

# A Time Series Modeling of the Morbidity Incidence of Pneumocystis Pneumonia among Farmers in Benue State, Nigeria

Peter Ogbeh<sup>1</sup> and David Adugh Kuhe<sup>2\*</sup>

<sup>1</sup>Department of Mathematics and Computer, Benue State University Makurdi, Benue State, Nigeria

<sup>2</sup>Department of Statistics, Joseph Sarwuan Tarka University, Makurdi, Benue State, Nigeria

\*Corresponding Author

DOI: <https://doi.org/10.51584/IJRIAS.2025.1005000104>

Received: 04 May 2025; Accepted: 16 May 2025; Published: 18 June 2025

## ABSTRACT

Pneumocystis pneumonia (PCP) is a severe opportunistic infection that poses significant public health challenges, particularly among immunocompromised individuals, necessitating accurate modeling and forecasting for effective disease control and prevention. This study aims to identify an optimal Autoregressive Integrated Moving Average (ARIMA) model for accurately predicting short-term trends in Pneumocystis pneumonia (PCP) infection cases in Benue State, Nigeria. Monthly time series data on PCP cases from January 2010 to December 2023 were analyzed. The stationarity properties of the data were examined using time series plots and the Augmented Dickey-Fuller (ADF) unit root test, which confirmed that the series is integrated of order one,  $I(1)$ . Following the Box-Jenkins methodology, an ARIMA (p,d,q) model was applied to the data. The results indicate that the ARIMA (5,1,2) model provided the best fit for modeling and forecasting PCP infection cases. The study identified a six-month infection cycle among the population, characterizing PCP as a chronic and potentially life-threatening condition if not properly managed. The selected ARIMA (5,1,2) model demonstrated dynamic stability and accounted for 76.16% of the variance in the data. It was subsequently used to generate short-term forecasts for 24 months (January 2024-December 2025). The projections reveal a fluctuating yet increasing trend in PCP cases, with an average of 698 infections per month. A forecast reliability test, comparing observed and predicted values, confirmed that the forecasted results were valid, accurate, and suitable for informing policy decisions. To enhance PCP infection control in Benue State, the study recommends that authorities should strengthen surveillance, improve early diagnosis and treatment, implement targeted public health interventions, utilize forecasting models for resource allocation, and encourage further research for improved predictive accuracy.

**Keywords:** Pneumocystis Pneumonia, Opportunistic Infections, Farmers, Public health, ARIMA model, Benue State, Nigeria

## INTRODUCTION

Pneumocystis pneumonia (PCP) is a life-threatening fungal infection caused by *Pneumocystis jirovecii*, primarily affecting individuals with compromised immune systems such as those living with HIV/AIDS, undergoing chemotherapy, or receiving immunosuppressive treatment for autoimmune diseases or organ transplants (Iriart, 2015; Chiliza *et al.*, 2020; Muñoz *et al.*, 2020). Initially identified in malnourished infants in Central Europe during World War II, PCP later emerged as a defining illness during the 1980s HIV epidemic (Roux *et al.*, 2014; Olugbenga *et al.*, 2020). Although highly active antiretroviral therapy (HAART) has reduced its prevalence among HIV patients, PCP remains a major opportunistic infection and a serious public health issue globally (Chiliza *et al.*, 2020).

PCP symptoms typically include a dry cough, progressive dyspnea, mild to moderate fever, and non-pleuritic or pleuritic chest pain (O'Donnell *et al.*, 2018; Chiliza *et al.*, 2020). Diagnosis often requires bronchoscopy with bronchoalveolar lavage, as *Pneumocystis* cannot be cultured (Chen *et al.*, 2019). First-line treatment and

prophylaxis usually involve trimethoprim-sulfamethoxazole, although drug resistance is an increasing concern (Muñoz *et al.*, 2020). Advances in molecular techniques have improved understanding of the organism's transmission and resistance mechanisms.

Though PCP is most strongly linked with HIV/AIDS and low CD4+ T-cell counts, where it often serves as an AIDS-defining illness (Stern, 2014; Broadhurst, 2021) its occurrence among HIV-negative individuals is rising. This is largely due to increasing use of immunosuppressive therapy for conditions like cancers, autoimmune diseases, and organ transplants (Iriart, 2015; Muñoz *et al.*, 2020). PCP remains the most common opportunistic infection in AIDS patients, affecting an estimated 3% to 15% of individuals with poorly controlled or untreated HIV (Polaczek, 2014).

Emerging evidence suggests that certain occupational groups, such as farmers, may be at elevated risk due to environmental exposures like dust, pesticides, and organic matter (Zahradnik *et al.*, 2016; Wolf *et al.*, 2020). These exposures, often combined with pre-existing respiratory conditions such as asthma or COPD both more common in farming populations increase susceptibility to respiratory infections like PCP (Ji *et al.*, 2016; Boubaker *et al.*, 2019). Farmers also report more respiratory symptoms and worse lung function compared to non-farmers (Ji *et al.*, 2016), which may reflect cumulative occupational hazards.

In regions such as Benue State, Nigeria, the burden of PCP among farmers is underreported and underexplored, despite a growing number of cases (Olugbenga *et al.*, 2020). Environmental and occupational conditions in this region may predispose farmers to infection, yet a lack of epidemiological data hinders timely diagnosis, prevention, and intervention efforts. There is thus a pressing need to investigate PCP as a potential occupational hazard in agricultural settings.

To address this gap, this study aims to model and forecast the morbidity incidence of PCP among farmers in Benue State using time series approaches, specifically the Autoregressive Integrated Moving Average (ARIMA) model (Box *et al.*, 2015). ARIMA models are effective for identifying trends, ensuring stationarity, and forecasting future events based on historical data. Such predictive tools are crucial for enabling health systems to plan interventions and allocate resources effectively.

Several empirical evidence regarding the subject matter are well documented in literature, for example, Bruns *et al.* (2022) conducted a retrospective analysis of *Pneumocystis pneumonia* (PCP) trends in Germany using hospital and national discharge data from 2014 to 2019. The incidence increased from 2.3 to 2.6 cases per 100,000 populations, driven predominantly by a rise in non-HIV-associated PCP and related mortality. However, the study was limited by its exclusive focus on hospitalized patients and the absence of causal data. These findings point to a need for further investigation into the drivers of non-HIV PCP incidence. Similarly, Koo *et al.* (2021) retrospectively reviewed 39 cases of *Pneumocystis jirovecii* pneumonia (PJP) among Korean patients with rheumatic diseases from 2005 to 2019. The incidence was 0.41 per 1,000 patient-years, with most cases occurring in women with rheumatoid arthritis. Fever and dyspnea were the predominant symptoms, and the study reported a mortality rate of 12.8%. The findings underscore the importance of prophylaxis and early recognition in immunocompromised patients to prevent adverse outcomes.

In China, Yang *et al.* (2021) assessed 48 cases of PJP in kidney transplant recipients across multiple centers between 2010 and 2020. The incidence was 1.2%, with fever and cough as the most frequently reported symptoms. Trimethoprim-sulfamethoxazole (TMP-SMX) was the primary treatment, and the study recorded a 14.6% mortality rate. These results highlight the susceptibility of transplant recipients to PJP and the critical need for early detection and preventive strategies. Similarly, Chen *et al.* (2021) reviewed 82 PCP cases in solid organ transplant recipients across Australia and New Zealand from 2000 to 2019. While overall PCP incidence declined over the study period, an increase was observed among non-lung transplant recipients, particularly those with kidney transplants. Despite its valuable insights, the study did not include data beyond 2019 and was limited to transplant populations, emphasizing the necessity for continuous surveillance and updated prophylactic protocols.

Zhang *et al.* (2023) examined 42 PCP cases in patients with rheumatoid arthritis-associated interstitial lung disease (RA-ILD) at a rheumatology center in China (2008-2018). Fever and cough were the most prevalent

symptoms, TMP-SMX remained the treatment of choice, and the study reported a mortality rate of 19%. Though rare, PCP in RA-ILD patients represents a significant clinical concern, reinforcing the need for timely diagnosis and targeted management. In South Africa, Chiliza *et al.* (2020) investigated 124 HIV-associated PCP cases from 2004 to 2015. Most patients had severely depleted CD4 counts, and tuberculosis co-infection was common. Notably, mortality among ICU-admitted patients reached 61.9%. The study highlights the devastating outcomes of PCP in the context of advanced HIV and calls for improved clinical management and up-to-date epidemiological surveillance.

