

## The Use of Markov Model in Continuous Time for Prediction of Rainfall for Crop Production

Abubakar U. Y., Lawal A., Muhammed A.

Department of Mathematics and Statistics, Federal University of Technology, Minna, Nigeria.

**Abstract:** In a first order Markov process, if the state is known for any specific values of the time parameter, that information is sufficient to predict the next behavior of the process beyond that point. This principle was used to formulate a four-state model in continuous time to study the annual rainfall data with respect to the annual rainfall distribution for crop production in minna. It was observed that if it is low rainfall in a given year, it would take at most 25%, 33% and 27% of the time to make a transition to moderate rainfall also well spread, high rainfall, and moderate rainfall but not well spread respectively in the far future. Thus given the rainfall in a year, it is possible to determine quantitatively the probability of finding rainfall in other states in the following year and in the long run. This is an important information that could assist the farmers to plan strategies for high crop production in the region.

**Keywords:** Markov Model, Transition probability, Annual rainfall, Crop production.

### I. Introduction

Crop production is the major activity of people that live in Minna and the surrounding. Rainfall is the main source of water for growing crops in Nigeria because bumper harvest and lean years depend on annual rainfall quantity and variability. Several attempts have been made to support the farmer with information services so that they could boost crop production in the face of variability and uncertainty of annual rainfall with mathematical models. In [1] was modified the Chapman-Kolmogorov Equation (CKE) and applied it to model the daily precipitation data of Abeokuta, Ogun State, Nigeria. Presented in [6] is the application of Semi-Markov model to study desert encroachment in Nigeria in view of the adverse effect of short rainfall on crop production in the region. Reported in [2] is a 7-State Markov Model to study the annual rainfall in Nadu, India. This paper considers the Markov chain model reported in [9] where the application was made in discrete state and time and then attempt to improve on it so as to provide some quantitative prediction in continuous time.

### The Model Equations

As contained in [5], a continuous time stochastic process  $\{X(t)\}$  is an infinite family of random variables indexed by the continuous real variable  $t$ , that is for any fixed  $t$ ,  $X(t)$  is a random variable and the collection of all these (for all  $t$ ) is a stochastic process. We shall take  $t$  as time, so we may expect  $X(t_1)$ , the random variable at time  $t_1$  to be dependent on  $X(t_0)$ , where  $t_0 < t_1$  but not upon  $X(t_2)$ , where  $t_2 > t_1$ . We refer to the value of  $X(t_1)$  as the state of the process at the time  $t$ . we assume  $X(t)$  are discrete state continuous time stochastic process.

We consider a finite (or countably infinite) set of point  $(t_0, t_1, \dots, t_n, t)$ ,  $t_0 < t_1 < t_2 < \dots < t_n < t$  and  $t, t_r \in T$  ( $r = 1, 2, \dots, n$ )

where  $T$  is the parameter space of the process  $\{X(t)\}$ . The dependence exhibited by the process  $X(t)$ ,  $t \in T$  is called "Markov dependence" for given values of  $X(t_1), X(t_2), \dots, X(t_n)$  depends only on  $X(t_n)$  which is the most recent known value of the process.

$$\begin{aligned} \text{That is, if } P[X(t) \leq X | X(t_n) = X_n, X(t_{n-1}) = X_{n-1}, \dots, X(t_0) = X_0] \\ = P[X(t) \leq X | X(t_n) = X_n] \\ = F(X_n, X; t_n, t) \end{aligned} \quad (1)$$

The stochastic process exhibiting this property is called a Markov process. In a Markov process, therefore, if the state is known for any specific values of the time parameter  $t$  that information is sufficient to predict the next behavior of the process beyond that point.

As a consequence of the property given by (1) we have the following relation

$$F(x_0, x; t_0, t) = \int_{y \in S} F(y, x, \tau, t) dF(x_0, y, t_0, \tau) \quad (2)$$

where  $t_0 < \tau < t$  and  $S$  is the state space of the process  $X(t)$ .

Equation (2) is called the Chapman Kolmogorov equation for the process. These are basic equation in the study of Markov processes they enable us to build a convenient relationship for the transition probabilities between any points in T at which the process exhibits the property of Markov dependence.

