Evaluating the Forecasting Performance of Time Series Approaches on Measles Data in Adamawa State

Paul Moses Medugu[1]; Yaska Mutah[2]; Nuhu Bata Malgwi[3]

[1;2;3]Department of Mathematics and Statistics, Federal Polytechnic Mubi, Adamawa State, Nigeria

Publication Date: 2025/08/26

Abstract: Accurate forecasting of measles incidence is crucial for optimizing vaccination campaigns and strengthening disease control efforts in Adamawa State, Nigeria. This study undertakes a comparative evaluation of multiple time series models to determine their relative performances in predicting measles cases. Monthly measles incidence data spanning 2020 to 2024 were analyzed using Autoregressive Integrated Moving Average (ARIMA), Seasonal ARIMA (SARIMA), and Holt–Winters exponential smoothing models. Parameter estimation was carried out via maximum likelihood, and model adequacy was verified through residual diagnostics and Ljung–Box tests. Comparative evaluation employed the Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and Root Mean Square Error (RMSE) to assess in-sample fit and out-of-sample forecast accuracy. The Holt–Winters model achieved superior performance, yielding the lowest RMSE, AIC, and BIC values, followed by SARIMA (2,1,1)(0,1,1)12_{12}12 and SARIMA (1,1,1)(0,1,1)12_{12}12. These results demonstrate the effectiveness of exponential smoothing in capturing both seasonal and trend components of measles dynamics in the state. The findings provide an evidence-based modeling framework to support public health decision-making, enabling more proactive epidemic preparedness and targeted intervention strategies.

Keywords: *Measles Incidence, Time Series Forecasting, Holt–Winters, SARIMA, Adamawa State.*

How to Cite: Paul Moses Medugu; Yaska Mutah; Nuhu Bata Malgwi (2025) Evaluating the Forecasting Performance of Time Series Approaches on Measles Data in Adamawa State. *International Journal of Innovative Science and Research Technology,* 10(8), 1276-1280. https://doi.org/10.38124/ijisrt/25aug554

I. INTRODUCTION

Measles is a highly infectious disease that affects millions of people worldwide, particularly in developing countries (World Health Organization, 2019). Accurate forecasting of measles outbreaks is crucial for effective disease control and prevention. Time series models have been widely used to forecast measles data (Chen et al., 2017; Zhang et al., 2019). Measles is a highly contagious viral disease that poses significant public health challenges, particularly in regions with inadequate vaccination coverage. Despite global efforts to control and eliminate measles, outbreaks persist in many developing regions, including Adamawa State, Nigeria. Effective forecasting of measles incidence is essential for proactive public health interventions, resource allocation, and vaccination campaigns (World Health Organization [WHO], 2021). Time series models provide a quantitative approach to analyzing disease trends, enabling health authorities to predict future outbreaks and implement timely control measures.

Several statistical and machine learning models have been used to forecast infectious disease trends. Traditional models such as the Autoregressive Integrated Moving Average (ARIMA) and its seasonal variant, Seasonal ARIMA (SARIMA), have been widely applied due to their ability to capture linear trends and seasonal patterns in epidemiological data (Box & Jenkins, 1976; Brockwell & Davis, 2016). Additionally, Exponential Smoothing State Space Models (ETS) offer alternative approaches that account for trends and seasonality with weighted past observations (Hyndman & Athanasopoulos, 2018). Recent advances in machine learning techniques, such as Long Short-Term Memory (LSTM) networks, have demonstrated promising results in disease prediction but require extensive datasets and computational resources (Zhou et al., 2020).

Time series models, including ARIMA, SARIMA, ES, and ANN models, have been widely used to forecast measles data. Each model has its strengths and weaknesses, and the choice of model depends on the specific characteristics of the data. By selecting the most appropriate model, public health officials can accurately predict measles outbreaks and take effective control measures. The ARIMA

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25aug554

model is a popular time series model used to forecast measles data (Box et al., 2015). Studies have shown that ARIMA models can accurately predict measles outbreaks (Chen et al., 2017; Zhang et al., 2019). For instance, Chen et al. (2017) used an ARIMA model to forecast measles cases in China and found that the model performed well in predicting the number of cases.

The SARIMA model is an extension of the ARIMA model that accounts for seasonal patterns in the data (Box et al., 2015). Studies have shown that SARIMA models can accurately predict measles outbreaks, particularly in areas with strong seasonal patterns (Zhang et al., 2019; Li et al., 2020). The ES model is a simple time series model that can be used to forecast measles data (Hyndman et al., 2008). Studies have shown that ES models can perform well in predicting measles outbreaks, particularly in areas with stable trends (Li et al., 2020). The ANN model is a machine learning model that can be used to forecast measles data (Zhang et al., 2003). Studies have shown that ANN models can accurately predict measles outbreaks, particularly in areas with complex patterns (Chen et al., 2017).