## MATERIALS AND METHODS

### Source of Data

The data used in this study comprises serological confirmed cases of Pneumocystis Pneumonia infection (PCP) cases in Benue state of Nigeria from January, 2010 to December, 2023. The data consists of 168 monthly observations of pneumocystis pneumonia infection. The data was obtained as secondary data from Benue State Epidemiological Unit, Makurdi. To reduce and stabilize the mean and variance, the original data on pneumocystis pneumonia infection among farmers in the study area was transformed to natural logarithm through the following formula:

$$Y'_t = \ln Y_t \quad (1)$$

where  $Y_t$  is the current Pneumonia infection at time  $t$ , while  $\ln$  is the natural logarithm.

### METHODS OF DATA ANALYSIS

The following statistical tools are employed in the analysis of data in this work. Let  $\{Y_t\}$  be a stochastic time series process.  $\{Y_t\}$  is defined as a sequence of monthly confirmed cases of pneumocystis pneumonia infection indexed by time and shall be used to refer to a series throughout this study.

### Preliminary tests

The following preliminary tests such as descriptive statistics and normality test and unit root test are employed before model specification.

### Descriptive statistics and Jarque-Bera test of normality

The mean of the monthly confirmed cases of pneumocystis pneumonia infection is computed as:

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i \quad (2)$$

The sample standard deviation of the monthly confirmed cases of pneumocystis pneumonia infection over a given period of time is computed using the following formula:

$$\hat{\sigma} = \sqrt{\frac{1}{n-1} \sum_{t=1}^n (y_t - \bar{y})^2} \quad (3)$$

where  $\bar{y}$  is the sample mean defined in (2) and  $n$  is the sample size.

Jarque-Bera test is a normality test of whether a given sample data have the skewness and kurtosis similar to that of a normal distribution. The test was proposed by Jarque and Bera (1980, 1987) and tests the null hypothesis that the series is normally distributed. Given a series  $\{Y_t\}$  the JB test statistic is defined as:

$$JB = \frac{T}{6} \left( g_1^2 + \frac{1}{4} (g_2 - 3)^2 \right) \quad (4)$$

where T is the number of observations,  $g_1$  is the sample skewness which is estimated by:

$$g_1 = \frac{\mu_3}{\mu_2^{3/2}} = T^{1/2} \frac{\sum_{t=1}^T (y_t - \bar{y})^3}{\left( \sum_{t=1}^T (y_t - \bar{y})^2 \right)^{3/2}} \quad (5)$$

and  $g_2$  is the sample kurtosis which is estimated by:

$$g_2 = \frac{\mu_4}{\mu_2^2} = T \frac{\sum_{t=1}^T (y_t - \bar{y})^4}{\left( \sum_{t=1}^T (y_t - \bar{y})^2 \right)^2} \quad (6)$$

The JB normality test checks the following pair of hypothesis:

$H_0: \hat{\mu}_3 = 0$  and  $\hat{\mu}_4 = 0$  (i.e.  $y_t$  is from a normal distribution) against the alternative

$H_1: \hat{\mu}_3 \neq 0$  and  $\hat{\mu}_4 \neq 0$  (i.e.  $y_t$  is not from a normal distribution)

The test rejects the null hypothesis of normality if the p-value of the JB test statistic is less than  $\alpha = 0.05$ , otherwise the null hypothesis is accepted.

### Autocorrelation function (ACF) and Partial autocorrelation function (PACF)

The ACF measures whether a variable is related to its own values as a function of a time difference. Mathematically, the Autocorrelation Function (ACF) of a stationary series  $\{Y_t\}$  is defined as:

$$\rho_k = \frac{Cov(Y_t, Y_{t+k})}{\sqrt{var(Y_t)}\sqrt{var(Y_{t+k})}} = \frac{\gamma_k}{\gamma_0} \quad (7)$$

where  $\gamma_k$  is the covariance between  $Y_t$  and  $Y_{t+k}$  given as

$$\gamma_k = Cov(Y_t, Y_{t+k}) = E(Y_t - \mu)(Y_{t+k} - \mu) \quad (8)$$

and  $\gamma_0 = E\{(Y_t - \mu)^2\} = var(Y_t) = \sigma^2$  for a stationary process,  $\rho_0 = 1$  and  $-1 \leq \rho_k \leq 1$  otherwise. The sample autocorrelation function can be estimated by:

$$\hat{\rho}_k = \frac{1}{T-k} \sum_{t=k+1}^T (Y_t - \bar{Y}_t)(Y_{t-k} - \bar{Y}_{t-k}) \bigg/ \frac{1}{T-k} \sum_{t=k+1}^T (Y_{t-k} - \bar{Y}_{t-k})^2 \quad (9)$$

which is the OLS estimator in  $Y_t = c + \rho_k Y_{t-k} + e_t$  and the 95% confidence bounds are given by  $\pm 1.96/\sqrt{T}$ , where T is the number of observations.

The partial autocorrelation between  $Y_t$  and  $Y_{t+k}$  is equal to the ordinary autocorrelation between  $(Y_t - \hat{Y}_t)$  and  $(Y_{t+k} - \hat{Y}_{t+k})$ . Let  $\phi_{kk}$  denotes the partial autocorrelation between  $Y_t$  and  $Y_{t+k}$ , then we have

$$\phi_{kk} = \frac{Cov[(Y_t - \hat{Y}_t)(Y_{t+k} - \hat{Y}_{t+k})]}{\sqrt{var(Y_t - \hat{Y}_t)}\sqrt{var(Y_{t+k} - \hat{Y}_{t+k})}} \quad (10)$$

A recursive procedure for computing the sample partial autocorrelation function (PACF) starting with  $\phi_{11} = \hat{\rho}_1$  for computing  $\phi_{kk}$  was given by Durbin (1960) as

$$\hat{\phi}_{k+1,k+1} = \hat{\rho}_{k+1} - \sum_{j=1}^k \hat{\phi}_{kj} \hat{\rho}_{k+1-j} / \left( 1 - \sum_{j=1}^k \hat{\phi}_{kj} \hat{\rho}_j \right) \quad (11)$$

and  $\hat{\phi}_{k+1,j} = \hat{\phi}_{kj} - \hat{\phi}_{k+1,k+1} \hat{\phi}_{k,k+1-j}$ ,  $j = 1, 2, \dots, k$ . This procedure also holds for computing the theoretical PACF  $\phi_{kk}$ .

### Ljung-Box Q-statistic test

The Ljung-Box Q-statistic test also called Portmanteau test is a test used to investigate the presence of serial correlation or autocorrelation in the residuals of a series. The test checks the following pairs of hypotheses:

$H_0: \rho_{k,1} = \rho_{k,2} = \dots = \rho_{k,T} = 0$  (all lags correlations are zero)

$H_1: \rho_{k,1} \neq \rho_{k,2} \neq \dots \neq \rho_{k,T} \neq 0$  (there is at least one lag with non-zero correlation). The test statistic is given by:

$$Q^{(LB)} = T(T+2) \sum_{k=1}^h \frac{\hat{\rho}_k^2}{T-k} \quad (12)$$

where T is the sample size, Q is the sample autocorrelation at lag k. We reject  $H_0$  if p-value is less than  $\alpha = 0.05$  level of significance (Ljung and Box, 1978).

### Augmented Dickey-Fuller (ADF) unit root test

The Augmented Dickey-Fuller (ADF) unit root test is used to check whether the given series contains a unit root or whether the given series is stationary or not (Dickey and Fuller, 1979). The Augmented Dickey-Fuller (ADF) test constructs a parametric correction for higher-order correlation by assuming that the series follows an AR(p) process and adding lagged difference terms of the dependent variable to the right-hand side of the test regression:

$$\Delta Y_t = \alpha Y_{t-1} + X_t' \delta + \beta_1 \Delta Y_{t-1} + \beta_2 \Delta Y_{t-2} + \dots + \beta_p \Delta Y_{t-p} + \varepsilon_t \quad (13)$$

where  $X_t$  are optional exogenous regressors which may consist of constant, or a constant and trend,  $\alpha$  and  $\delta$  are parameters to be estimated, and the  $\varepsilon_t$  are assumed to be white noise. The null and alternative hypotheses are written as:

$$H_0: \alpha = 0 \text{ versus } H_1: \alpha < 0$$

and evaluated using the conventional  $t$ -ratio for  $\alpha$ :

$$t_\alpha = \hat{\alpha} / \{se(\hat{\alpha})\} \quad (14)$$

where  $\hat{\alpha}$  is the estimate of  $\alpha$ , and  $se(\hat{\alpha})$  is the coefficient standard error.