Another statement of the Chapman Kolmogorov equation and the proof is given below, [7].

$P_i(t+s) = P(X(t+s) = j | X(0) = i)$  (definition)

$$\begin{aligned}
 &= \sum_k P(X(t+s) = j, X(t) = k | X(0) = i) \text{ (Marginal from joint)} \\
 &= \sum_k P(X(t+s) = j | X(t) = k, X(0) = i) P(X(t) = k | X(0) = i) \text{ (Joint from Conditional)} \\
 &= \sum_k P(X(t+s) = j | X(t) = k) P(X(t) = k | X(0) = i) \text{ (Markov assumption)} \\
 &= \sum_k P(X(s) = j | X(0) = k) P(X(t) = k | X(0) = i) \text{ (Stationary)} \\
 &= \sum_k P_{ik}(s) P_{kj}(t) \text{ (definition)}
 \end{aligned}$$

Thus  $P_{ij}(t+s) = \sum_k P_{ik}(t) P_{kj}(s)$

This is the Chapman Kolmogorov equation in general form. We shall use it in the special form of

$$P_{ij}(t + \delta t) = \sum_k P_{ik}(t) P_{kj}(\delta t)$$

This equation requires the Markov assumption to permit a multiplication of the probabilities referring to events during t and to event during  $\delta t$ , it also requires stationary to permit use of the same probability functions for this interval t and for the later interval  $\delta t$ .

A Markov process is said to be time homogeneous or stationary if

$$P(X(t_2) = j | X(t_2) = i) = P(X(t_2 - t_1) = j | X(0) = i)$$

For all i, j,  $t_1$ , and  $t_2$  such that  $t_1 < t_2$

In words the process is Stationary if these conditional probabilities depend on the interval between the events rather than an absolute time.

A Stationary Markov Process is completely described by its transition probability functions, denoted  $P_{ij}(t)$  where

$$P_{ij}(t) = P(X(t) = j | X(0) = i)$$

We observe that  $P_{ij}(t)$  function is probabilities, for all t, they are non negative bounded function because they must lie between 0 and 1

$$P_{ij}(0) = P(X(0) = j | X(0) = i) \text{ clearly, for } i \text{ different from } j$$

$P_{ij}(0) = 0$  and for i equal to j,  $P_{jj}(0) = 1$ , if we fix i and vary j over all states, the sum of the  $P_i(t)$  must equals 1 (for all t)

$$\begin{aligned}
 \sum_j P_{ij}(t) &= \sum_j P(X(t) = j | X(0) = i) \\
 &= P(X(t) = \text{any of its possible states} | X(0) = i) \\
 &= 1
 \end{aligned}$$

Under the assumption that the  $P_{ij}(t)$  are continuous functions of time we can express  $P_{ij}$  for small  $\delta t$  by the use of Maclaurin's series.

$$p_{ij}(\delta t) = p_{ij}(0) + p'_{ij}(0)\delta t + o(\delta t^2)$$

where  $o(\delta t^2)$  represents all terms of the order of  $(\delta t)^2$  or higher. If we consider this expression for  $i \neq j$ , and

Let  $q_{ij} = p'_{ij}(0)$  we obtain

$$p_{ij}(\delta t) = q_{ij}\delta t + o(\delta t^2)$$

We may think of this as a linear approximation to  $P_{ij}(t)$  which is a good approximation as long as  $\delta t$  is small. The  $q_{ij}$  is called the transition rate from  $i$  to  $j$  since  $P_{ij}(0) = 1$  and this is the minimum value. We should be certain that  $q_{ij}$  non-negative.

For  $i = j$  the Maclaurin's series expansion yields

$$p_{jj}(\delta t) = 1 + p'_{jj}(0)\delta t + o(\delta t^2)$$

$$\text{let } q_{jj} = p'_{jj}(0)$$

we get the linear approximation

$$p_{jj}(\delta t) = 1 + q_{jj}\delta t + o(\delta t^2)$$

Since we know that  $p_{ij}(0) = 1$  and that is the maximum value we may be certain that  $q_{ij}$  is non-positive.

