

Markov Chain Modelling of Pneumonia Cases in the Builsa North Municipality of the Upper East Region of Ghana

Abdulai Jafar, Alhassan Faisal*

Department of Statistics, Faculty of Physical Sciences, University for Development Studies, Tamale, Ghana

*Corresponding author: afaisal@uds.edu.gh

Received September 03, 2025; Revised October 05, 2025; Accepted October 12, 2025

Abstract The aim of the study is to model pneumonia cases using Markov Chain. Secondary data on monthly pneumonia cases spanning January, 2015 to December, 2020 was obtained from Builsa North Municipal hospital. Descriptive statistics of the data revealed the minimum and maximum number of pneumonia cases were 9 and 370 respectively with an approximate mean of 121. The data was classified into Low, Moderate and High states based on quartiles of the dataset. The analysis with the aid of R software revealed that, the transition probability of remaining in the low, moderate and high states were 61.10%, 62.9% and 50% respectively, but the probability of transitioning from the low state to the moderate and high states were 33.30% and 5.60% respectively. Also, in the long-run there will be higher chances of recording moderate pneumonia cases with a probability of approximately 51% as compared to 26% and 23% for low and high states respectively. Again, the estimates for Mean Recurrence Time for the low, moderate and high states were 4, 2 and 4 months, respectively and their respective sojourn times were 3, 3 and 2 months. The estimated first passage time probabilities also revealed that when low pneumonia cases are observed in the first month, there is higher probability of low pneumonia cases being observed in the next month. The analysis also revealed that it will take the states twenty-eight (28) months to be in equilibrium and in the long-run the municipality will record, on average, approximately 117 pneumonia cases monthly.

Keywords: Steady state probabilities, first-order Markov Chain, transition probability, mean recurrence time, sojourn time, first passage time probabilities

Cite This Article: Abdulai Jafar, and Alhassan Faisal, "Markov Chain Modelling of Pneumonia Cases in the Builsa North Municipality of the Upper East Region of Ghana." *American Journal of Applied Mathematics and Statistics*, vol. 13, no. 3 (2025): 47-56. doi: 10.12691/ajams-13-3-1.

1. Introduction

Pneumonia is an infection of the lungs that causes inflammation in the alveoli making it fill with fluid resulting in difficulty in breathing and limits oxygen intake and affects one or both lungs [1]. It is a contagious disease and a leading cause of death in the world among children below five years of age [2]. It kills an approximated number of 1.4 million children below the age of five years, accounting for about 18% of all children's deaths under five years old worldwide [3]. According to [4] there are about 120 million cases of pneumonia in children below five years per year and about 14 million progresses to severe cases. Also, [5] states that pneumonia causes death of children and families everywhere, but it is most prevalent in sub-Saharan Africa and South Asia. Furthermore, [6] has indicated that older people have higher risk of getting pneumonia and are more likely to die from it. Ventola indicated further that in the United States, pneumonia is the most common cause of hospitalization in adults and

that about one million adults seek care in hospitals every year and about 50,000 dies from this disease. According to [7], despite progress against pneumonia, over 800,000 children still die from it every year. However, by 2017, this number decreased by almost two-thirds due to enhancements in the major risk factors such as childhood wasting, falling global poverty, poor sanitation, air pollution and a better availability of health technology such as antibiotics and pneumococcal vaccines.

There are three types of pneumonia based on the causative pathogen which include; fungal pneumonia, bacterial pneumonia and viral pneumonia [8]. It is transmitted when germs from infected person spread to another person either through inhaling infected particles from coughs and sneezes of an infected person, when a person touches an infected surface or when an infected person coughs into his or her hand and then shakes an uninfected person who touches the mouth or eyes without washing [9]. Anyone can get pneumonia, however those with high risk of getting infected include; children less than 5 years of age, people 65 and above years of age and individuals with other medical conditions such as diabetes, respiratory diseases, malnourishment and immune

deficiency disorders. Smoking and alcoholism which reduces amount of white blood cells count in the bloodstream are also factors [10]. While most healthy children can fight the infection with their natural defenses, children with compromised immune systems as a result of malnutrition, especially due to insufficient breastfeeding and pre-existing illness such as measles increases a child's risk of contracting pneumonia [11].

In Ghana, pneumonia is rated as the leading cause of morbidity and mortality with an annual death record of 16,200 children, accounting for 20% of deaths per year [11]. The statistics is alarming and so there is the need for the reversal of the trend. This study therefore, seeks to model monthly pneumonia cases in Ghana using three-state discrete-time Markov chain, using Builsa North Municipality as a case study. Other authors such as [12,13,14,15,16] have also applied Markov chain in various disciplines of natural science, engineering, economics and management. The findings of this study can help policy makers institute appropriate intervention measures to minimize the prevalence of pneumonia in the Builsa North Municipality in particular and the country as a whole.

2. Materials and Methods

2.1. Source of Data

Secondary data on monthly pneumonia cases from January 2015 to December 2020 were obtained from the Builsa Municipal hospital.

2.2. Statistical Methods and Tools Used

Descriptive statistics of the data indicated the presence of over dispersion (where variance > mean) in the dataset, and so, the negative binomial (NB) regression was used, instead of the Poisson, to explore the effect of the months on the pneumonia cases since NB can directly estimate incidence rate ratios (IRR) from count data while accommodating over dispersion. If Y_t is the monthly pneumonia cases at month t and the months are treated as categorical predictor with one month as reference (say January), then the Negative Binomial Regression model will be written with dispersion parameter $\alpha > 0$ as $Y_t \sim NB(\mu_t, \alpha)$ with the log-linear predictor written as

$$\log(\mu_t) = \log(E_t) + \beta_0 + \sum_{m=2}^{12} \beta_m I(\text{month} = m) \quad (1)$$

Where E_t is the exposure at month t (population at risk $\times$ duration of the time unit, 1 month); $I(\text{month} = m)$ are dummy indicators for months 2, 3, ..., 12 (if month 1 is reference); β_m gives the log IRR of month m compared with the reference month; $\exp(\beta_m)$ is the incidence rate ratio (IRR) for month m versus reference.

Time series plot was also employed to ascertain the trend of the pneumonia cases over the period of the study. The monthly pneumonia cases were modeled using the

first discrete-time Markov chain analysis. R software and Excel were used for the analysis.