### The autoregressive (AR) model

A stochastic time series process  $\{Y_t\}$  is an autoregressive process of order p, denoted AR(p) if it satisfied the difference equation

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \varepsilon_t \quad (15)$$

where  $\varepsilon_t$  is a white noise and  $\phi_1, \phi_2, \dots, \phi_p$  are constants to be determined.

## Moving average (MA) model

A time series  $\{Y_t\}$  which satisfies the difference equation

$$Y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q} \quad (16)$$

where  $\theta_1, \theta_2, \dots, \theta_q$  are fixed constants with  $\varepsilon_t$  as white noise is called a moving average process of order  $q$ , denoted MA( $q$ ).

## Autoregressive moving average (ARMA) model

A stochastic time series process  $\{Y_t\}$  which results from a linear combination of autoregressive and moving average processes is called an Autoregressive Moving Average (ARMA) process of order  $p, q$ , denoted ARMA( $p, q$ ) if it satisfies the following difference equation:

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q} \quad (17)$$

where  $\phi_1, \phi_2, \dots, \phi_p$  are fixed constants associated with the AR terms and  $\theta_1, \theta_2, \dots, \theta_q$  are fixed constants associated with the MA terms with  $\varepsilon_t$  being a white noise. The stationarity of an ARMA( $p, q$ ) process is guaranteed if the roots of the polynomial

$$1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p = 0 \text{ lie outside the unit circle.}$$

An ARMA(2,2) model is specified as:

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} \quad (18)$$

## Autoregressive integrated moving average (ARIMA) model

Autoregressive (AR), Moving Average (MA) or Autoregressive Moving Average (ARMA) model in which differences have been taken are collectively called Autoregressive Integrated Moving Average or ARIMA models. A time series  $\{Y_t\}$  is said to follow an integrated autoregressive moving average model if the  $d^{\text{th}}$  difference  $W_t = \nabla^d Y_t$  is a stationary ARMA process. If  $\{W_t\}$  follows an ARMA( $p, q$ ) model, we say that  $\{Y_t\}$  is an ARIMA( $p, d, q$ ) process. For practical purposes, we can usually take  $d = 1$  or at most 2.

Consider then an ARIMA( $p, 1, q$ ) process, with  $W_t = Y_t - Y_{t-1}$ , we have

$$W_t = \phi_1 W_{t-1} + \phi_2 W_{t-2} + \cdots + \phi_p W_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q} \quad (19)$$

In terms of the observed series,

$$\begin{aligned} Y_t - Y_{t-1} &= \phi_1 (Y_{t-1} - Y_{t-2}) + \phi_2 (Y_{t-2} - Y_{t-3}) \cdots + \phi_p (Y_{t-p} - Y_{t-p-1}) \\ &+ \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q} \end{aligned} \quad (20)$$

## Model Order Selection

We use the following information criteria for model order selection in conjunction with log likelihood function: Akaike information criterion (AIC) due to Akaike (1978), Schwarz information Criterion (SIC) due to (Schwarz, 1978) and Hannan-Quinn information Criterion (HQC) due to (Hannan, 1980). The formula for the information criteria are:

$$AIC(P) = -2 \ln(L) + 2P \quad (21)$$

$$SIC(P) = -2 \ln(L) + P \ln(T) \quad (22)$$



$$HQC(P) = 2 \ln[\ln T] P - 2 \ln L \quad (23)$$

where  $P$  is the number of free parameters to be estimated in the model,  $T$  is the number of observations and  $L$  is the likelihood function defined as:

$$L = \prod_{i=0}^n \left( \frac{1}{2\pi\sigma_i^2} \right)^{1/2} \exp \left[ - \sum_{i=1}^n \frac{(y_i - \mu)^2}{2\sigma_i^2} \right] \quad (24)$$

$$\ln(L) = \ln \left[ \prod_{i=1}^n \left( \frac{1}{2\pi\sigma_i^2} \right)^{1/2} \right] - \frac{1}{2} \sum_{i=1}^n \frac{(y_i - \mu)^2}{\sigma_i^2} \quad (25)$$

Thus given a set of estimated ARIMA models for a given set of data, the preferred model is the one with the minimum information criteria and maximum log likelihood.

### Model Forecast Evaluation

We employed Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) accuracy measures to select an optimal model mode that is both parsimonious and accurately forecast the data based on minimum values of the accuracy measures.

#### Root Mean Square Error (RMSE)

The Root Mean Square Error is a statistical tool for measuring the accuracy of a forecast method. It is computed as:

$$RMSE = \frac{1}{n} \sum_{t=1}^n (\hat{Y}_t - Y_t)^2 \quad (26)$$

Where  $\hat{Y}_t$  is the forecast value of the series and  $Y_t$  is the actual series and  $n$  is the number of forecast observations.

#### Mean Absolute Error (MAE)

The mean absolute error (MAE) is a statistical tool for measuring the average size of the errors in a collection of predictions, without taking their directions into account. It is measured as the average absolute difference between the predicted values and the actual values and is used to assess the effectiveness of a model. It is given as:

$$MAE = \frac{1}{n} \left( \sum_{t=1}^n |\hat{Y}_t - Y_t| \right) \quad (27)$$

where  $Y_t$  is the actual value of the series at time  $t$ ,  $\hat{Y}_t$  is the forecasted value of the series and  $n$  is the number of observations. The lower the value of RMSE and MAE, the better the model is able to forecast future values.

### Paired Sample t-Test

To examine whether the actual cases of pneumocystis pneumonia are significantly different from the forecast values, a paired sample t-test is employed. A paired sample t-test is a statistical test used to determine whether there is a significant difference between the means of two related groups (actual and forecast). It is designed specifically for situations where each observation in one sample can be paired or matched with a specific observation in the other sample.

Let  $d_i = x_i - y_i$ ,  $i = 1, 2, 3, \dots, n$  denote the difference in the observations for the  $i$ th unit. Where  $x_i$  is the set of actual series and  $y_i$  is the set of forecast series. Under the following pair of hypothesis:

$$H_0: \mu_x = \mu_y \text{ versus } H_1: \mu_x \neq \mu_y \quad (28)$$

The test statistic for paired sample t-test is given by:

$$t = \frac{\bar{d}}{s/\sqrt{n}} = \frac{\bar{d}}{\sqrt{s^2/n}} \sim t_{n-1} \quad (29)$$

where

$$d = x - y; \bar{d} = \frac{1}{n} \sum d_i \text{ and } s^2 = \frac{1}{n-1} \sum (d - \bar{d})^2 = \frac{1}{n-1} \left[ \sum d^2 - \frac{(\sum d)^2}{n} \right]$$

The  $(1 - \alpha)100\%$  confidence interval for  $\mu_x = \mu_y$  is given by:

$$\left[ (\bar{x} - \bar{y}) \pm t_{n_1+n_2-2}(\alpha/2) \cdot \sqrt{s^2 \left( \frac{1}{n_1} + \frac{1}{n_2} \right)} \right] \quad (30)$$

where  $t_{n_1+n_2-2}(\alpha/2)$  is the two-tailed critical value of t for  $(n_1 + n_2 - 2)$  degree of freedom and at  $\alpha$  level of significance.

## RESULTS AND DISCUSSION

### Summary Statistics and Normality Measures

To better understand the summary statistics and distributional characteristics of the series under investigation, we compute the descriptive statistics such as monthly mean, maximum and minimum, standard deviation as well as normality measures such as skewness, kurtosis and Jarque-Bera statistic of Pneumocystis Pneumonia Infection data in Benue State. The results are presented in Table 1.

The summary statistics results reported in Table 1 show that the monthly mean of pneumocystis pneumonia infection in Benue State is approximately 613 infections with an approximate standard deviation of 175 infections which indicate a high level of dispersion from the average monthly infection for the period under review. The wide gap between the maximum and minimum infection gives supportive evidence for the high level of variability of Pneumonia Infection in the study area over the period under investigation.

Table 1: Summary Statistics of Pneumocystis Pneumonia Infection in Benue State

Variable	Value
Mean	613.0774
Maximum	1131.000
Minimum	370.0000
Standard Deviation	175.0535
Skewness	0.858833
Kurtosis	3.584584



Jarque-Bera statistic	23.04481
p-value	0.000010
Number of Observations	168

The skewness coefficient of the series which is greater than zero indicates that the distribution of pneumocystis pneumonia infection in the study area is substantially positively skewed, the kurtosis coefficient, which is a measure of the thickness of the tail of the distribution of the series exhibit a kurtosis which is greater than 3. The skewness and kurtosis coefficients of the series show that the pattern of pneumonia infection in the study area during the study period does not follow a normal distribution. The null hypothesis of normality for Jarque-Bera test at 5% level of significance is rejected for this series since the p-value of the Jarque-Bera test statistic is 0.000010 which is less than  $\alpha = 0.05$ . In conclusion, the distribution of pneumocystis pneumonia infection among farmers in Benue State does not follow normal distribution.