If we now consider the Chapman-Kolmogorov equation

$$p_{ij}(t + \delta t) = \sum_k p_{ik}(t) p_{kj}(\delta t)$$

for small  $\delta t$ , and substitute our linear approximation, we get

$$p_{ij}(t + \delta t) = p_{ij}(t)[1 + q_{jj}\delta t + o(\delta t^2)] + \sum_{k \neq j} p_{ik}(t)[q_{kj}\delta t + o(\delta t^2)]$$

$$\frac{p_{ij}(t + \delta t) - p_{ij}(t)}{\delta t} = p_{ij}(t)q_{jj} + \frac{p_{ij}(t)o(\delta t^2)}{\delta t} + \sum_{k \neq j} \left[ p_{ik}(t)q_{kj} + \frac{p_{ik}(t)o(\delta t^2)}{\delta t} \right]$$

$$= \sum_k p_{ik}(t)q_{kj} + \sum_k \frac{p_{ik}(t)o(\delta t^2)}{\delta t}$$

Taking limit as  $\delta t \rightarrow 0$

$$\frac{dp_{ij}(t)}{dt} = \sum_k p_{ik}(t)q_{kj}$$

$$\text{In matrix form we have } \frac{dp(t)}{dt} = p(t)Q \tag{3}$$

where  $\frac{dp(t)}{dt}$  is the matrix whose  $(i, j)^{th}$  element is  $\frac{dp_{ij}(t)}{dt}$ ,

$p(t)$  is the matrix whose  $(i, j)^{th}$  element is  $p_{ij}(t)$ , and  $Q$  is the matrix whose  $(i, j)^{th}$  element is  $q_{ij}$

The elements of  $Q$  may be further related by extending the properties of  $P(t)$

In particular, since for each  $i$

$$\sum_j p_{ij}(t) = 1$$

$$\frac{d}{dt} \left[ \sum_j p_{ij}(t) \right]_{t=0} = \left[ \frac{d}{dt} (1) \right]_{t=0}$$

$$\sum_j \frac{d}{dt} p_{ij}(t) \Big|_{t=0} = 0$$

$$\sum_j q_{ij} = 0$$

In words, each row of Q must sum to zero. Since off-diagonal element in non negative, the diagonal element  $q_{ii}$ , must be equal in magnitude and opposite in sign to the sum of others in the same rows. That is

$$q_{ii} = - \sum_{i \neq j} q_{ij} \tag{4}$$

To obtain the solution to (3) the initial condition  $P_i(0), i = 1,2,3,4...$  must be specified

Taking the Laplace transform of (3) we obtained

$$P(s) = p(0)(SI - Q)^{-1} \tag{5}$$

Thus p(t) is obtained as the inverse transform of P(s), [4].

APPLICATION

Suppose that the amount of annual rainfall in Minna in a year is considered as a random variable  $X$ , the collection of this random variable over the years n constitutes a stochastic process  $X_n$ ,  $n = 0,1,2,3.....$  It is assumed that this stochastic process satisfies first order Markov property, [9].

It has been observed over the past ten years that the annual rainfall in Minna varied from 698mm to 1423.4mm. It is known that the rainfall requirement for the major crops grown in the environment is within the range of 1000mm-1200mm and should spread fairly across from May to October in the year. The annual rainfall below or above this range and not spread well may lead to poor harvest, contained in [8].

Let the annual rainfall be modeled by four state Markov model.