2.2.1. Markov Chain Modelling

A Markov chain is a stochastic model describing a sequence of possible events in which the probability of each event depends on the state of the previous event. Markov chain can be of first, second or higher orders. This chain possesses the memory-less property, which states that the probability of the next state of a system is only dependent on the present state of the system and not on any prior states [17]. A discrete-time process $X_t, t = 0, 1, 2, 3, \dots$ is said to be a Markov chain if for all $i, j, t, \dots, n$, the following holds:

$$\begin{aligned} P[X_t = j | X_{t-1} = i, X_{t-2} = \alpha, \dots, X_0 = \gamma] \\ = P[X_t = j | X_{t-1} = i] = P_{ij}(t) \end{aligned} \quad (2)$$

Where $P_{ij}(t)$ is known as state transition probability, which is a conditional probability that the process will be in state j at time t given that the process is in state i at time $t-1$. Thus, the current

state is sufficient to determine the next state. A homogenous Markov chain is one in which the transition probabilities are independent of time and thus, its transition probability is defined as

$$\begin{aligned} P[X_t = j | X_{t-1} = i, X_{t-2} = \alpha, \dots, X_0 = \gamma] \\ = P[X_t = j | X_{t-1} = i] = P_{ij} \end{aligned} \quad (3)$$

and p_{ij} holds for all times. The homogeneous transition probability satisfies these conditions;

$0 \leq p_{ij} \leq 1$ and $\sum_j p_{ij} = 1$. This follows that the states are mutually exclusive and exhaustive.

2.2.1.1. Transition Probability Estimation Using Maximum Likelihood Estimation (MLE)

The MLE method was used to estimate the transition probabilities, p_{ij} , which is, the conditional probability that a process will be in state j given that it is in state i . For a given initial state i and a number of trials n_i , the sample of transition counts, n_{i1}, n_{i2}, n_{i3} can be considered as a sample of size n_j from a trinomial distribution with probabilities p_{i1}, p_{i2} and p_{i3} such that $\sum_{j=1}^3 p_{ij} = 1$ [18]. The probability of this outcome can therefore be given as;

$$\begin{aligned} f(n_{ij} | p_{ij}) &= \frac{n_i!}{n_{i1}! n_{i2}! n_{i3}!} p_{i1}^{n_{i1}} p_{i2}^{n_{i2}} p_{i3}^{n_{i3}} \\ &= n_i! \prod_{i=1}^3 \frac{p_{ij}^{n_{ij}}}{n_{ij}}, i, j = 1, 2, 3 \end{aligned} \quad (4)$$

The likelihood of the transition Probabilities, p_{ij} follows a trinomial distribution given by;

$$L(n_{ij} | p_{ij}) = n_i! \prod_{i=1}^3 \frac{p_{ij}^{n_{ij}}}{n_{ij}}, i, j = 1, 2, 3$$

and the loglikelihood will therefore be given

$$l(p_{ij}) = \log \left[L(p_{ij} | n_{ij}) \right] = \log \left[n_i! \prod_{i=1}^3 \frac{p_{ij}^{n_{ij}}}{n_{ij}!} \right]$$

by

$$\Rightarrow l(p_{ij}) = \log n_i! + \log \prod_{i=1}^3 \frac{p_{ij}^{n_{ij}}}{n_{ij}!} = \log n_i! + \sum_{j=1}^3 \log \frac{p_{ij}^{n_{ij}}}{n_{ij}!}$$

$$= \log n_i! + \sum_{j=1}^3 n_{ij} \log p_{ij} - \sum_{j=1}^3 \log n_{ij}!$$

Maximizing $l(p_{ij})$ using Lagrange multiplier with the constraint $\sum p_i = 1$ then,

$$L(p_{ij}, \lambda) = l(p_{ij}) + \lambda \left(1 - \sum_{j=1}^3 p_i \right) \Rightarrow L(p_{ij}, \lambda)$$

$$= \log n_i! + \sum_{j=1}^3 n_{ij} \log p_{ij} - \sum_{j=1}^3 \log n_{ij}! + \lambda \left(1 - \sum_{j=1}^3 p_i \right)$$

Taking partial derivative with respect to p_{ij} and equating to zero,

$$\frac{\partial}{\partial p_{ij}} [L(p_{ij}, \lambda)] = \frac{\partial}{\partial p_{ij}} \left[\log n_i! + \sum_{j=1}^3 n_{ij} \log p_{ij} - \sum_{j=1}^3 \log n_{ij}! + \lambda \left(1 - \sum_{j=1}^3 p_i \right) \right] = 0$$

$$\Rightarrow \frac{\partial}{\partial p_{ij}} \log n_i! + \frac{\partial}{\partial p_{ij}} \sum_{j=1}^3 n_{ij} \log p_{ij} - \frac{\partial}{\partial p_{ij}} \sum_{j=1}^3 \log n_{ij}! + \frac{\partial}{\partial p_{ij}} \lambda - \frac{\partial}{\partial p_{ij}} \left(\lambda \sum_{j=1}^3 p_i \right) = 0$$

$$\Rightarrow \frac{n_{ij}}{p_{ij}} - \lambda = 0 \Rightarrow \frac{n_{ij}}{p_{ij}} = \lambda \Rightarrow p_{ij} = \frac{n_{ij}}{\lambda}$$

Now applying the constraint

$$\sum_{j=1}^3 p_{ij} = 1 \Rightarrow \sum_{j=1}^3 \frac{n_{ij}}{\lambda} = 1$$

$$= 1 \Rightarrow \frac{1}{\lambda} \sum_{j=1}^3 n_{ij} = 1 \Rightarrow \lambda = \sum_{j=1}^3 n_{ij} = n_i$$

Substituting back, we obtain:

$$p_{ij} = \frac{n_{ij}}{\lambda} = \frac{n_{ij}}{\sum_{j=1}^3 n_{ij}} = \frac{n_{ij}}{n_i} \quad (5)$$

2.2.1.2. State-transition Probability and Diagram

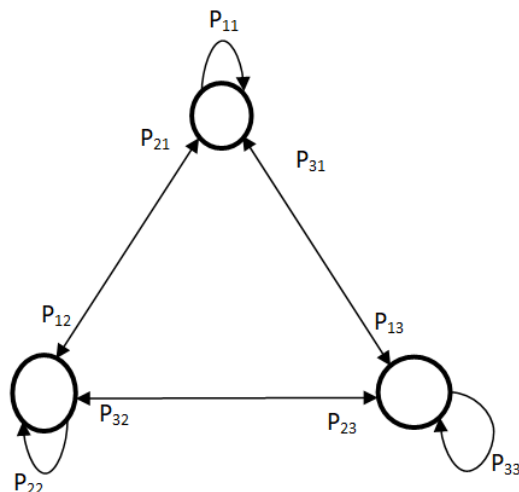


Figure 1. Three-state Transition Diagram

For a Markov chain of $n=1,2,3$ states with transition probabilities p_{ij} , then the transition probability matrix will

be $P = \begin{bmatrix} P_{11} & P_{12} & P_{13} \\ P_{21} & P_{22} & P_{23} \\ P_{31} & P_{32} & P_{33} \end{bmatrix}$ which must satisfy

$$p_{ij} \geq 0, 0 \leq p_{ij} \leq 1 \text{ and } \sum_j p_{ij} = 1 \text{ for } i, j = 1, 2, 3 \text{ (Ibe,}$$

2014). The corresponding transition diagram is shown in Figure 1 where the states are represented by circles and the transitions by arrows.