### Graphical Properties of the Series

The first step in analyzing time series data is to plot the original series in level against time and observe its graphical properties. This help in understanding the trend as well as pattern of movement of the original series. Here we plot the original series (monthly infection cases of pneumocystis pneumonia) as a function of time. The time plot is presented in Figure 1. The time plot of pneumocystis pneumonia infection cases reported in Figure 1 represents the raw series in level with high mean and variance. To reduce this high mean and variance in the series, we transform the series to natural log which stabilizes the mean and variance of the series. The time plot of the natural log transform of the series is reported in Figure 2. We observe from the natural log transform series in Figure 2 that the trending in the series is not smooth which indicates that the series do not have a constant mean and variance (i.e., the series is not mean reverting). The variability in the series appears not to be uniform which raises the possibility that the variance is changing with time (heteroskedastic). These observations suggest that the series is non-stationary and contains a unit root. The series is thus subjected to differencing and the result is presented in Figure 3. From the result of the first difference series reported in Figure 3, we observe that the trending in the series is smoother which indicates that the series have a constant mean (i.e., the series is mean reverting). The variability in the series appears to be uniform which raises the possibility that the variance is not changing with time (homoskedastic). These observations suggest that the series is weakly or covariance stationary. The series also exhibits some gradual rise and fall, which indicates the presence of some degree of autocorrelation.

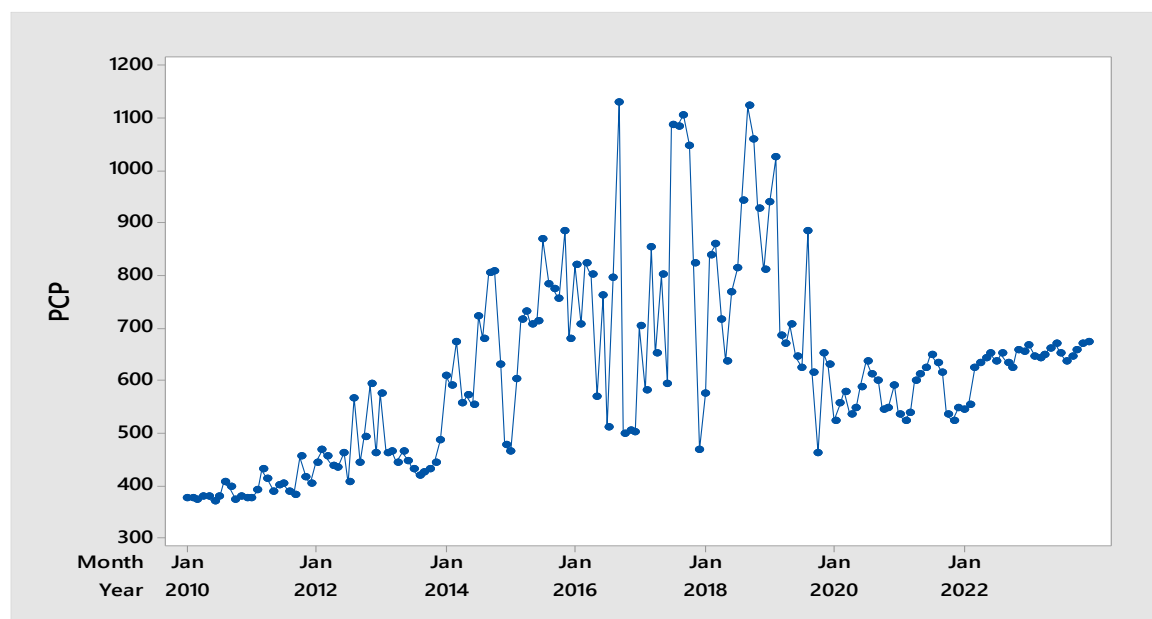


Figure 1: Time Plot of Monthly Pneumocystis Pneumonia Infection Cases in Benue State (Level Series)

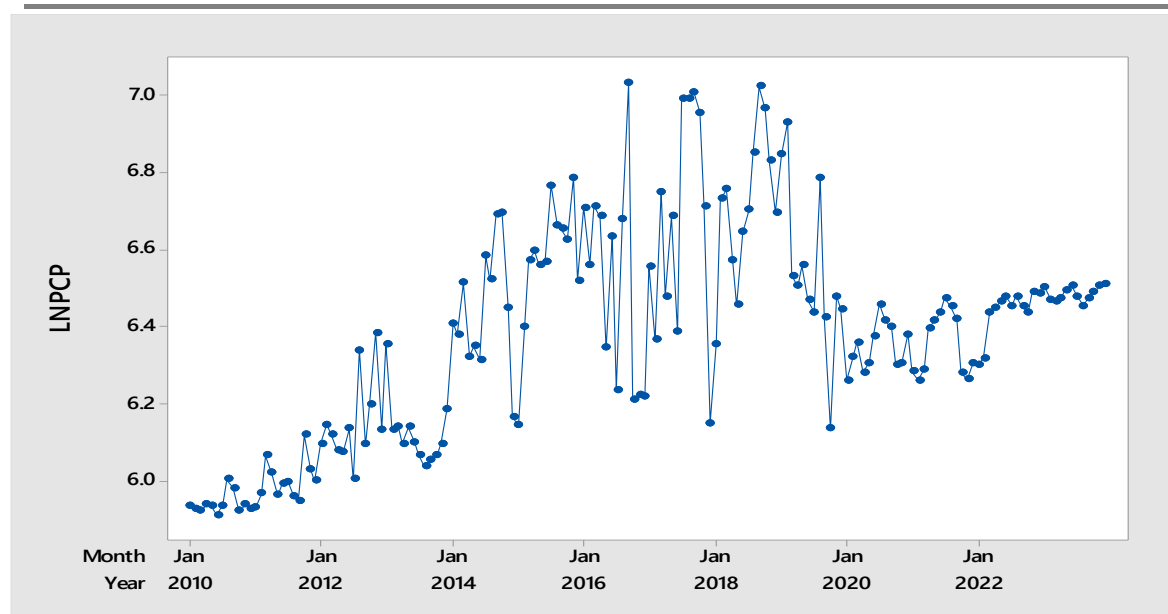


Figure 2:Time Plot of Monthly Pneumocystis Pneumonia Infection Cases in Benue State (Log Transformed Series)

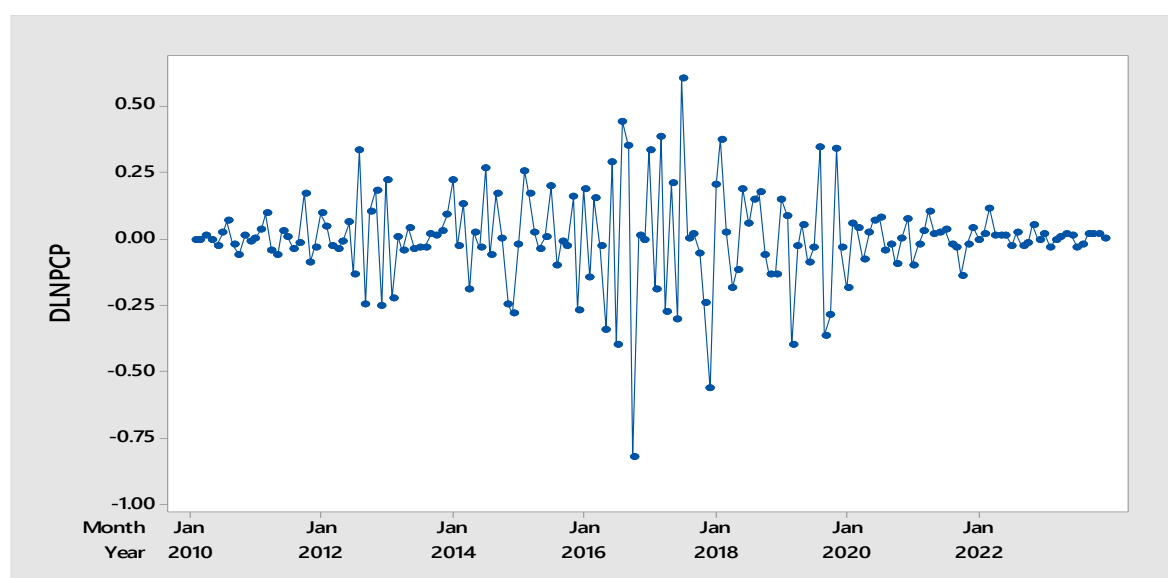


Figure 3:Time Plot of Monthly Pneumocystis Pneumonia Infection Cases in Benue State (First Differenced Series)

### Augmented Dickey-Fuller (ADF) Unit Root Test Result

This study employs Augmented Dickey-Fuller (ADF) unit root test to determine the order of integration and stationarity characteristics of the series. The result of the ADF test is reported in Table 2.