In Nigeria, measles outbreaks remain a public health concern due to periodic vaccination gaps and inconsistent surveillance systems. Studies have explored the application of time series models to forecast infectious diseases, but limited research has focused on their comparative performance in predicting measles incidence in Adamawa State (Adegboye et al., 2017).

This study aims to evaluate the effectiveness of various time series models—ARIMA, SARIMA, and ETS—in approximately predicting measles cases using historical data from Adamawa State. By assessing the models' forecasting accuracy based on statistical criteria such as the Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), Mean Absolute Error (MAE), and Root Mean Square Error (RMSE), this research seeks to identify the most suitable model for measles prediction. The findings of this study will contribute to improved disease surveillance and epidemic preparedness in Adamawa State, providing a data-driven approach for public health policymakers. The results will support better decision-making regarding immunization campaigns, outbreak response strategies, and healthcare resource allocation.

II. METHODOLOGY

This study employs a systematic approach to evaluate the forecasting performance of time series models for predicting measles incidence among under five children in Adamawa state of Nigeria. The methodology involves the following steps: This research thesis adopts a descriptive research method and using time series analysis to make adequate model selection and predictions. Box-Jenkings Seasonal Auto-Regressive Integrated Moving Average Model and the Holt-winters Exponential Smoothing model approach will be used in modelling the measles data and comparisons will be made among them. The data used in this study was obtained from Nigeria Demographic and Health Survey (NDHS). The data consists of monthly measles cases reported from 2014 to 2024.The data was cleaned and preprocessed to ensure that it was suitable for analysis. This involved checking for missing values, outliers, and inconsistencies. Any missing values were imputed using the mean of the surrounding values. Time series models were used to forecast measles cases:

- Seasonal Autoregressive Integrated Moving Average (SARIMA) Model: This model was used to capture the seasonal patterns in the data.
- Exponential Smoothing (ES) Model: This model was used to capture the exponential trends in the data.

➢ *Population and Sampling Procedure*

The population for this study consists of all reported measles cases in Adamawa State, Nigeria from 2014 to 2024. The population is estimated to be around 10,000 cases. A sample of 5 years (2020-2024) of monthly measles cases was selected from the population. This sample consists of 60 data points (12 months x 5 years). The sample was chosen to represent the most recent and reliable data available. A random sampling technique was used to select the sample. This involves selecting a subset of data points from the population at random. The sample size was determined using the following formula:

$n = (Z^2 \times \sigma^2) / E^2$

Where:

n = sample size

Z = Z-score (1.96 for 95% confidence level)

σ = standard deviation of the population

E = margin of error (0.05)

Using this formula, a sample size of 60 was determined to be sufficient for this study. A sample size of 60 was chosen because it provides a good balance between precision and feasibility. A larger sample size would provide more precise estimates, but it would also increase the complexity and cost of the study. A smaller sample size would be less precise, but it would also be more feasible and less expensive.

➢ *Model Selection*

- *Autoregressive Process*

An autoregressive process may be used to forecast a time series. As mentioned earlier, a first-order autoregressive model is denoted AR(1) and is Y_t regressed on Y_{t-1}. An autoregressive model of the pth-order is denoted AR(p) and takes the form of where the constant is denoted by δ and $u_t$is white noise (Gujarati & Porter 2008).

$$Y_t = \delta_1 + \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \cdots + \alpha_p Y_{t-p} + u_t \qquad (1)$$

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25aug554

- *Moving Average Process*

In a moving average process, the dependent variable is regressed on current and lagged error terms and is therefore estimated through a constant and a moving average of the error terms. If the dependent variable is regressed on the current and one lagged error term, it follows a first-order moving average process, denoted MA(1). Moreover, a model that includes q number of error terms follows a qth-order moving average process, denoted MA(q). A MA(q) process is defined as where the error terms u are assumed to be white noise and μ is the constant (Gujarati & Porter 2008). In a MA model the error terms are usually scaled to make $\beta_{:}$ equal to one (Chatfield 2003).

$$Y_t = \theta + \alpha_1 y_{t-1} + \beta_0 u_t + \beta_1 u_{t-1} \quad (2)$$

- *Seasonal ARMA Models*

Seasonal data may be also modelled. The numbers of seasonal AR and MA terms are usually denoted by P and Q respectively. Thus, a general seasonal ARMA model may be and S is the seasonal span, hence quarterly data S = 4 and for monthly data S = 12.