2.2.1.4. n-Step Transition Probability Matrix and Limiting State Transition Probability

An n -step transition probability $P_{ij}^{(n)}$ is the conditional probability of a system moving to state j after n transitions given that it is currently in state i , is provided for by Chapman-Kolmogorov equations defined as follows;

$$p_{ij}^{(n)} = p[X_n = j | X_0 = i] = \sum_k p[X_n = j, X_m = k | X_0 = i] \quad (6)$$

From the principle of total probability,

$$p_{ij}^{(n)} = \sum_k p[X_n = j | X_m = k, X_0 = i] p[X_m = k | X_0 = i] = \sum_k p_{kj}^{(n-m)} p_{ik}^{(m)} = \sum_k p_{ik}^{(m)} p_{kj}^{(n-m)}$$

for any $m=1,2,\dots,n$ and $k=1,2,\dots,n-1$ (7)

Equation (6) implies that the probability of going to state j from state i after the n^{th} transition is the product of the probability of first going from state i to an intermediary state k after m transitions and from state k to state j after $(n-m)$ transitions. The summing is done over all possible k states. For limiting state transition probability, if P is the transition probability matrix of a Markov chain, $X_n, n=1,2,3,\dots$. Then, the n^{th} transition probability matrix, is defined as

$$P^{(n)} = P^n = \begin{bmatrix} P_{11}^n & P_{12}^n & P_{13}^n & \dots \\ P_{21}^n & P_{22}^n & P_{23}^n & \dots \\ P_{31}^n & P_{32}^n & P_{33}^n & \dots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix} \quad (8)$$

When $n \rightarrow \infty$, the Markov chain becomes stationary where the state probabilities do not change with n any longer. If the limit of the chain exists, then the limiting state probability is given by;

$$\lim_{n \rightarrow \infty} P[X_n = j] = \pi_j, j = 1, 2, 3, \dots, N \quad (9)$$

If the limiting state probabilities exists without depending on the initial state, then,

$$\lim_{n \rightarrow \infty} P_{ij}^{(n)} = \lim_{n \rightarrow \infty} \sum_k P_{ik}^{(n)} P_{kj} = \lim_{n \rightarrow \infty} \sum_k \pi_k P_{kj} = \pi_j \quad (10)$$

Equation (9) can be interpreted as the long-run probabilities that the process is in state j . If $\pi' = [\pi_1, \pi_2, \pi_3, \dots, \pi_N]$ denotes the limiting-state probability vector, then the $\pi_j = \pi_k P_{kj}$; $\pi = \pi P$ and

$$\sum_j \pi_j = 1 \text{ holds.}$$

2.2.1.5. Sojourn Time (ST)

The expected time of stay of a process in a given state is referred to as sojourn time. If for state i with $P_{ii} > 0$ being the probability of the process remaining in the state for d time units and the random variable D_i is the number of time units the process remains in state i before leaving it, given that the process enters the state, then the probability mass function of D_i is defined as

$$P_{D_i}(d) = P[D_i = d] = P[X_0 = i, X_1 = i, X_2 = i, X_3 = i, \dots, X_{d-1} = i] = P_{ii}^{(d-1)}(1 - P_{ii}) \quad (11)$$

where $P_{D_i}(d)$ represents the probability that the sequence remains unchanged $d-1$ times before changing. Since the random variable D_i is geometrically distributed, then the expected sojourn time of the process in state i is defined as

$$E[D_i] = \frac{1}{1 - P_{ii}}, \text{ for } P_{ii} \neq 1 \quad (12)$$

However, if the state i is an absorbing state, then $P_{ii} = 1$ and $E[D_i] = \infty$.

2.2.1.6. First Passage Time (FPT) Probabilities

The FPT probability denoted by $f_{ij}(n)$ is the conditional probability that, the first time a process enters state j occurs in just n transitions, given that the process is currently in state i . Thus, the probability of first passage from state i to state j is defined by [19] as

$$f_{ij} = \sum_{n=1}^{\infty} f_{ij}(n) \quad (13)$$

2.2.1.7. The Mean Recurrence Time (MRT)

The MRT for a recurrence state i , is defined as the reciprocal of the stationary probability which can also be interpreted as the average time it takes for the chain to return to state i given that it was initially in state i and it is given by;

$$T_i = \frac{1}{\pi_i} \quad (14)$$

where π_i is the stationary probability of being in state i . If i is transient, T_i is considered infinite.

2.2.1.8 Long-Run Expectation

Consider a Markov chain with $n = 1, 2, 3, \dots$ states. If the limiting state probability vector and the mean values of the various states of the chain are, respectively, $\pi' = [\pi_1, \pi_2, \pi_3, \dots]$ and $\mu' = [\mu_1, \mu_2, \mu_3, \dots]$, then the long-run expected value of the chain is defined as

$$E[L_T] = \pi' \mu \quad (15)$$

2.3. Empirical Framework of the Study

2.3.1. Definition of States of the Pneumonia Data Set

The monthly pneumonia data collected from the Builsa Municipal hospital were grouped into three states as low (L), moderate (M) and high (H) based on the computed values of the first quartile (Q1) and upper quartile (Q3) of the pneumonia dataset. Monthly recorded cases less than or equal to the Q1 were considered low, moderate for cases between the Q1 and Q3 and high for cases greater than or equal to the Q3.

2.3.2. Verification of the First-order Markov Chain Assumption

The Chi square test of independence was used to verify whether the three states of the pneumonia cases satisfy the first-order Markov chain assumption before proceeding with the analysis. The assumption states that the future state of pneumonia cases depends only on its current state and that the transition probabilities are independent of time. The set hypothesis was;

H_0 : the first order Markov chain assumption does not hold for the data on monthly pneumonia cases

H_1 : the first-order Markov chain assumption holds for the data on monthly pneumonia cases

Table 1 below presents how the monthly pneumonia cases will be laid out. The entries n_{ij} representing the number of pneumonia cases in the i^{th} month belonging to the j^{th} state. We compared the estimated p-value to the α and reject H_0 if p-value $< \alpha$ and accept H_0 if otherwise.