The ADF unit root test result which is conducted with intercept only and with intercept and linear trend reported in Table 2 fails to reject the null hypothesis of unit root in the level of the series. This means that the series is non-stationary and contains a unit root in level since the ADF test statistics are greater than the critical values of the ADF test at the 5% levels of significance with statistically non-significant p-values. However, the ADF unit root test result of the first difference of the series which is also conducted with intercept only and with intercept and linear trend rejects the null hypothesis of unit root in the series. This means that the first difference of series is stationary and do not contains a unit root since the ADF test statistics are smaller than the critical values of the ADF test at the 5% levels of significance with statistically significant p-values.

Table 2: ADF Unit Root Test Result

Variable	Option	ADF test stat.	p-value	5% critical value	Remark
<i>lnpcp</i>	Intercept only	-2.3675	0.1526	-2.4707	Non-stationary
	Intercept & trend	-2.4555	0.3499	-3.4376	Non- stationary
$\nabla \lnpcp$	Intercept only	-9.1211	0.0000	-2.8793	Stationary
	Intercept & trend	-9.1439	0.0000	-3.4378	Stationary

### Ljung-Box Q-statistic Test for Serial Correlation

To investigate the presence of autocorrelation in the series, we employ Ljung-Box Q-statistic serial correlation test for both the natural log level series and the first difference of the natural log transformed series and the results are presented in Table 3.

The results of Ljung-Box Q-statistic reported in Table 3 show the presence of autocorrelations in both the level series and the log transformed series as the p-values in both series are highly statistically significant. This indicates that the series are serially correlated. When a series exhibits serial correlation, it implies that the values of the series at different time points are not independent of each other, but rather, they depend on the preceding values. The implication of serial correlation in a series underscores the importance of appropriately modeling and analyzing time series data to account for the dependence structure between observations and obtain reliable results.

Table 3: Autocorrelation functions and Ljung-Box Q-statistics for PCP

Lag	ACF	PACF	Q-statistic	p-value	ACF	PACF	Q-statistic	p-value
Natural Log of PCP					First Difference of Natural Log of PCP			
1	0.7974	0.7974	108.7448	0.0000	-0.2994	-0.2994	15.2392	0.0000
2	0.7107	0.2055	195.6465	0.0000	-0.0822	-0.1888	16.3960	0.0000
3	0.6557	0.1193	270.0733	0.0000	-0.1035	-0.2156	18.2397	0.0000
4	0.6415	0.1591	341.7415	0.0000	0.0152	-0.1341	18.2796	0.0001
5	0.6213	0.0848	409.3871	0.0000	-0.0717	-0.1930	19.1752	0.0002
6	0.6278	0.1410	478.8629	0.0000	0.1100	-0.0334	21.2968	0.0002
7	0.5919	-0.0107	541.0028	0.0000	0.1126	0.1158	23.5342	0.0001
8	0.5147	-0.1381	588.2851	0.0000	-0.0838	0.0050	24.7814	0.0002
9	0.4693	-0.0279	627.8458	0.0000	-0.2000	-0.2098	31.9274	0.0000
10	0.4994	0.1700	672.9193	0.0000	0.0225	-0.1537	32.0187	0.0000
11	0.5210	0.1090	722.3047	0.0000	0.0840	-0.0444	33.2957	0.0000

12	0.5092	0.0010	769.7807	0.0000	-0.0132	-0.0884	33.3272	0.0001
13	0.5027	0.0461	816.3393	0.0000	0.0623	-0.0332	34.0379	0.0001
14	0.4730	-0.0072	857.8326	0.0000	0.1055	0.1155	36.0929	0.0001
15	0.4062	-0.1297	888.6381	0.0000	-0.2180	-0.1027	44.9163	0.0000
16	0.4222	0.0841	922.1300	0.0000	-0.0053	-0.0468	44.9214	0.0000
17	0.4378	0.0197	958.3744	0.0000	0.0493	-0.0328	45.3781	0.0000
18	0.4354	0.0078	994.4625	0.0000	0.2163	0.1785	54.2381	0.0000
19	0.3488	-0.1915	1017.7842	0.0000	-0.0842	0.0894	55.5911	0.0000
20	0.2936	-0.0980	1034.4216	0.0000	-0.1508	-0.1603	59.9545	0.0000
21	0.2965	0.1259	1051.4984	0.0000	-0.0244	-0.1392	60.0698	0.0000
22	0.3149	0.1120	1070.8943	0.0000	0.0932	0.1080	61.7594	0.0000
23	0.2952	-0.1115	1088.0570	0.0000	0.0140	0.1423	61.7980	0.0000
24	0.2690	-0.1292	1102.4079	0.0000	0.0335	-0.0158	62.0194	0.0000

### Autocorrelation and Partial Autocorrelation Functions of the Series

We also examine the autocorrelation function (ACF) and partial autocorrelation function (PACF) of the first differenced natural log transformed series to see the degree of correlation in the data points of the series. The ACF and PACF are the approximate two standard error bounds (the upper confidence bound and the lower confidence bound) computed as  $\pm 1.96\sqrt{T}$ , where T is the number of observations. If the sample autocorrelation and partial autocorrelation are within these bounds, it is not significantly different from zero at (approximately) 5% significance level. That is, if all lag values of the data or most of the lag values fall within these confidence bounds, then, the series is stationary and independent of time but non-stationary and time dependent if otherwise. The ACF and PACF plots are reported in Figure 4.

From the ACF and PACF plots of the first differenced log transformed series reported in Figure 4, it is observed that most of the lag values are inside the confidence bounds and we conclude that Pneumocystis Pneumonia infection series in Benue state is stationary in the first difference and time independent (i.e., the series do not contain a unit root in the first difference and the infection in the present month does not depend on the infection of the previous month and vice versa).

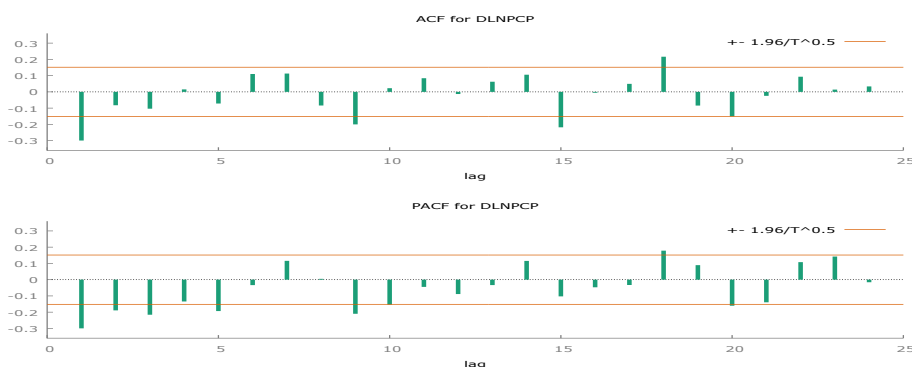


Figure 4: Autocorrelation Function Plots of the First Differenced Log Transformed Series

## Model Order Selection

To search for an optimal time series model that will best model and forecasts Pneumocystis pneumonia infection cases in Benue State, Schwarz information criterion (SIC) and Hannan Quinn criterion (HQC) in conjunction with the log likelihood (LogL) have been employed to select the optimal model. The best fitting model is the one with the least information criteria and highest log likelihood value. This model will produce the best fit as well as the forecast. The model search result is presented in Table 4.

