.276 & 0.345 & 0.242 & 0.137 \\ 0.276 & 0.345 & 0.242 & 0.137 \end{bmatrix} = [0.276 \;\; 0.345 \;\; 0.242 \;\; 0.137]$$

From the above iterations, it can be stated that for any given year there is 27.6% of chances that the rainfall will be in Category- 1 (0 to 700 mm), 34.5% of chances that the rainfall will be in Category-2 (700 to 900 mm), 24.2% of chances that the rainfall will be in Category-3 (900 to 1150 mm) and 13.7% of chances that it will be Category-4 (more than 1150 mm).

C. Prediction of Rainfall Magnitude by Linear Regression Model:

Considering the annual rainfall from year 1989 to 2003 from table -3, a linear regression line is plotted for this data as show in fig-3. The regression equation and r^2 value has been mention in the graph.

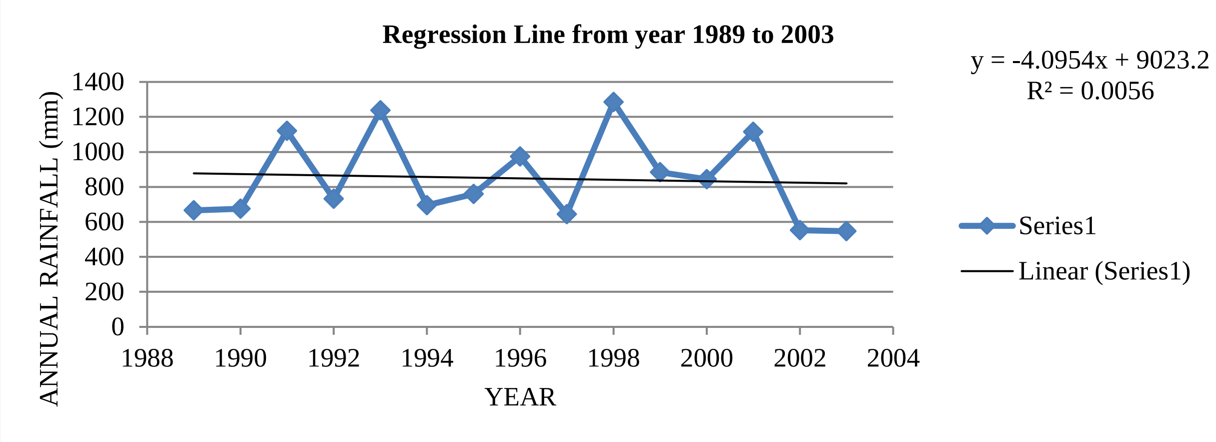

Fig – 3: Linear Regression plots of Kolar region for 15 years from 1989 to 2003.

Table – 4: Comparison of predicted rainfall with actual rainfall for successive five years 2004 to 2008.

YEAR	Actual Rainfall	Predicted Rainfall	% of Error
2004	992.1	816.02	17.1
2005	1264.8	811.9	35.2
2006	662.2	807.8	29.4
2007	702.7	803.7	14.4
2008	1061.8	799.6	24.7

With the help of this regression equation, rainfall for the future 5 years from 2004 to 2008 is predicted and compared to the actual values. The percentage error of the actual value and the predicted value is calculated in table -4.

From table – 4, the average error of the prediction value for the next five years with respect to its actual value is 24.16%. Similarly, it can be possible to fit the regression lines for the next fifteen years 1994 to 2008 and 1999 to 2013. With the help of the regression line, rainfall of the next respective future five years can be estimated and compared with the actual rainfall values as well as the percentage of error is also calculated. These values are represented in fig – 4 and table -5. The statistical parameters of every 15 years data is calculated and tabulated in table -6.

Table – 5: Regression statistics of annual rainfall for fifteen years data.

S.No	Year	Regression Equation	r^2	% of average error in the predicted values of next 5 years
1	1989 to 2003	Y = - 4.095X + 9023.2	0.0056	24.16
2	1994 to 2008	Y = 4.6368X – 8412.6	0.0073	20.66
3	1999 to 2013	Y = - 4.5182 + 9909.1	0.0088	22.85

Table – 6: Statistical Parameters of every 15 years.

S.No	Year	Mean	Standard deviation	Co-efficient of Variation	Co-efficient of Skewness
1	1989 to 2003	848.82	243.91	0.287	0.588
2	1994 to 2008	865.64	242.19	0.280	0.411
3	1999 to 2013	845.41	215.81	0.255	0.450

From the table – 5, it can be noticed that the error in the prediction of successive 5 years rainfall is 20% to 25% and on an average, the error will be 22.56%.