Table 1. Frequency of Monthly Pneumonia Cases in Each State

States of Pneumonia Cases				
Month	Low (L)	Moderate (M)	High (H)	Row Total
1	n_{1l}	n_{1m}	n_{1h}	R_1
2	n_{2l}	n_{2m}	n_{2h}	R_2
3	n_{3l}	n_{3m}	n_{3h}	R_3
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
12	n_{12l}	n_{12m}	n_{12h}	R_{12}
Column Total	C_l	C_m	C_h	N

2.3.3. Estimation of State Transition Probabilities of Pneumonia Cases

The state space for the Markov Chain is $S = L, M, H$ with the assumption that the state of pneumonia cases for the next month is expected to depend only on the state of the pneumonia cases in the current month. The observed frequency of pneumonia cases n_{ij} from state i to state j

and $\sum_j n_{ij} = n_i$ are presented in Table 2.

Table 2. Transition Frequencies of Monthly Pneumonia Cases from State i to State j

		Current Month			
Previous Month		Low (L)	Moderate (M)	High (H)	Total
	Low (L)	n_{ll}	n_{lm}	n_{lh}	n_l
	Moderate (M)	n_{ml}	n_{mm}	n_{mh}	n_m
	High (H)	n_{hl}	n_{hm}	n_{hh}	n_h

The transition probabilities, p_{ij} for the various states were then estimated using (4) and the results will be displayed as in Table 3.

Table 3. Transition Probabilities of Pneumonia Cases from State i to State j

		Current Month		
		Low (L)	Moderate (M)	High (H)
Previous Month	Low (L)	p_{ll}	p_{lm}	p_{lh}
	Moderate (M)	p_{ml}	p_{mm}	p_{mh}
	High (H)	p_{hl}	p_{hm}	p_{hh}

The transition probability matrix can therefore be written as below

$$P = \begin{matrix} & \begin{matrix} L & M & H \end{matrix} \\ \begin{matrix} L \\ M \\ H \end{matrix} & \begin{bmatrix} p_{ll} & p_{lm} & p_{lh} \\ p_{ml} & p_{mm} & p_{mh} \\ p_{hl} & p_{hm} & p_{hh} \end{bmatrix} \end{matrix} \quad (16)$$

where $\sum_j p_{ij} = 1, j = l, m, h$ and p_{ij} is the probability of monthly pneumonia cases being in the j^{th} state given that it was in i^{th} state.

2.3.4. Estimation of Long-run Probabilities

If the long-run probabilities of low, moderate and high pneumonia cases are π_l, π_m and π_h , respectively, then (17) below holds;

$$\pi' P = \pi' \Rightarrow [\pi_l, \pi_m, \pi_h] \begin{bmatrix} p_{ll} & p_{lm} & p_{lh} \\ p_{ml} & p_{mm} & p_{mh} \\ p_{hl} & p_{hm} & p_{hh} \end{bmatrix} = [\pi_l, \pi_m, \pi_h] \quad (18)$$

where $\sum \pi_i = 1, i = l, m, h$ (Raheem et al, 2015).

2.3.5. Estimation of Stationary Transition Matrix

Let P be the transition probability matrix of the low, moderate and high states, then the matrix after the n^{th} transition P^n can be obtained by multiplication as follows;

$$P = \begin{bmatrix} p_{ll} & p_{lm} & p_{lh} \\ p_{ml} & p_{mm} & p_{mh} \\ p_{hl} & p_{hm} & p_{hh} \end{bmatrix}$$

$$P^2 = P \times P = \begin{bmatrix} p_{ll} & p_{lm} & p_{lh} \\ p_{ml} & p_{mm} & p_{mh} \\ p_{hl} & p_{hm} & p_{hh} \end{bmatrix}^2$$

$$P^3 = P \times P^2 = \begin{bmatrix} p_{ll} & p_{lm} & p_{lh} \\ p_{ml} & p_{mm} & p_{mh} \\ p_{hl} & p_{hm} & p_{hh} \end{bmatrix}^3$$

$$P^4 = P \times P^3 = \begin{bmatrix} p_{ll} & p_{lm} & p_{lh} \\ p_{ml} & p_{mm} & p_{mh} \\ p_{hl} & p_{hm} & p_{hh} \end{bmatrix}^4$$

$$P^n = P \times P^{n-1} = \begin{bmatrix} p_{ll} & p_{lm} & p_{lh} \\ p_{ml} & p_{mm} & p_{mh} \\ p_{hl} & p_{hm} & p_{hh} \end{bmatrix}^n$$

If after the n^{th} transition, the rows of the matrix become the same and does not change with n anymore, then P^n is the stationary (equilibrium) transition probability matrix.

2.3.6. Estimation of Long-run Expected Value

Let $\pi' = [\pi_l, \pi_m, \pi_h]$ and $\mu' = [\mu_l, \mu_m, \mu_h]$ represent the steady state probabilities and the expected values of the low, moderate and high states of the pneumonia cases. The average pneumonia cases in the long-run, $E[L_r]$ will be estimated as

$$E[L_r] = \pi' \mu \quad (19)$$

3. Results and Discussions

3.1. Descriptive Statistics

Table 4. Descriptive Statistics of All Recorded Pneumonia Cases in the Municipality

Statistic s	Mean	Variance	Min	Max	SD	CV
Value	120.54 2	5,675.85 7	9.00 0	370.00 0	75.33 8	62.49 9

Table 4 displays the mean, variance, minimum, maximum, standard deviation (SD) and coefficient of variation (CV) of the pneumonia cases recorded in the Builsa North Municipality during the period under study. It can be seen that the minimum and maximum number of pneumonia cases are 9 recorded in August and 370 recorded in March, respectively, with an overall approximate mean and variance of 121 and 5,676, respectively. Since the variance is greater than the mean, it suggests the presence of over-dispersion in the monthly pneumonia dataset.