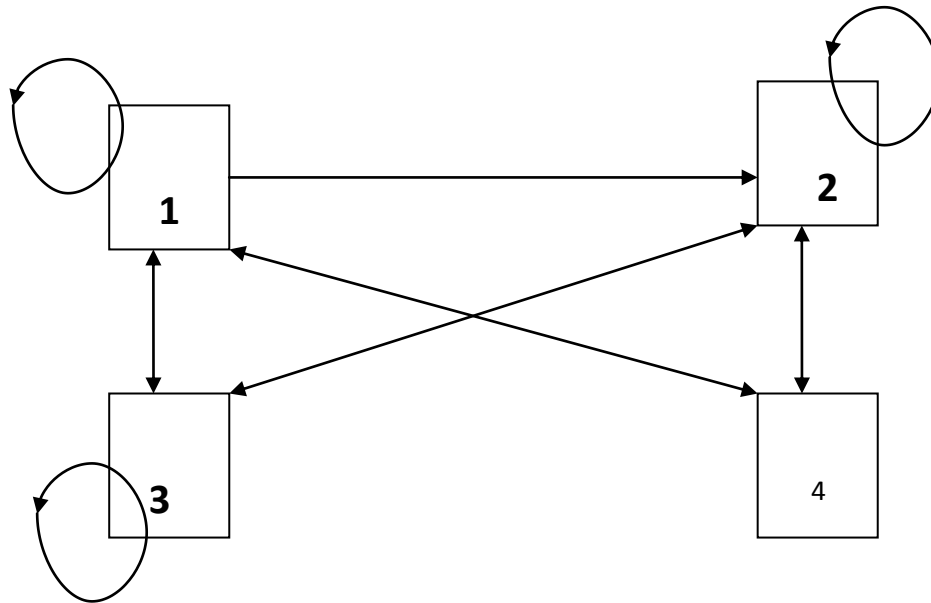
**State1:** Low rainfall (Annual rainfall below 1000mm)

**State2:** Moderate rainfall (Annual rainfall within the range 1000mm-1201mm also well spread)

**State3:** High rainfall (Annual rainfall above 1201mm)

**State4:** Moderate rainfall (Annual rainfall within the range 1000mm-1201mm but not well spread)

On the basis of the observed data, the transition between the states is described by the transition diagram shown in figure1.



**Fig 1 Transition diagram for the annual rainfall.**

Consequently, we have the following transition matrix p.

$$p = \begin{bmatrix} P_{11} & P_{12} & P_{13} & P_{14} \\ 0 & P_{22} & P_{23} & P_{24} \\ P_{31} & P_{32} & P_{33} & 0 \\ P_{41} & P_{42} & 0 & 0 \end{bmatrix}$$

The matrix P is homogeneous transition stochastic matrix and the transition probabilities  $p_{ij}$  are fixed and independent of time. The transition probabilities must satisfy

$$P_{ij} \geq 0, i, j = 1, 2, 3, 4. \text{ and } \sum_{j=1}^4 P_{ij} = 1, \quad i = 1, 2, 3, 4. \quad [8 \text{ and } 9]$$

The summary statistics of the annual rainfall in Minna for a period of forty one years is presented in table1.

Table1: A summary statistics of annual rainfall in Minna between 1970-2010 and states distribution (Source; from [8] and [9]).

Annual rainfall in mm	Frequency	State
Above 1201	18	3
1000-1201 (WS)	14	2
1000-1201(NS)	3	4
Below 1000	6	1

**WS:** mean the annual rainfall that spread fairly across the rainy season

**NS:** mean annual rainfall that does not spread evenly across the rainy season

The transition between the states is recorded in the transition count matrix M shown below.

$$M = \begin{bmatrix} 2 & 2 & 1 & 1 \\ 0 & 6 & 6 & 1 \\ 3 & 4 & 11 & 0 \\ 1 & 2 & 0 & 0 \end{bmatrix}$$

Normalizing this matrix using equation (4) we obtain a Matrix A

$$A = \begin{bmatrix} -4 & 2 & 1 & 1 \\ 0 & -7 & 6 & 1 \\ 3 & 4 & -7 & 0 \\ 1 & 2 & 0 & -3 \end{bmatrix}$$

The Matrix A can be expressed as the expected value of the exponential distribution to get a Matrix Q.

$$Q = \begin{bmatrix} -2.50 & 0.5 & 1 & 1 \\ 0 & -1.17 & 0.17 & 1 \\ 0.33 & 0.25 & -0.58 & 0 \\ 1 & 0.5 & 0 & -1.5 \end{bmatrix}$$

Solving equation (5) with the Matrix Q using maple software, we have the following set of equations.