Represented as;

$$\varphi(B)\phi(B)X_t = \theta(B)\Theta(B)\varepsilon_t$$

Where

$$\varphi(B) = 1 - \phi_{1s}B^{1s} - \phi_{2s}B^{2s} - \cdots - \phi_{ps}B^{ps}, \quad (3)$$

$$\Theta(B) = 1 + \Theta_{1s}B^{1s} + \Theta_{2s}B^{2s} + \cdots + \Theta_{Qs}B^{Qs}, \quad (4)$$

- *Holt-Winters Exponential Smoothing Model*

The data used in this study consist of trend and seasonal component. It is however appropriate to apply necessary smoothing technique to model the data used. Smoothing can be seen as a technique to separate the signal and the noise as much as possible and in that a smoother acts as a filter to obtain an "estimate" for the signal (Montgomery *et al.*, 2008).

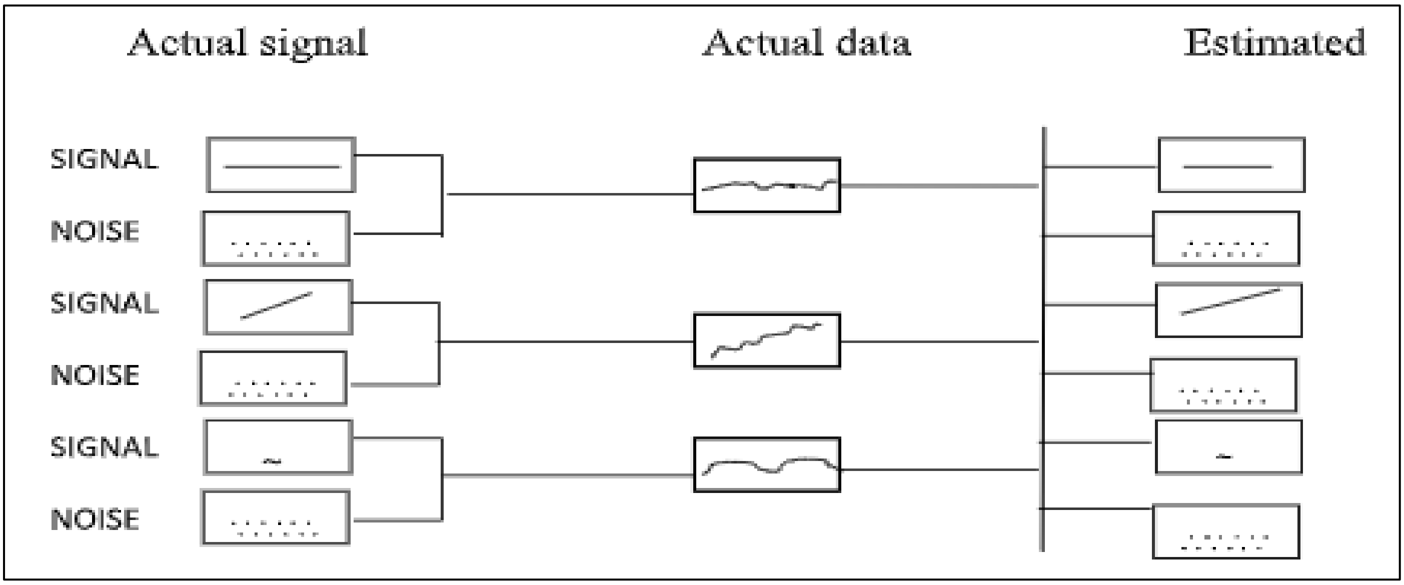

Fig 1 Presence of Signal and Noise

The fig. 1 above shows the presence of signal and noise in the actual data. The signal represents any pattern caused by the intrinsic dynamics of the process from which the data is collected and it can assume various forms. Exponential Smoothing could be Single Exponential Smoothing, Double Exponential Smoothing and Triple Exponential Smoothing.

- *Single Exponential Smoothing*

The single exponential smoothing is also referred to as **2**simple exponential smoothing. It assumes that the data fluctuates around a reasonably stable mean. The model is given below;

$$S_{t+1} = \alpha y_t + (1-\alpha)S_t \quad 0 < \alpha \le 1, t > 0 \quad (5)$$

Each successive observation in the series that the above is applied to gives each new smoothed value computed as the weighted average of the current observation and the previous smoothed observation. The weights being applied to get each smoothed value decrease exponentially depending on the value of the parameter α. New forecast is previous plus an error adjustment; this can be written as:

$$S_{t+1} = S_t + \alpha\varepsilon_t \quad (6)$$

Where ε_t is the forecast error for period *t*. However single exponential smoothing is not effective when there is a trend. The single parameter α does not accommodate this.