Table 5. Descriptive Statistics of Monthly Recorded Pneumonia Cases in the Municipality

Month	Mean	CV	Min	Max
January	109.000	42.949	43.000	223.000
February	74.333	60.885	12.000	130.000
March	146.500	83.509	39.000	370.000
April	90.667	45.243	35.000	151.000
May	77.333	57.047	24.000	152.000
June	107.167	42.201	63.000	192.000
July	118.667	67.848	45.000	270.000
August	132.000	80.583	9.000	305.000
September	144.667	69.586	76.000	346.000
October	179.500	49.494	81.000	318.000
November	146.833	40.225	88.000	246.000
December	119.833	49.269	61.000	223.000

Table 5 displays the monthly descriptive statistics. Comparing the means of the monthly recorded cases from January to December, February recorded the least mean of approximately 75 and the month of October recorded the highest mean of 180 followed by the months of March and November each recording the mean of approximately 147. Also, March had the highest variability followed by August while November had the lowest variability of

pneumonia cases as shown by their coefficient of variations.

3.2. Time Series Plot of Pneumonia Cases

The time series plot of pneumonia cases is shown in Figure 2. It can be seen that pneumonia exhibited an increasing trend from May, 2015 to September, 2015 and thereafter, started to sharply decline to March, 2016. Again, there were fluctuations from April, 2016 till February, 2018. It increased in March, 2018 and dropped sharply in April, 2018. Fluctuations were then noticed thereafter but increased in October, 2018 and started to decline till January, 2019. It remained stable but fluctuates thereafter till December, 2020. Thus, the graph generally indicates a decreasing linear trend of the pneumonia cases over the period of study from 2015 to 2020.

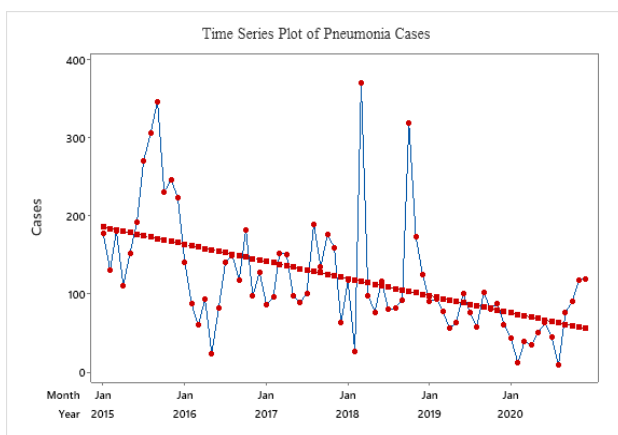


Figure 2. Time Series Plot of the Pneumonia Cases

Table 6. Monthly Effect on Pneumonia Incidence Rate

Month	IRR	SE	Z-value	P-value
January	1.202	0.387	0.570	0.567
February	0.820	0.265	-0.620	0.538
March	1.616	0.519	1.490	0.135
May	0.853	0.275	-0.490	0.622
June	1.182	0.380	0.520	0.603
July	1.309	0.421	0.840	0.402
August	1.456	0.468	1.170	0.242
September	1.596	0.512	1.460	0.146
October	1.980	0.635	2.130	0.033
November	1.619	0.520	1.500	0.133
December	1.322	0.425	0.870	0.386
Constant	90.667	20.651	19.790	0.000
Chi Square	2176.950			

To explore the effect of the months on the pneumonia cases, the negative binomial regression model was fitted since the monthly pneumonia count dataset exhibited the presence of over dispersion. The results of the fitted negative binomial regression are displayed in Table 6 which presents the estimates of the effect of the months on the incidence rate ratios (IRR) of pneumonia cases with April as the reference month. April was used as the reference month because April marks the end of the dry

season and as well the beginning of the rainy season in the Builsa North Municipality. The p-values revealed that only the month of October was significant with reference to the month of April. Also, when other months are held constant, the pneumonia cases incidence rates were expected to increase for all the Months with the exception of February and May as compared to the month of April. October had the highest expected incidence rate of pneumonia cases since it was expected to have 1.98 times greater pneumonia incidence rate compared to that of April. The overall goodness of fit test was tested using the likelihood ratio test which tests whether the months are really significant in explaining the IRR. We can conclude that the months collectively explain the variations in the rates since the likelihood ratio chi square test value of 2,176.950 (indicated at the bottom) is large.

3.3. Markov Chain Modelling of Pneumonia Cases

3.3.1. Identification of the three States for the Pneumonia Data

The Low (L), Moderate (M) and High (H) states were identified using the first and third quartiles of the pneumonia dataset. A low state was defined as a state of months of pneumonia cases less than or equal to the first quartile (approximately 76) whilst cases greater than or equal to the third quartile (approximately 152) were defined as high and finally, cases recorded between the 76 and 152 were defined as moderate. The transition frequencies from one state to another of pneumonia cases were observed and compiled.

3.3.2. Identification of Transition Frequencies for the three States of the Pneumonia Data

The observed transition frequencies between the three states of the pneumonia cases are presented in Table 7. The entries in the table denote the number of pneumonia cases that are in a particular state in the current month given that the pneumonia cases were in a particular state in the previous month. For instance, it can be seen that, given that the pneumonia cases were in the low state in the previous month, the transition frequencies with which the cases will make to the current month in the low, moderate and high states were 11, 6 and 1, respectively. Other rows in the table can be interpreted appropriately in a similar fashion. It can therefore, be deduced from Table 7 that when low pneumonia cases are recorded in the previous month then, there is high possibility that low pneumonia cases will be recorded in the current month. Also, when moderate pneumonia cases are recorded in the previous month then there is a high likelihood that moderate pneumonia cases will be recorded in the current month and when high pneumonia cases are recorded in the previous month then, there is high likelihood that high pneumonia cases will be recorded in the current month. Therefore, when measures are put in place to reduce pneumonia cases to the low state then, there is high probability of recording low pneumonia cases.

Table 7. Transition Frequencies of Monthly Pneumonia Cases

		Current Month			
Previous Month		Low (L)	Moderate (M)	High (H)	Total
	Low (L)	11	6	1	18
	Moderate (M)	6	22	7	35
	High (H)	1	8	9	18

3.3.3. Verification of the First-order Markov Chain Assumption

The first-order Markov chain assumed that the state of pneumonia cases for the next month is expected to depend only on the state of the pneumonia cases in the current month. The pneumonia dataset was tested to investigate whether this assumption holds for the three states by performing chi square test of independence. Using the data in Table 7, the computed chi square test statistic yielded a value of 22.44 with p-value of 0.0000, indicating that the first-order Markov chain assumption holds for the states of the pneumonia cases.