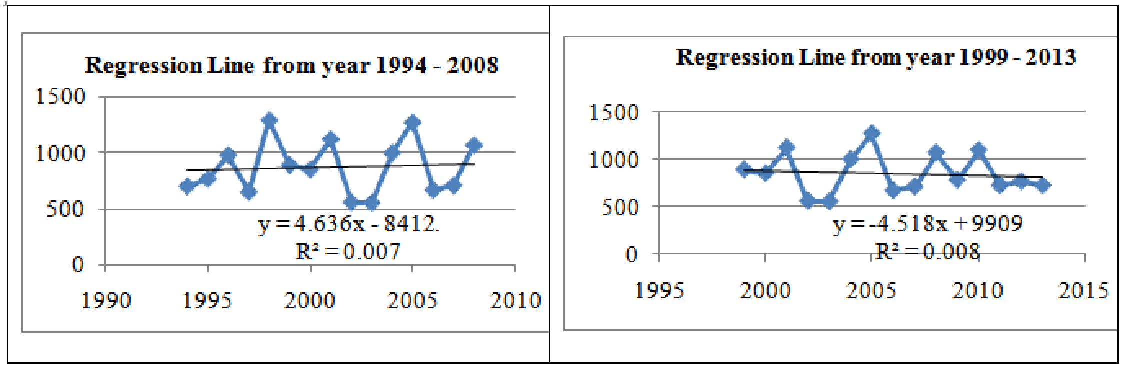

Fig – 4: Linear Regression plots of Kolar region for 15 years from 1994 to 2008 and 1999 to 2013

Now considering the last fifteen years rainfall data from 2004 to 2018, a regression line is fitted for these values as shown in the fig–5 and the obtained regression equation is Y= -16.027X + 33079. From this regression equation, the rainfall of successive five years during year 2019 to 2023 is predicted and tabulated in table – 7.

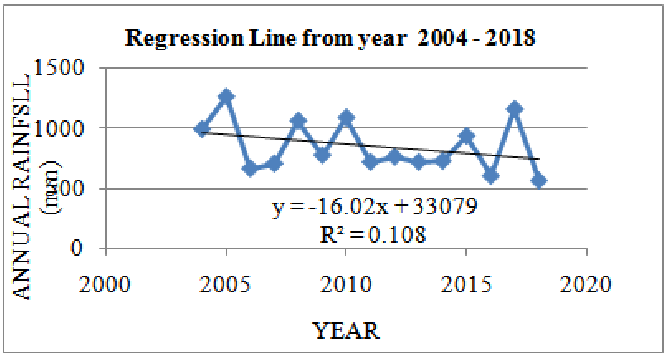

Fig – 5: Linear Regression plot of Kolar region for 15 years from 2004 to 2018

Table – 7: Predicted rainfall for the years 2019 to 2023:

S.No	YEAR	Predicted annual rainfall(mm)
1	2019	720.5
2	2020	704.5
3	2021	688.4
4	2022	672.4
5	2023	656.3

V. CONCLUSION

This paper explains the implementation of the Markov Chain probability method to develop the models for analyzing and predicting the probability of occurrence of rain in any given day and any given year for the study area. As rainfall is not uniform in every year, with the help of these models the occurrence of rainfall in any given day or given year can be predicted with better accuracy for crop planning. With Markov Chain Techniques, the probable of occurrence of maximum rainfall and discharge for a given period can be calculated and this can be used for any hydraulic and engineering designs. Also, Linear Regression analysis is used to predict the magnitude of the rainfall intensity of 5 years with the existing previous 15 years rainfall data. It is observed that the average error of the predicted and observed data is 22.5%. The next successive five years rainfall magnitudes are also estimated with 77.5% of accuracy.

ACKNOWLEDGMENT

We are very thankful to Karnataka State Natural Disaster Monitoring Centre (KSNDMC), for providing the information of the 30 years rainfall data.

REFERENCE:

[1]. Abubakar Usman Yusuf, Lawal Adamu, Muhammed Abdullahi; "Markov Chain Model and its Application to Annual Rainfall Distribution for Crop Production", American Journal of Theoretical and Applied Statistics; 2014, 3(2).

[2]. G. Arvind, P. Ashok Kumar, S. Girish Karthik and C. R. Suribabu, "Statistical Analysis of 30 Years Rainfall Data: A Case Study"; IOP Conf. Series: Earth and Environment Science 80 (2017) 012067.

[3]. Joachim Osuagwu, "A Markov Model for Prediction of Annual Rainfall", Journal of Engineering and applied sciences, November 2018.

[4]. Katkar Surendrakumar, Nagrale Prashant, Patil Mayuresh, “Application of Markovian Probabilistic Process to Develop a Decision Support System for Pavement Maintenance Management”; International Journal of Scientific & Technology Research; Volume 2; Issue 8, August 2013.
[5]. M. Nyatuame, V. Owusu-Gyimah and F. Ampiaw; “Statistical Analysis of Rainfall Trend for Volta Region in Ghana”; International Journal on Atmospheric Sciences, Volume 2014, Article ID 203245.
[6]. Mohammad Shohidul Islam, Sultana Easmin Siddika, S. M. Injamamul HaqueMasum; “Rainfall Prediction in South Eastern Part of Bangladesh by Linear Regression Method.” International Journal of Emerging Research in Management & Technology, ISSN: 2278-9359, Volume -6, Issue – 6.
[7]. Saurabh Singh, Babita Pandey, “Forecasting of Monthly and Annual Rainfall at Rajasthan Region”, Journal of Civil Engineering Technology, p-ISSN: 2349-8404; e-ISSN: 2349-879X; Volume 5; Issue 2.
[8]. Higher Engineering Mathematics (textbook) by Dr. B. S. Grewal.
[9]. Stochastic Hydrology (textbook) by Dr. P. Jayarami Reddy.

Mohammed A Rasiq, etal. "Prediction of Daily and Yearly Rainfall of Kolar Region of Karnataka State By Markov Chain and Regression Model." *IOSR Journal of Engineering (IOSRJEN)*, 10(2), 2020, pp. 07-16.