- *Double Exponential Smoothing*

The single exponential smoothing has only one constant, α as indicated in the equation above, which brings about the limitation in handling the presence of trend. However this situation is improved in the double

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25aug554

exponential smoothing by the introduction of another equation with additional parameter, a second constant is shown in the equations below

$$S_t = \alpha y_t + (1-\alpha)(S_{t-1} + b_{t-1})\ 0 < \alpha < 1 \quad (7)$$

$$b_t = y(S_t - b_{t-1})\ 0 < \alpha < 1 \quad (8)$$

The current value of the series is used to calculate its smoothed value replacement in double exponential smoothing. There are several methods of setting the initial values for S_t and $b_t S_1$is in general set to y_1. For b_1 the following could be adopted.

$$b_2 = y_2 - y_1 \quad (9)$$

$$b_1 = \frac{[(y_2-y_1)+(y_3-y_2)+(y_5-y_4)]}{3} \quad (10)$$

$$b_1 = \frac{y_n - y_1}{3}$$

III. MODEL EVALUATION AND COMPARISON

The performance of each model was evaluated using the following metrics: Root Mean Squared Error (RMSE): This metric measures the square root of the average squared difference between the predicted and actual values and information criteria like AIC and BIC. The performance of each model was compared using the metrics mentioned above. The model with the best performance was selected as the most suitable model for forecasting measles cases.

➢ *Data Analysis*

The analysis was performed using R software, version 4.2.3. The following packages were used using the following package: forecast, stats and t series.

➢ *Test for Stationarity (Augmented Dickey Fuller Test)*

The Augmented Dickey Fuller test, Dickey and Fuller (ADF) was employed to test for stationarity of the original time series data and the differenced time series data for the measles cases in Adamawa state.

Table 1 Unit Root Test for the Measles Cases

Null Hypothesis: Measles data has a unit root				
Exogenous: Constant				
Lag Length: 1 (Automatic - based on SIC, maxlag=13)				
			t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic			-12.10092	0.0000
Test critical values:	1% level		4.1023450	
	5% level		4.0569176	

The table 1 above shows the tests of stattionarity of the measles data and it was observed that the null hypothesis of a unit root is rejected due to the their p-values equal 0.0000 is less 5% and 1% levels of significance and therefore the data is stationary.

Table 2 Postulated SARIMA Models for Measles Data

Model	AIC	BIC	RMSE	Normality Test		Serial Correlation	
				JB Test	P-value	Q-statistic	P-value
$SARIMA(2,1,1)(0,1,1)_{12}$	-1.222	-1.104	0.129	2.269	0.322	26.027	0.762
$SARIMA(2,1,1)(1,1,1)_{12}$	1.157	1.004	0.132	1.066	0.587	23.667	0.587
$SARIMA(2,1,1)(1,1,1)_{12}$	1.157	1.004	0.132	1.066	0.587	23.667	0.587
$SARIMA(1,1,1)(1,1,1)_{12}$	1.123	0.997	0.135	1.587	0.452	34.654	0.342
$SARIMA(1,1,1)(0,1,1)_{12}$	1.203	1.109	0.13	2.102	0.35	32.94	0.47
$SARIMA(1,1,2)(1,1,1)_{12}$	1.135	0.984	0.133	1.172	0.557	29.161	0.561

From Table 2, considering the models performance in terms of AIC, BIC and RMSE. $SARIMA(2,1,1)(0,1,1)_{12}$ and $SARIMA(1,1,1)(0,1,1)_{12}$ with lower AIC and BIC values performed competitively better than the other SARIMA models. The L-jung Box portmanteau test for residual autocorrelation shows that the residuals are not serially correlated and the Jarque-Bera normality test confirms that the residuals are normal. Both models should be adequate in modelling the measles data. The $SARIMA(1,1,1)(0,1,1)_{12}$ is the selected model since it is more parsimonious.