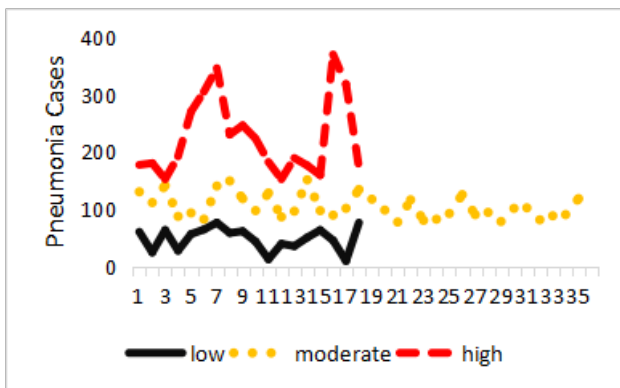
3.3.4. Estimation of Correlation between the States of the Pneumonia Cases

The correlation between the low, moderate and high states of the pneumonia cases was investigated using correlation coefficients as shown in Table 8 below. It can be seen that there exists positively weak correlation between the states since their correlation coefficients ranges from 0.1 to 0.3 approximately. This means that an increase in pneumonia cases in a state does not necessarily leads to an increment of pneumonia cases in another state.

Table 8. Correlation Between the Three States of the Pneumonia Cases

State	Low	Moderate	High
Low	1.0000		
Moderate	0.0913	1.0000	
High	0.2059	0.2599	1.0000

The pattern of correlation between the low, moderate and high states of the monthly pneumonia cases is shown by Figure 3. It can be seen from the figure 3 that an increment or decrement in any of the states shows no significant change in the other states.

**Figure 3. Correlation Between the three States of the Pneumonia Cases**

3.3.5. Estimation of State Transition Probabilities for Pneumonia Cases

The transition probabilities, p_{ij} were estimated using (4), where n_{ij} is the number of pneumonia cases in the j^{th} state given that it was in i^{th} state; and $n_i = \sum_{j=1}^3 n_{ij}$ = total number of pneumonia cases in the i^{th} state. Using the information in Table 7, the transition probability matrix P can be extracted from the calculated transition probabilities of the pneumonia cases from one state to another as follows:

$$P = \begin{matrix} & \begin{matrix} L & M & H \end{matrix} \\ \begin{matrix} L \\ M \\ H \end{matrix} & \begin{bmatrix} 0.6110 & 0.3330 & 0.0560 \\ 0.1710 & 0.6290 & 0.2000 \\ 0.0560 & 0.4440 & 0.5000 \end{bmatrix} \end{matrix}$$

It can be seen for instance that the transition probability of remaining in the low, moderate and high state of pneumonia cases were 61.10%, 62.9% and 50% respectively. Also, the probability of transitioning from the low state to the moderate and high states were 33.30% and 5.60%, respectively. Similarly, it can be inferred that there is high probability of remaining in a particular state of the pneumonia cases than moving to a different state. This means that if preventive measures of pneumonia are adhered to reduce the prevalence of pneumonia cases, then the likelihood of pneumonia cases remaining low would be high. It is therefore prudent for stakeholders to create awareness of the lethal impact of pneumonia as well put in place appropriate preventive measures so as to reduce pneumonia cases in the Builsa North municipality.

3.3.6. Estimation of Steady State Probabilities, Mean Recurrence and Sojourn Times

Table 9 shows the estimated values of the Steady State Probabilities (SSP), Mean Recurrence Time (MRT) and Sojourn Time (ST) for the three states. The SSP values indicate that in the long-run, there are higher chances of pneumonia cases to be in the moderate state than the low and high states. Also, the MRT values show that it takes approximately 4, 2 and 5 months for pneumonia cases to return to the low, moderate and high states after leaving those states. Furthermore, from the estimated ST values, the length of time pneumonia cases is expected to stay in the low and moderate states were estimated to be approximately 3 months each and 2 months for the high state. Hence the complete cycle of the pneumonia cases in the Builsa North municipality was estimated to be 8 months.

Table 9. The SSP, MRT and ST of Pneumonia Cases for the Builsa North Municipality

State	SSP	MRT	ST
Low	0.258	3.883	3
Moderate	0.510	1.962	3
High	0.233	4.297	2

3.3.7. Estimation of Stationary (Equilibrium) Transition Matrix of the Pneumonia States

It was known from the estimation that it will take the states at least twenty-eight (28) months to be in equilibrium with estimated probabilities for the low, moderate and high states as 0.2575645, 0.5097059 and

0.2327296 respectively.

$$\begin{aligned}
 P &= \begin{matrix} & \begin{matrix} L & M & H \end{matrix} \\ \begin{matrix} L \\ M \\ H \end{matrix} & \begin{bmatrix} 0.6110 & 0.3330 & 0.0560 \\ 0.1710 & 0.6290 & 0.2000 \\ 0.0560 & 0.4440 & 0.5000 \end{bmatrix} \end{matrix} \\
 P^2 &= \begin{matrix} & \begin{matrix} L & M & H \end{matrix} \\ \begin{matrix} L \\ M \\ H \end{matrix} & \begin{bmatrix} 0.43340 & 0.437784 & 0.128816 \\ 0.22324 & 0.541384 & 0.235376 \\ 0.13814 & 0.519924 & 0.341936 \end{bmatrix} \end{matrix} \\
 P^3 &= \begin{matrix} & \begin{matrix} L & M & H \end{matrix} \\ \begin{matrix} L \\ M \\ H \end{matrix} & \begin{bmatrix} 0.3468822 & 0.4768826 & 0.1762352 \\ 0.2421574 & 0.5193764 & 0.2384662 \\ 0.1924590 & 0.5248524 & 0.2826886 \end{bmatrix} \end{matrix} \\
 P^4 &= \begin{matrix} & \begin{matrix} L & M & H \end{matrix} \\ \begin{matrix} L \\ M \\ H \end{matrix} & \begin{bmatrix} 0.3033611 & 0.4937194 & 0.2029195 \\ 0.2501256 & 0.5132052 & 0.2366692 \\ 0.2231727 & 0.5197347 & 0.2570925 \end{bmatrix} \end{matrix} \\
 P^{27} &= \begin{matrix} & \begin{matrix} L & M & H \end{matrix} \\ \begin{matrix} L \\ M \\ H \end{matrix} & \begin{bmatrix} 0.2575645 & 0.5097059 & 0.2327296 \\ 0.2575645 & 0.5097059 & 0.2327296 \\ 0.2575644 & 0.5097059 & 0.2327296 \end{bmatrix} \end{matrix} \\
 P^{28} &= \begin{matrix} & \begin{matrix} L & M & H \end{matrix} \\ \begin{matrix} L \\ M \\ H \end{matrix} & \begin{bmatrix} 0.2575645 & 0.5097059 & 0.2327296 \\ 0.2575645 & 0.5097059 & 0.2327296 \\ 0.2575645 & 0.5097059 & 0.2327296 \end{bmatrix} \end{matrix}
 \end{aligned}$$