Table 4: Model Order Selection

S/n	Model	LogL	AIC	SIC	HQC
1.	ARIMA(0,1,1)	64.2335	-0.7618	-0.7431	-0.7542
2.	ARIMA(1,1,0)	71.3238	-0.8422	-0.8235	-0.8346
3.	ARIMA(1,1,1)	75.3786	-0.8720	-0.8158	-0.8492
4.	ARIMA(0,1,2)	57.8475	-0.6808	-0.6621	-0.6732
5.	ARIMA(2,1,0)	56.1855	-0.6689	-0.6501	-0.6732
6.	ARIMA(1,1,2)	71.9178	-0.8303	-0.7744	-0.8075
7.	ARIMA(2,1,1)	70.6238	-0.8197	-0.7632	-0.7968
8.	ARIMA(2,1,2)	56.2259	-0.6452	-0.5887	-0.6224
9.	ARIMA(1,1,3)	65.7082	-0.7676	-0.7301	-0.7524
10.	ARIMA(3,1,1)	70.9362	-0.6407	-0.8029	-0.8253
11.	ARIMA(3,1,2)	75.1936	-0.8731	-0.7353	-0.7577
12.	ARIMA(2,1,3)	75.9485	-0.8478	-0.7349	-0.8020
13.	ARIMA(3,1,3)	75.0993	-0.8305	-0.6982	-0.7767
14.	ARIMA(1,1,4)	76.9713	-0.8551	-0.7426	-0.8094
15.	ARIMA(4,1,1)	72.8071	-0.8197	-0.7058	-0.7734
16.	ARIMA(2,1,4)	74.8369	-0.8223	-0.6905	-0.7688
17.	ARIMA(4,1,2)	74.2619	-0.8253	-0.6924	-0.7713
18.	ARIMA(3,1,4)	76.3251	-0.8332	-0.6820	-0.7718
19.	ARIMA(4,1,3)	74.8969	-0.8208	-0.6689	-0.7592
20.	ARIMA(4,1,4)	80.3973	-0.8760	-0.7052	-0.8067
21.	ARIMA(1,1,5)	75.8244	-0.8292	-0.6980	-0.7759
22.	ARIMA(5,1,1)	72.1822	-0.8047	-0.6713	-0.7506

23.	ARIMA(2,1,5)	82.6483	-0.9048	-0.7542	-0.8437
24.	ARIMA(5,1,2)**	97.9527	-0.9776	-0.7911	-0.9355
25.	ARIMA(3,1,5)	81.8532	-0.8885	-0.7183	-0.8194
26.	ARIMA(5,1,3)	79.7985	-0.8741	-0.7025	-0.8044
27.	ARIMA(4,1,5)	78.4717	-0.8401	-0.7631	-0.7631
28.	ARIMA(5,1,4)	79.9707	-0.8638	-0.6732	-0.7865
29.	ARIMA(5,1,5)	91.7794	-0.9073	-0.7676	-0.9122

Following the result of Table 4 on model order selection, ARIMA (5,1,2) model seems to provide statistically adequate representation of the given data since it has the highest log likelihood value as well as the smallest AIC, SIC and HQC values. Hence ARIMA (5,1,2) model has been retained as the optimal and best candidate to model and forecast Pneumocystis Pneumonia infection cases in Benue State.

### Model Estimation Result

After the best model has been chosen, the next thing to do is to estimate the parameters of the model. The result of the parameter estimates of the optimal ARIMA (5,1,2) model is presented in Table 5.

Table 5: Parameter Estimates of ARIMA (5,1,2) Model

Variable	Coefficient	Std. Error	t-Statistic	P-value
AR(1)	0.232551	0.098942	2.350372	0.0200
AR(2)	0.843050	0.087982	9.582104	0.0000
AR(3)	-0.424496	0.098911	-4.291705	0.0000
AR(4)	0.231631	0.080086	2.892272	0.0044
AR(5)	-0.234496	0.086340	-2.715967	0.0074
MA(1)	-0.710015	0.063880	-11.11474	0.0000
MA(2)	0.895595	0.058281	15.36684	0.0000
R-squared	0.761643	AIC	-0.973573	
Adjusted R <sup>2</sup>	0.633061	SIC	-0.791106	
Log likelihood	97.95273	HQC	-0.935502	
Durbin-Watson stat.	1.996233			

From the result of the parameter estimates of Table 5, the data fits an ARIMA (5,1,2) model which is presented below:

$$(1 - L)Y_t = 0.2326Y_{t-1} + 0.8431Y_{t-2} - 0.4245Y_{t-3} + 0.2316Y_{t-4} - 0.2345Y_{t-5} + \varepsilon_t - 0.7100\varepsilon_{t-1} + 0.8956\varepsilon_{t-2} \quad (31)$$



where

$Y_t$  =Pneumocystis Pneumonia infection response (dependent) variable at time  $t$

$Y_{t-1}, \dots, Y_{t-5}$  =Pneumocystis Pneumonia infection response variables at time  $t - 1, \dots, t - 5$  respectively

$\varepsilon_t$  = Error term at time  $t$

$\varepsilon_{t-1}, \varepsilon_{t-2}$  = Error terms in the previous time periods which are incorporated in the response variable  $Y_t$ .

The result of the estimated ARIMA (5,1,2) model presented in Table 5 and Equation (31) shows that the AR and MA slope coefficients of the model are all statistically significant at 5% significance levels and satisfy the stationarity and stability constraints of the model as the sums of AR and MA terms are all less than unity (i.e.,  $\phi_i + \theta_i < 1$ ). The coefficient of determination ( $R^2$ ) of the regression model is 0.7616 indicating that about 76.16% of the total variations in Pneumocystis Pneumonia infection in Benue state has been explained by independent variables while the remaining 23.84% unexplained variations is being accounted for by the error term or by factors not included in the model. The value of Durbin Watson statistic is 1.9962 which is approximately 2 indicating that the model is not spurious and there is no positive serial correlation in the residuals of the estimated model.

### Stability and invertibility analysis of the model

An evidence to show that the estimated ARIMA model is dynamically stable is that the inverse roots of AR/MA polynomials should all lie within a unit circle. The table of AR/MA polynomial roots of estimated ARIMA (5,1,2) model is reported in Table 6.

From the results of the AR/MA polynomial roots of the estimated model reported in Table 6, it is observed that all the roots lied inside a unit circle and the model is dynamically stable and satisfied the stability and invertibility conditions. From the root of AR and MA polynomials of the fitted model presented in Table 6, sum of AR roots = 4.0605 and sum of MA roots = 1.9016 and we estimate that  $\tan \theta = y/x = 4.0605/1.9016 = 2.1353$  and  $\theta = 64.3^\circ$ . Thus, the life cycle of pneumocystis pneumonia infection among the farming population in Benue state is computed as  $360^\circ/64.3^\circ = 5.5988 \approx 6$  months and we say that pneumocystis pneumonia infection among the farming population of Benue state has a life cycle of 6 months which could be described as chronic, a disease condition in which if not properly controlled, prevented and treated will have severe complications and high risk of developing into serious infection and death.

Table 6: AR/MA Polynomial Roots of Estimated ARIMA (5,1,2) Model

	Root	Real	Imaginary	Modulus
AR	Root 1	0.44	-0.86	0.9660
	Root 2	0.44	0.86	0.9660
	Root 3	-0.04	-0.66	0.6612
	Root 4	-0.04	0.66	0.6612
	Root 5	-0.57	0.57	0.8061
MA	Root 1	0.36	0.88	0.9508
	Root 2	0.36	-0.88	0.9508

## ARIMA (5,1,2) model validation and diagnostic checks

After the model fit, the adequacy of the model is being checked by examining the goodness of fit by means of plotting the ACF and PACF of residuals of the fitted model. If most of the sample autocorrelation coefficients of the residuals are within the confidence limits  $\pm 1.96/\sqrt{T}$  where T is the number of observations upon which the model is based, then the residuals are said to be white noise indicating that the model is a good fit. The Ljung-Box Q-statistics test is also conducted on the residuals to check the presence of serial correlation (autocorrelation) in the residuals of the estimated ARIMA (5,1,2) model. The ACF and PACF plots are reported in Figure 5 while the Q-statistic test results are presented in Table 7.