$$P_{12}(t) = 0.2495604333 - 0.0090440077760e^{-0.7708862413t} - 0.19657749e^{-1.796647256t} - 0.04394867763e^{-3.182466503t}$$

$$P_{13}(t) = 0.3329363054 + 0.03252636670e^{-0.7708862413t} - 0.2658836081e^{-3.182466503t} - 0.099579064e^{-1.796647256t}$$

$$P_{14}(t) = 0.2668254778 - 0.3895921738e^{-3.182466503t} - 0.01881100472e^{-0.7708862413t} + 0.1415777007e^{-1.796647256t}$$

$$P_{23}(t) = 0.3329363054 - 0.5476416039e^{-0.7708862413t} - 0.09639085915e^{-3.182466503t} + 0.311096156e^{-1.796647256t}$$

$$P_{24}(t) = 0.2668254778 + 0.3167180879e^{-0.7708862413t} - 0.1412389612e^{-3.182466503t} - 0.4423046045e^{-1.796647256t}$$

$$P_{42}(t) = 0.259560433 + 0.09201897185e^{-0.770886241256t} + 0.0308565067e^{-3.182466503t} - 0.3724359119e^{-1.796647256t}$$

$$P_{43}(t) = 0.3329363054 - 0.3309420955e^{-0.7708862413t} - 0.1886719456e^{-1.79664256t} + 0.1866777357e^{-3.182466503t}$$

$$P_{41}(t) = 0.1506777838 + 0.04752906760e^{-0.7708862413t} - 0.4910681609e^{-3.182466503t} + 0.2928613099e^{-1.796647256t}$$

These equations are computed to have the transition probabilities contained in table2.

Table 2: The Transition Probabilities

t	p <sub>12</sub> (t)	p <sub>13</sub> (t)	p <sub>14</sub> (t)	p <sub>23</sub> (t)	p <sub>24</sub> (t)	P <sub>42</sub> (t)	p <sub>43</sub> (t)	p <sub>41</sub> (t)
0	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
1	0.210952	0.320438	0.265443	0.127194	0.334123	0.231639	0.156294	0.200866
2	0.242142	0.33670	0.266024	0.224132	0.322194	0.259061	0.257246	0.168060
3	0.247762	0.335683	0.265581	0.280133	0.296152	0.256972	0.299326	0.156684
4	0.248997	0.33435	0.26607	0.308091	0.280995	0.253493	0.317638	0.153075
5	0.249344	0.333613	0.266445	0.321373	0.27348	0.251462	0.325901	0.151721
6	0.249468	0.333253	0.266644	0.327576	0.26992	0.250455	0.329689	0.151150
7	0.249519	0.330834	0.266741	0.330455	0.26826	0.249976	0.331435	0.150894
8	0.249541	0.333004	0.266786	0.331789	0.267489	0.249753	0.332242	0.150778
9	0.249552	0.332968	0.266807	0.332405	0.267133	0.249650	0.332615	0.150724
10	0.249556	0.332951	0.266682	0.33269	0.266968	0.249602	0.332788	0.150699
11	0.249559	0.332943	0.266822	0.332823	0.266891	0.24958	0.332868	0.150688
12	0.249560	0.332939	0.266824	0.332884	0.266856	0.249569	0.332905	0.150682
13	0.24956	0.332938	0.266825	0.332912	0.26684	0.249565	0.332922	0.15068

The transition probabilities are illustrated with graph shown in figure2.