3.3.8. Estimation of First Passage Time (FPT) Probabilities for any given State of Cases

3.3.8.1. The FPT Probabilities Assuming the Current State of Pneumonia Cases is Low

The FPT probabilities for the low, moderate and high states assuming the cases are in the low state were estimated and the results are displayed in Table 10. It can be observed for instance that if low pneumonia cases are observed in the first month, then the probability of observing low, moderate and high pneumonia cases in the next month will be 0.6110, 0.3330 and 0.0560, respectively. This implies that assuming low pneumonia cases are observed in the first month then, there is higher probability of low pneumonia cases being observed in the next month. It can generally be observed that from the first to the fourth months when low pneumonia cases are observed then there is high probability of observing moderate pneumonia cases in the next month than observing high pneumonia cases and from the fifth to the twelfth months there is high probability of observing high pneumonia cases than moderate pneumonia cases. This finding could be attributed to the fact that the first to fourth months constitutes the hot season and from fifth to the tenth month constitutes the rainy season and from the eleventh to twelfth months onward constitutes the harmattan season characterized by coldness and mosquito-borne infections in the municipality.

Table 10. FPT Probabilities Assuming the current state of Pneumonia Cases is Low

Month	States		
	Low	Moderate	High
1	0.6110	0.3330	0.0560
2	0.0601	0.2280	0.1008
3	0.0454	0.1530	0.1067
4	0.0377	0.1009	0.0993
5	0.0323	0.0659	0.0882
6	0.0279	0.0427	0.0768
7	0.0243	0.0275	0.0664
8	0.0211	0.0176	0.0572
9	0.0183	0.0113	0.0492
10	0.0160	0.0072	0.0423
11	0.0139	0.0046	0.0363
12	0.0120	0.0029	0.0312

3.3.8.2. The FPT Probabilities Assuming the Current State of Pneumonia Cases is Moderate

The FPT probabilities for the low, moderate and high states were also estimated assuming the cases are in the moderate state and the results are displayed in Table 11. It can be seen, for instance, that in the first month, given that the current pneumonia cases are in the moderate state, the probability of observing low, moderate and high pneumonia cases in the next month will be 0.1710, 0.6290 and 0.2000, respectively. This implies that when current pneumonia cases in the first month are moderate, then there is higher probability of having moderate pneumonia cases in the next month than low and high pneumonia cases in the municipality. It can generally be observed that when current pneumonia cases are moderate, then there is high probability of observing high pneumonia cases than low pneumonia cases in the next month from the first to the fourth months, whilst from the fifth to the twelfth months the likelihood of observing low pneumonia cases is higher than observing high pneumonia cases.

Table 11. FPT Probabilities Assuming the current state of Pneumonia Cases is Moderate

Month	State		
	Low	Moderate	High
1	0.1710	0.6290	0.2000
2	0.1188	0.1457	0.1354
3	0.0955	0.0872	0.1024
4	0.0810	0.0528	0.0826
5	0.0699	0.0323	0.0690
6	0.0606	0.0199	0.0585
7	0.0527	0.0124	0.0499
8	0.0458	0.0077	0.0427
9	0.0398	0.0048	0.0367
10	0.0346	0.0030	0.0315
11	0.0301	0.0019	0.0270
12	0.0262	0.0012	0.0232

3.3.8.3. The FPT Probabilities Assuming the Current State of Pneumonia Cases is High

The FPT probabilities for the low, moderate and high

states assuming the cases are in the high state were also estimated and the results are displayed in Table 12. It can, for instance, be seen that, given that the current pneumonia cases in the Builsa North Municipality are high, the probability of observing low, moderate and high pneumonia cases in the next month will be 0.0560, 0.0444 and 0.5000, respectively. This means that in the first month there is higher probability of recording high pneumonia cases in the municipality in the next month given that the municipal currently has high pneumonia cases. It can generally be observed in the Table 12 that, with the exception of the second and third months, there is high probability of observing low pneumonia cases than moderate pneumonia cases in the next month in the municipality when pneumonia cases are in the high state.

Table 12. FPT Probabilities Assuming the current state of Pneumonia cases is High

Month	State		
	Low	Moderate	High
1	0.0560	0.0444	0.5000
2	0.1039	0.2406	0.0919
3	0.1047	0.1331	0.0658
4	0.0947	0.0751	0.0514
5	0.0833	0.0432	0.0423
6	0.0727	0.0253	0.0356
7	0.0633	0.0150	0.0303
8	0.0550	0.0091	0.0258
9	0.0478	0.0055	0.0222
10	0.0416	0.0034	0.0190
11	0.0361	0.0021	0.0163
12	0.0314	0.0013	0.0140

3.3.9. Estimation of Long-run Expected Value

The estimated long-run (steady state) probabilities vector of the three states of the pneumonia cases is $\pi' = [\pi_l, \pi_m, \pi_h] = [0.258, 0.510, 0.233]$ and that of the mean vector is $\mu' = [\mu_l, \mu_m, \mu_h]$. Therefore,

the expected number of pneumonia cases in the municipality in the long-run was estimated

$$\text{as } E[L_r] = \pi' \mu = [0.258, 0.510, 0.233] \begin{bmatrix} 47.833 \\ 103.629 \\ 220.579 \end{bmatrix}. \text{ This implies that}$$

$$= 116.458 \approx 117$$

the Builsa North Municipality in the long-run would record an average pneumonia case of approximately 117 monthly.

4. Conclusion

The main objective of the study was to model the pneumonia data using the first-order Markov chain modeling. Descriptive statistics of the data revealed that the minimum and maximum number of pneumonia cases were 9 and 370, respectively. The mean and variance were approximately 121 and 5,676, respectively, suggesting the presence of over-dispersion. As a result, the negative binomial regression model was fitted to explore the effect of the months on the incidence rate ratios (IRR) of

pneumonia cases using April as the reference. The fitted model revealed that only the month of October was significant and when all other months are held constant. The pneumonia cases incidence rates were expected to increase for all the Months with the exception of February and May. Also, the time series plot of Malaria cases revealed a decreasing non-seasonal trend of the pneumonia cases over the period of study. The chi square test of independence confirmed that the pneumonia data satisfied the first-order Markov chain assumption.