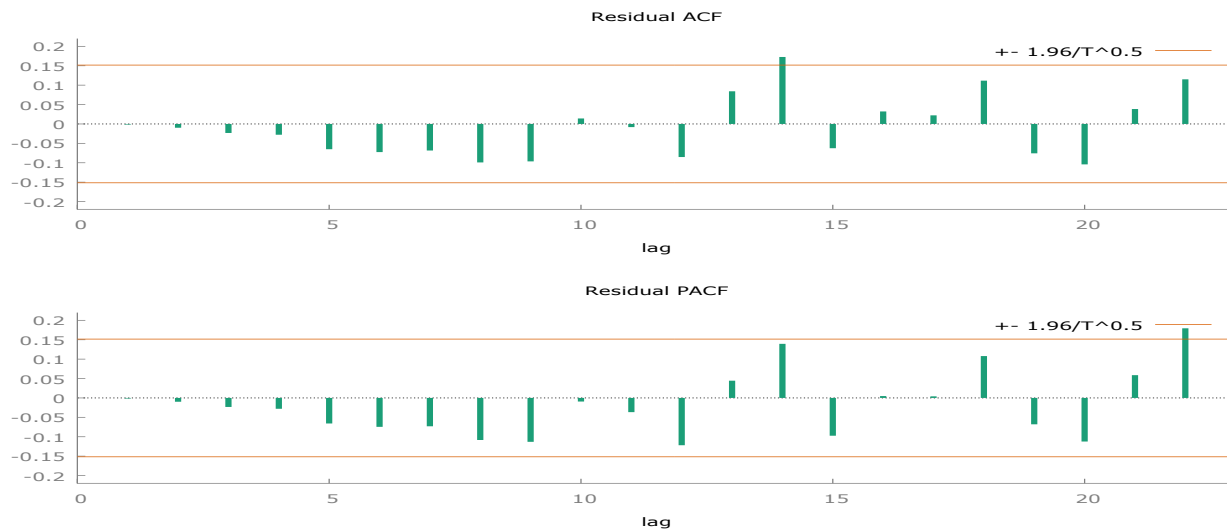


Figure 5: ACF and PACF of Residuals of the Estimated ARIMA (5,1,2) Model

Table 7: Autocorrelation functions and Ljung-Box Q-statistics Test for Residuals

Lag	ACF	PACF	Q-Statistics	p-value
1	-0.091	-0.107	5.9935	0.150
2	0.019	-0.004	6.0545	0.109
3	-0.009	-0.037	6.0701	0.194
4	-0.091	-0.127	7.5269	0.184
5	0.080	0.040	8.6746	0.193
6	0.175	0.141	14.166	0.148
7	-0.055	-0.091	14.716	0.165
8	0.036	0.009	14.950	0.192
9	0.017	-0.001	15.003	0.132
10	0.104	0.100	17.012	0.108
15	-0.006	0.013	22.691	0.122

20	-0.074	-0.122	24.869	0.253
24	0.074	-0.014	28.321	0.293

The ACF and PACF plots in Figure 5 shows that all the sample autocorrelation coefficients of the residuals are within the confidence bounds indicating that the residuals are white noise and the fitted model is stable and stationary. It is therefore concluded that the model is adequate, valid and good and should be used for forecasting. From the results of Table 7, the null Hypothesis of no serial correlation in the residuals of the fitted ARIMA (5,1,2) model at all lags is not rejected since the p-values of the Q-statistics are all greater than 0.05. This shows that the estimated model is stationary and dynamically stable.

### Forecast evaluation results

Having validated our model, we now seek an appropriate forecast mode that best forecast future relevant series. Here we consider in-sample and out-of-sample forecasts using two accuracy measures namely: Root Mean Square Error (RMSE) and Mean Absolute Error (MAE). The forecast mode with the least accuracy measures stands as the best to predict pneumocystis pneumonia infection cases Benue state of Nigeria. The result of forecast comparison is presented in Table 8.

From Table 8, we consider two benchmarks: Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) to compare the in-sample and out of sample forecasts performance of the estimated ARIMA (5,1,2) model to evaluate its forecast ability and to decide on which mode of forecast is better for the model. We observe that the RMSE and MAE of the out-of sample forecast are smaller than those of the in-sample forecast, and the decision is that the smaller the forecast errors, the better the forecasting ability of that model, according to the criterion, our model is good for future forecast.

Table 8: Forecast Comparison using Accuracy Measures

Performance Metrics	In-Sample Forecast	Out-of-Sample Forecast **
RMSE	0.174360	<b>0.149735</b>
MAE	0.112941	<b>0.104349</b>

Note: \*\* denotes forecast mode selected by accuracy measures.

### Short-Term Forecast of Pneumocystis Pneumonia Infection Cases in Benue State

Having selected the out-of-sample forecast approach for the series, we use the estimated ARIMA (5,1,2) model to forecast future values of pneumocystis pneumonia infection cases in the study area for the period of 2 years (24 months) starting from January 2024 to December 2025. The result of the forecast is presented in Table 9.

Table 9: Forecast of Pneumocystis Pneumonia Infection Cases in Benue State from January 2024-December, 2025

Year: Month	Forecast (natural log form)		Actual Forecast (No. of Infections)		
	Forecast	Std. error	LCL	Forecast	UCL
2023:12	6.51026	-	-	672	-
2024:01	6.52156	0.147412	509	<b>680</b>	907
2024:02	6.51315	0.166426	486	<b>674</b>	934

2024:03	6.50082	0.179964	468	666	947
2024:04	6.50402	0.190752	460	668	971
2024:05	6.52355	0.201601	459	681	1011
2024:06	6.53840	0.209167	459	691	1041
2024:07	6.53746	0.221301	448	691	1066
2024:08	6.52734	0.237553	429	684	1089
2024:09	6.52296	0.250826	416	681	1113
2024:10	6.53150	0.258418	414	687	1139
2024:11	6.54672	0.264286	415	697	1170
2024:12	6.55600	0.272062	413	704	1199
2025:01	6.55373	0.283060	403	702	1222
2025:02	6.54672	0.294677	391	697	1242
2025:03	6.54626	0.303677	384	697	1263
2025:04	6.55592	0.309951	383	703	1291
2025:05	6.56841	0.315667	384	712	1322
2025:06	6.57408	0.322980	380	716	1349
2025:07	6.57119	0.332267	372	714	1370
2025:08	6.56703	0.341431	364	711	1389
2025:09	6.56962	0.348567	360	713	1412
2025:10	6.57930	0.354090	360	720	1441
2025:11	6.58903	0.359669	359	727	1471
2025:12	6.59226	0.366617	356	729	1497
<b>Total</b>	<b>157.13703</b>			<b>16745</b>	
<b>Average</b>	<b>6.5473625</b>			<b>697.70833</b>	

Note: For 95% confidence intervals,  $Z_{0.025} = 1.96$ . LCL and UCL denote lower and upper confidence limits respectively.

The forecast presented in Table 9 that the forecast of pneumocystis pneumonia infection value for the month of January, 2024 was 680 cases with a 95% confidence interval of [509, 907] cases. By this we are 95% confident that the outcome of pneumocystis pneumonia infection in the study area for the next period will fall within this interval. Comparing with the monthly infection in the month of December 2023 which was (672 cases), it is predicted that in January 2025 the pneumocystis pneumonia infection will slightly increase from the current

month by 8 cases. The interval [509, 907] cases imply that the increase in the number of pneumocystis pneumonia infection cases in the month of January, 2024 may lie between 509 and 907 persons (i.e., it may reduce by at least 163 cases or increase by at most 235 cases) in the month of January, 2024. The forecast shows that at least 666, 691, 681 and 704 persons will be infected with pneumocystis pneumonia in Benue state in the months of March, 2024, June, 2024, September, 2024 and December, 2024 respectively. Also, about 697, 716, 713 and 729 persons are predicted to be infected with the disease in the months of March, 2025; June, 2025; September, 2025 and December, 2025 respectively. The forecast also shows that at least a total of 16,745 persons will be infected with pneumocystis pneumonia in the study area within the period of two years from January, 2024 to December, 2025 with an average of 698 infections cases per month. The forecast shows an increasing and fluctuating trend in the level of pneumocystis pneumonia infection in Benue state of Nigeria over the forecasted period typical of the trend found in the original series. The confidence intervals of the forecast also follow this increase in the trend level of infection during the forecasted period from January 2024 to December 2025.

### Forecast Reliability Test using Paired Sample t-test

We test the reliability and accuracy of the pneumocystis pneumonia infection forecast values to ascertain whether a significant difference exist between the observed actual infection cases and the forecast values by a mere comparison of the actual infection cases and the point forecast values through a paired sample t-test as presented in Table 10. We have restricted the test to a few forecast values (January 2018 to December 2019) because long term forecast is not advisable for obvious reasons.

Let  $\mu_0$  represents the observed mean of pneumocystis pneumonia infection cases and let  $\mu_F$  represents the forecast mean. Then we state the following hypothesis:

$H_0: \mu_0 - \mu_F = 0$ : There is no significant difference between the observed mean and forecast mean.

$H_1: \mu_0 - \mu_F \neq 0$ : There is a significant difference between the observed mean and forecast mean.

Failure to reject the null hypothesis implies that our forecast is reliable, accurate and valid. The result is presented in Table 10.

Table 10: Paired-Sample t-Test between Actual and Forecast Values

	Paired Differences					t-stat.	Df	p-value
Actual-Forecast	Mean	SD	SE	95% of the Diffs.				
	-0.01875	0.16674	0.03404	-0.089	0.05166	-0.551	23	0.587

From the result of the paired-sample t-test presented in Table 10, we fail to reject the null hypothesis at 5% level of significance and reasonably conclude with 95% confident that the difference between the observed pneumocystis pneumonia infection values and the forecast values are equal to zero or more technically not statistically significantly different from zero. This conclusion is made on the basis of the large p-value of 0.587 which is greater than 0.05. Alternatively, we can equally infer from the 95% confidence interval (-0.089, 0.05166) that the interval contains zero. Since we failed to reject the null hypothesis, we therefore conclude that our forecast values of pneumocystis pneumonia infection cases in Benue state are reliable, valid and accurate and can be relied upon for policy implementation.