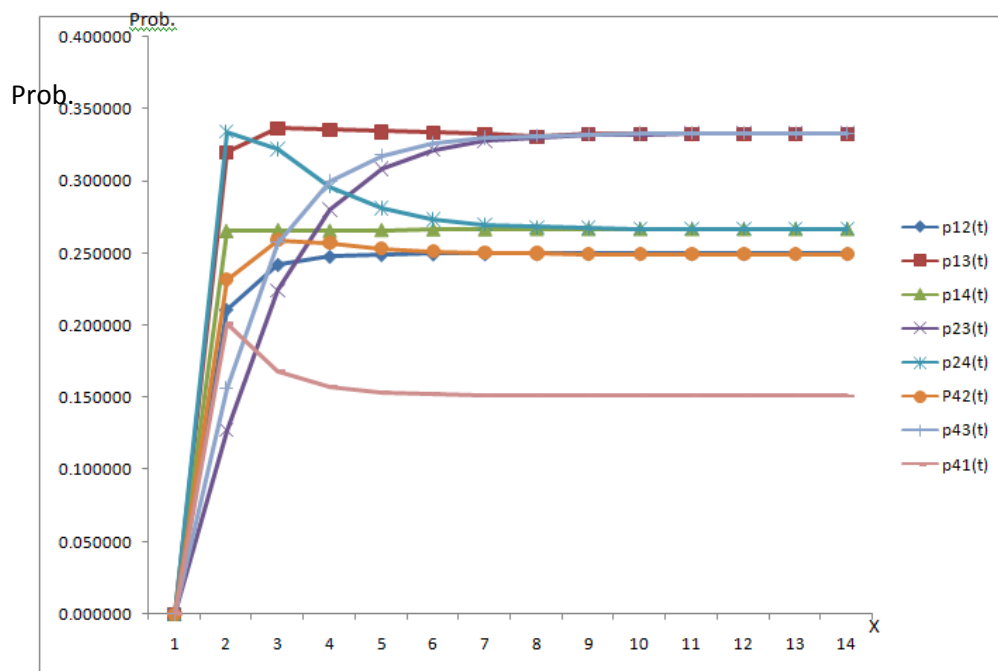


Figure 2: The graph of the transition probabilities.

## II. Discussion of Results

The result is presented in Table 2 and illustrated with graph in figure 1. The model enables us to determine the values of  $P_{11}(t)$ ,  $P_{12}(t)$ ,  $P_{13}(t)$ , .....  $P_{44}(t)$  respectively at any time  $t$ .

It is observed that  $p_{23}(t)$  and  $p_{43}(t)$  rose steadily and become stable at about  $t=7$ , but  $p_{12}(t)$  become stable about  $t=4$ . This is unlike others that rose and fall to become stable sharply for  $t < 2$  respectively. It can be seen from the table that it takes at most 0.25, 0.33 and 0.27 of the time to make transition from state1 to states2,3 and 4 respectively. It takes also at most about 0.33 and 0.27 of the time to make a transition from state2 to state3 and state4 respectively. Similarly it takes at most 0.25 and 0.33 of the time to make a transition from state4 to state2 and state3 respectively. Finally, it takes at most 0.20 of the time to make a transition from state4 to state1. These are the equilibrium transition probabilities. For instance, if it low rainfall in a given year, it would take at most 25%,33% and 27% to have moderate rainfall also well spread, high rainfall, and moderate rainfall but not well spread in the respectively in the long run. Thus given the rainfall in a year it is possible to determine quantitatively the probability of finding rainfall in other states in the following year(s) and in the long run. This is important information that could be useful to the farmers in planning strategy for high crop production in minna and the environment.

## References

- [1] Akintunde A.A, Asiribo O.E. Adebajji A.O. Adelokun A.A Agwuegbo S.O.N. (2008) Stochastic modelling of daily precipitation in Abeokuta, proceedings of the third Conference on science and National Development pp108-118 [www.unaab.edu.ng/journal/index.php/COLNAS/article/.../150/153](http://www.unaab.edu.ng/journal/index.php/COLNAS/article/.../150/153). Accessed 15/10/2012.
- [2] Tamil S. and Samuel S.(2011).Universal Journal of Environmental Research and Technology vol.1 Issue4 pp566-570.
- [3] Howard R.A.(1960) Dynamic programming and Markov processes. The M.I.T.PRESS, Massachusetts.
- [4] Korve K.N.(2000). A Three-State Continuous Time Markov Model for asthma process. ABACUS, The journal of Mathematical Association of Nigeria. Volume27, Number2, pp 33-46.
- [5] Bhat,U.N. (1984) ‘‘Element of Applied Stochastic processes’’, John Wiley, New York.
- [6] Okorie C. (2012). Semi-Markov Model in Discrete time and States and the Application to Desertification Study in Nigeria. Unpublished MSc. Thesis. Department of Mathematics and Statistics, Federal University of Technology, Minna, Nigeria.
- [7] Korve K N(1994). Operations Research MAT537, Unpublished lecture notes. Department of Mathematics, University of Jos Nigeria.
- [8] Lawal Adamu(2012), The Use of Markov Model in Predicting the Rainfall Pattern forCropProduction. Unpublished MTech Thesis, Department of Mathematics and Statistics, Federal University of Technology,Minna. Nigeria
- [9] Abubakar U.Y. and Lawal A. Muhammed A., (2013), Markov Chain Model and Its Application to annual Rainfall Distribution for Crop Production. Submitted.