The data was classified into Low (L), Moderate (M) and High (H) states based on quartiles of the dataset. The analysis revealed that, the transition probability of remaining in the low, moderate and high states were 61.10%, 62.9% and 50%, respectively. However, the probability of transitioning from the low state to the moderate and high states were 33.30% and 5.60% respectively. Also, the probabilities of moving from the moderate state to the low state and from the high state to the low state were 17.10% and 5.60% respectively. The study also discovered that in the long-run there will be higher chances of recording moderate pneumonia cases in the municipality with a probability of approximately 51% as compared to low and high pneumonia cases with probabilities of 26% and 23% respectively.

Furthermore, the Mean Recurrence Time revealed that it takes the pneumonia cases shorter time to return to the moderate state than to the low and high states. Also, the expected length of time pneumonia cases will stay in the low and moderate states were estimated to be approximately 3 months each and 2 months for the high state. The estimated first passage time probabilities also revealed that when low pneumonia cases are observed in the first month, there is higher probability of low pneumonia cases being observed in the next month. Moreover, it will take the states twenty-eight (28) months to be in equilibrium and in the long-run the municipality will record on average approximately 117 pneumonia cases monthly.

5. Recommendations

From the findings of the study, the following recommendations are made;

- i. Since there is high probability of pneumonia cases being in the same state as its current state, it is recommended that adequate attention be given to pneumonia by the government and other agencies by putting in place more efficacious measures and ensuring their adherence to minimize pneumonia cases in the municipality and the country at large.
- i i. Also, public health education be organized to create awareness of the lethal impact of pneumonia especially on children so as to intensify compliance in order to minimize the prevalence of pneumonia infections.
- i i i. Women should be encouraged to practice exclusive breast feeding and reducing malnutrition, especially in under-5 children, which weakens immunity.
- iv. Reducing overcrowding in rooms and improving ventilation indoors, especially during rainy or cold

- seasons.
- v. Community health workers should be properly trained to identify and refer pneumonia cases early.
 - vi. Policy makers in the health sector should always work with routine data (e.g., monthly incidence, seasonal peaks) to anticipate outbreaks and allocate resources.
 - vii. Furthermore, future studies can consider other states such as susceptibility, infection and recovery states of pneumonia as well as gender in estimating the relevant metrics of pneumonia
 - viii. Future studies should consider other statistical models to estimate relevant epidemiological quantities of pneumonia and comparisons made to ascertain which one would be most appropriate.

Statement of Competing Interests

The authors have no competing interests (financial or others) to disclose.

References

- [1] World Health Organization, "Pneumonia", 2019 . <https://www.who.int/news-room/fact-sheets/detail/pneumonia>. [Accessed Dec. 6, 2020].
- [2] Le Roux, D. M., Zar, H. J "Community-acquired pneumonia in children—a changing spectrum of disease", *Pediatric radiology*, 47, 1392-1398, 2017.
- [3] Mhandu, E, *Building a low-cost biomedical device to improve accuracy in pneumonia diagnosis in under five children*. Dissertation, Ashesi University. 2019.
- [4] Lassi, Z. S., Imdad, A. and Bhutta, Z. A, "Short-course versus long-course intravenous therapy with the same antibiotic for severe community-acquired pneumonia in children aged two months to 59 months", *Cochrane Database of Systematic Reviews*, 2017.
- [5] Pant, A., Jain, A., Nayak, K.C., Gandhi, D. and Prasad, B.G "Pneumonia detection: An efficient approach using deep learning", In 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT) IEEE - 49239.
- [6] Ventola, C. L, "Immunization in the United States: recommendations, barriers, and measures to improve compliance", *Part 1: childhood vaccinations. Pharmacy and Therapeutics*, 41:426-436, 2016.
- [7] Swedberg E, Shah R, Sadruddin S, Soeripto J (2020) Saving young children from forgotten killer: pneumonia. *Am J Physiol Lung Cell Mol Physiol*.
- [8] Jill, S.S. (2018). Why pneumonia can be deadly for some people. <https://www.healthline.com/health/pneumonia/can-you-die-from-pneumonia#risk>. [Accessed Dec. 10, 2020].
- [9] Zawn, V. (2017). "Is Pneumonia contagious? Causes and Transmission" *MedicalRecords.com*. [Accessed November 25, 2020].
- [10] Backhaus, E., Berg, S, Andersson, R., Ockborn, G., Malmström, P., Dahl, M., ...and Trollfors, B, "Epidemiology of invasive pneumococcal infections: manifestations, incidence and case fatality rate correlated to age, gender and risk factors", *BMC infectious diseases*, 16,1-12, 2016.
- [11] Iddrisu, A. K., Alhassan, A. and Amidu, N, "Survival analysis of birth defect infants and children with pneumonia mortality in Ghana", *Hindawi Advances in Public Health*, 1-7, 2019.
- [12] Datong, G.M, "A Markov Chain Model Analysis of GSM Network Service Providers of Marketing Mix", *International Journal of Engineering & Technology IJET-IJENS*, 11(04), 38 – 43, August 2011.
- [13] Kassa, A.M., Abrham, E. and Seid, T. "Application of Markov Chain Analysis Model for Predicting Monthly Market Share of Restaurants", *International Journal of Recent Engineering Research and Development*, 2, 48-55, 2017.
- [14] Liu, X., Ren, L., Yuan, F., Yang, B. and Nanjing, P.R. (2009), "Meteorological drought Forecasting using Markov Chain model" In 11th International Conference on Environmental Science and Information Application Technology, Chania, Crete, Greece, pp 23–26 <http://toc.proceedings.com/18229webtoc.pdf>. [Accessed November 10, 2020].
- [15] Pobb, M.A, *A three-state Markov Chain Model of plasmodium falciparum parasitemia transmission in Ghana*, Dissertation, Kwame Nkrumah University of Science and Technology, 2012.
- [16] Twumasi, C. L, *Markov Chain Modeling of HIV, Tuberculosis and Hepatitis-B Transmission: A Study of a Regional Hospital in Ghana*. Dissertation, University of Ghana, 2018.
- [17] Grimmett, G. R. and Stirzaker, D. R, *Probability and Random Processes*. Oxford University Press Inc., New York, 2001.
- [18] Teodorescu, I, "Maximum likelihood estimation for Markov Chains", 2009.
- [19] Ibe, O. C, *Fundamentals of Applied Probability and Random Processes*. Elsevier Academic Press, Boston, 2005.



© The Author(s) 2025. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).