Table 3 Parameter of Estimated SARIMA Models

$SARIMA(3,1,1)(2,1,2)_{12}$		
Variables	**Estimates**	**P-value**
$AR(1)$	-0.39	0.005
$AR(2)$	-0.324	0.009
$AR(3)$	0.176	0.042
$SAR(12)$	-0.195	0.029
$SAR(24)$	0.275	0.007

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25aug554

MA(1)	-0.459	0.04
SMA(12)	0.665	0
SMA(24)	0.872	0
JB Test	4.352	0.113
Q-Test	16.82	0.952
AIC	-1.968	
BIC	-1.721	

Table 3 presents the estimated parameters of the seasonal models for both measles cases. The diagnostics checks performed indicate that the models should be adequate in modelling the data. The performances of these models were evaluated in comparison with the Holt-winters Exponential smoothing.

➢ *Estimation of the Holt-Winter's Exponential Smoothing Model*

The time series data obviously indicate the presence of seasonality as shown previously. The multiplicative and additive Holt-winters exponential smoothing models would be examined on the variable to see which models fit best.

Table 4 Estimates of the Holt-Winters Exponential Smoothing Parameters

Model	α (Level)	β (Trend)	γ (Seasonal)	BIC	AIC
Multiplicative	0.46791	0.0000	0.21912	192.4516	146.3267
Additive	0.23097	0.0000	0.12071	166.4400	120.3152

Estimates of the parameters in the Table 4 shows that the additive Holt-winters model is best suited for modelling the measles for Adamawa state based on AIC and BIC criteria. The parameters are estimated objectively rather than subjectively, by choosing the values that best minimize the sum of squared errors. The value of the parameter β being zero in the models for both models indicates that the slope is relatively constant over time. The values of α and γ show that emphasis is only fairly placed on the recent observation. The residuals of the best of these models would be examined for presence of non-zero autocorrelations, using the Ljung-Box test and also the forecast performance characteristics.

IV. CONCLUSION

This study aimed to evaluate the forecasting performance of two time series models (SARIMA and ES) in predicting measles cases in Adamawa State, Nigeria. The results showed that Holt-winters model is best suited for modelling the measles for Adamawa state, followed by the $SARIMA(2,1,1)(0,1,1)_{12}$ and $SARIMA(1,1,1)(0,1,1)_{12}$ with lower RMSE, AIC and BIC values. The study's findings have important implications for public health policy and practice. Accurate forecasting of measles cases can help healthcare authorities prepare for potential outbreaks, allocate resources effectively, and implement targeted interventions to prevent the spread of the disease. Based on the study's findings, the following recommendations are made: Healthcare authorities should consider using ES and $SARIMA(2,1,1)(0,1,1)_{12}$ models for forecasting measles cases, as they have been shown to perform well in this study. The forecasting models should be integrated with existing surveillance systems to provide timely and accurate predictions of measles cases. The forecasting models should be regularly updated to reflect changes in the data and to ensure that they remain accurate and reliable.
Generally

REFERENCES

[1]. Adegboye, O. A., Adegboye, M., & He, J. (2017). Forecasting the dynamics of measles in Nigeria: Model comparison and implications for control. *BMC Infectious Diseases, 17*(1), 1-12.

[2]. Box, G. E. P., Jenkins, G. M., & Reinsel, G. C. (2015). Time series analysis: Forecasting and control. John Wiley & Sons.

[3]. Box, G. E., & Jenkins, G. M. (1976). *Time Series Analysis: Forecasting and Control*. Holden-Day.

[4]. Brockwell, P. J., & Davis, R. A. (2016). *Introduction to Time Series and Forecasting*. Springer.

[5]. Chen, X., Zhang, Y., & Li, Z. (2017). Measles prediction based on ARIMA model. Journal of Medical Systems, 41(10), 210.

[6]. Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and Practice*. OTexts.

[7]. Hyndman, R. J., Koehler, A. B., Ord, J. K., & Snyder, R. D. (2008). Forecasting with exponential smoothing: The state space approach. Springer.

[8]. Li, Q., Zhang, Y., & Chen, X. (2020). Measles prediction based on SARIMA model. Journal of Intelligent Information Systems, 56(2), 257-269.

[9]. World Health Organization. (2019). Measles. Retrieved from (link unavailable)

[10]. World Health Organization (WHO). (2021). Measles surveillance and outbreak response. Retrieved from https://www.who.int

[11]. Zhang, G., Patuwo, B. E., & Hu, M. Y. (2003). Forecasting with artificial neural networks: The state of the art. International Journal of Forecasting, 19(3), 361-377.

[12]. Zhang, Y., Chen, X., & Li, Z. (2019). Measles prediction based on ARIMA model with seasonal component. Journal of Medical Systems, 43(10), 210.

[13]. Zhou, L., Wang, Y., & Liu, Q. (2020). Predicting infectious disease using deep learning models: A review. *International Journal of Environmental Research and Public Health, 17*(17), 6318.