## CONCLUSION AND RECOMMENDATIONS

This study successfully applied time series analysis using the ARIMA (5,1,2) model to forecast short-term trends of Pneumocystis pneumonia (PCP) infection cases in Benue State, Nigeria. The findings reveal that PCP infections follow a cyclical pattern with a six-month recurrence period, highlighting the chronic and potentially life-threatening nature of the disease if not effectively managed. The ARIMA model demonstrated strong

predictive power, explaining 76.16% of the variance in the data, and was used to project future infection trends for the period of January 2024 to December 2025. The forecast indicates a fluctuating yet increasing trend in PCP cases, with an average monthly infection rate of 698 cases. The study underscores the importance of data-driven public health planning, demonstrating that time series forecasting can provide reliable insights for disease surveillance, healthcare resource allocation, and policy development. The validated forecast model offers a valuable tool for health authorities to anticipate and mitigate potential surges in PCP infections through targeted interventions. This study contributes to the field of epidemiological modeling by providing a robust framework for forecasting infectious diseases, demonstrating that time series analysis can play a critical role in enhancing public health preparedness and response strategies.

The following recommendations are presented based on the study's findings:

1. Health authorities in Benue State should enhance data collection and surveillance mechanisms for PCP infections to ensure timely detection of trends and outbreaks. Regular updates to the time series data will improve the accuracy of future forecasts and inform better policy decisions.
2. Given the chronic and life-threatening nature of PCP infections, healthcare facilities should be equipped with advanced diagnostic tools and trained personnel to facilitate early detection and prompt treatment, reducing disease progression and mortality rates.
3. The study's findings on the six-month PCP infection cycle highlight the need for periodic public health interventions, such as awareness campaigns, preventive treatment strategies, and community outreach programs, to mitigate transmission risks and reduce infection rates.
4. The validated ARIMA model should be integrated into public health planning to anticipate healthcare demands. This will help allocate resources effectively, ensuring sufficient medical supplies, hospital capacity, and healthcare personnel in anticipation of projected PCP infection trends.

## REFERENCES

1. Akaike, H. (1974). A new look at statistical model identification. Institute of Electrical and Electronics Engineers Transmission on Automatic Control, AC-19, 716-723.
2. Boubaker, S., Trabelsi, S., de Lavallade, H., & Zahar, J. R. (2019). The changing face of *Pneumocystis jirovecii* pneumonia: Shifting epidemiology and challenges in diagnosis. *Current Opinion in Pulmonary Medicine*, 25(2), 142-148.
3. Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time series analysis: forecasting and control* (5th ed.). Wiley.
4. Broadhurst, A. (2021). The diagnostic challenge of pneumocystis pneumonia and COVID-19 co-infection in HIV. *International Journal of Medical Biology*, 13(4), 38-51.
5. Bruns, S., Suppan, R., Reichelt, M., Koczulla, A. R., Krüger, C., Kern, W. V., ... & Vehreschild, M. J. (2022). Evolving epidemiology of *Pneumocystis pneumonia*: Findings from a longitudinal population-based study and a retrospective multi-center study in Germany. *BMC Infectious Diseases*, 20(1), 882-895.
6. Chen, S., Tseng, Y., & Li, W. (2021). *Pneumocystis jirovecii* Pneumonia in Solid Organ Transplant Recipients in Australia and New Zealand (2000-2019): A Retrospective Multicentre Analysis. *Clinical Infectious Diseases*, 72(9), 1553-1562.
7. Chen, D. Y., Huang, Y. C., Liu, C. Y., & Lin, W. C. (2019). *Pneumocystis jirovecii* pneumonia: A review of current diagnostic and therapeutic approaches. *Journal of Microbiology, Immunology, and Infection*, 52(2), 1-8.
8. Chiliza, M., Mathebula, M., Moodley, S., & Moloi, P. L. (2020). Outcomes of HIV-associated pneumocystis pneumonia at a South African referral hospital. *PLoS One*, 15(2), e0230222.
9. Dickey, D. A., & Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, 74(366), 427-431.
10. Durbin, J. (1960). The fitting of time series models. *Review of the Institute of International Statistics*, 28, 233-244.



11. Hannan, E. (1980). The estimation of the order of ARMA process. *Annals of Statistics*, 8, 1071-1081.
12. Iriart, X. (2015). Pneumocystis Pneumonia in Solid-Organ Transplant Recipients. [10.3390/jof1030293](https://doi.org/10.3390/jof1030293).
13. Jarque, C. M. & Bera, A. K. (1980). Efficient test for normality, heteroskedasticity and serial independence of regression residuals. *Econometric Letters*, 6, 255-259.
14. Jarque, C. M. & Bera, A. K. (1987). A test for normality of observations and regression residuals. *International Statistical Review*, 55(2), 163-172.
15. Ji, W., Liu, R., Yong, F., & Guo, L. (2016). Respiratory health status and its influencing factors of farmers in the south of China. *International Journal of Environmental Research and Public Health*, 13(1), 103-111.
16. Koo, B. S., Hong, S., & Kim, Y. J. (2021). Epidemiology and clinical features of *Pneumocystis jirovecii* pneumonia in Korean patients with rheumatic diseases. *Biomedical Journal*, 26(4), 127-139.
17. Ljung, G. M. & Box, G. E. P. (1978). On a measure of lack of fit in time series models. *Biometrika*, 65, 297-303.
18. Muñoz, P., Giannella, M., Montesinos, P., Berenguer, J., & Bouza, E. (2020). *Pneumocystis jirovecii* pneumonia in non-HIV immunocompromised patients: A series of 21 cases and review of the literature. *Medicine (Baltimore)*, 99(32), e21443.
19. O'Donnell, A. C., Bow, E. J., Tilley, J. A., Allan, D. S., Beyene, J., Ethier, M. C., Lehrnbecher, T., & Sung, L. (2018). *Pneumocystis jirovecii* pneumonia prophylaxis for patients with hematologic malignancies and undergoing hematopoietic stem cell transplantation: a systematic review and meta-analysis of randomized controlled trials. *Leukemia & Lymphoma*, 59(5), 1133-1143.
20. Olugbenga, G. A., Ojo-Osagie, M. A., Adetiloye, V. A., Idigbe, O. O., Ogbonna, A. C., & Omitolade, A. O. (2020). Pneumonia hospitalizations and mortality in children 3 - 24-month-old in Nigeria from 2013 to 2020: Impact of pneumococcal conjugate vaccine ten valent (PHiD-CV-10). *PubMed Central*, 11(8): e02162289.
21. Polaczek, M. (2014). *Pneumocystis pneumonia* in HIV-infected patients with cytomegalovirus co-infection. Two case reports and a literature review. *Pneumonologia i Alergologia Polska*, 82(6), 517-523.
22. Roux, A., Gonzalez, F., Roux, M., Mehrad, M., Menotti, J., Zahar, J. R., Canet, E., & Azoulay, E. (2014). Update on pulmonary *Pneumocystis jirovecii* infection in non-HIV patients. *Medical & Malária Infections*, 44(5), 185-198.
23. Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6(2): 461-464.
24. Stern, A. (2014). Prophylaxis for *Pneumocystis pneumonia* (PCP) in non-HIV immune compromised patients. *Medical Journal*, 24(2), 118-129.
25. Wolf, D. G., & Pappas, P. G. (2020). *Pneumocystis jirovecii* Pneumonia. In *Infectious Diseases Society of America (Ed.), Mandell, Douglas, and Bennett's Principles and Practice of Infectious Diseases (Eighth Edition)* (pp. 2883-2896). Elsevier.
26. Yang, H. Y., Sun, T. T., & Li, Z. L. (2021). *Pneumocystis jirovecii* pneumonia in kidney transplant recipients: a multicenter case series study. *BMC Infectious Diseases*, 21(1), 53-62.
27. Zahradnik, E., Raulf, M., van Kampen, V., & Bünger, J. (2016). Respiratory allergens from furred mammals: Environmental and occupational exposure. *Veterinary Sciences*, 3(3), 17-28.
28. Zhang, M., Yin, J., Zhang, X. (2023), Factors associated with interstitial lung disease in patients with rheumatoid arthritis: A systematic review and meta-analysis. *PLoS One*, 18(6), e0286191.