Evaluating Time-Series Models for Dengue Haemorrhagic Fever Prediction: Arima vs. Sarima and Exponential Smoothing

Farhan Haziq Baderul Hisam[1], Muhammad Harith Syafiq Zainudin[1], Muhammad Hakim Ikram Ismail[1], Nur Afriza Binti Baki[2], Nur Hanisah Binti Abdul Malek[2], Amiruddin Bin Ab Aziz[2,*]

[1]Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA Cawangan Terengganu, Kampus Kuala Terengganu, 21080 Kuala Terengganu, Terengganu, Malaysia.

[2]Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA Cawangan Terengganu, Kampus Bukit Besi, 23200 Bukit Besi, Terengganu, Malaysia

***Corresponding Author**

DOI: https://doi.org/10.47772/IJRISS.2026.100800190

Received: 15 August 2026; Accepted: 20 August 2026; Published: 29 August 2026

ABSTRACT

Accurately forecasting the transmission of infectious diseases, such as Dengue Haemorrhagic Fever (DHF), is essential for robust public health planning. This study assesses the performance of three time-series forecasting models—Exponential Smoothing, Autoregressive Integrated Moving Average (ARIMA), and Seasonal ARIMA (SARIMA)—in predicting annual DHF cases in Malaysia from 2011 to 2021. Model accuracy was evaluated using the Mean Absolute Percentage Error (MAPE). Results indicate that ARIMA and SARIMA yielded the lowest error rates (73.34%), whereas Exponential Smoothing performed less effectively (76.79%). ARIMA was preferred due to its simplicity and comparable predictive power. Although these error rates exceed the standard 10–20% threshold for reliable forecasting, the study demonstrates that these models have significant potential when integrated with more granular data and external covariates. Furthermore, a three-segment piecewise nonlinear regression was employed to identify structural changes in DHF trends. This combined analytical approach provides deeper insights into disease progression and strengthens the evidence base for effective public health interventions.

Keywords: ARIMA, curve fitting, dengue forecasting, exponential smoothing, SARIMA

INTRODUCTION

Dengue Haemorrhagic Fever (DHF) is a severe, potentially fatal condition caused by the dengue virus, which is primarily transmitted by *Aedes aegypti* and *Aedes albopictus* mosquitoes (Guzman et al., 2025). Tropical nations, including Malaysia, frequently face significant dengue outbreaks driven by a complex interplay of environmental and socioeconomic factors (Shepard et al., 2013). Consequently, accurate forecasting of dengue incidence is vital for optimizing public health planning and resource allocation.

Time-series forecasting methods, such as Exponential Smoothing, Autoregressive Integrated Moving Average (ARIMA), and Seasonal Autoregressive Integrated Moving Average (SARIMA), are widely utilized in epidemiological research (Hasan et al., 2024). These models are particularly effective at capturing temporal patterns and seasonal variations in disease incidence (Box et al., 2015). Exponential Smoothing models, which prioritize recent observations, are well-suited for short-term predictions (Hyndman & Athanasopoulos, 2018). In contrast, ARIMA and SARIMA integrate autoregressive, differencing, and moving average components; notably, SARIMA extends this framework by explicitly accounting for seasonal components within the data (Makridakis et al., 2020).

These forecasting models have proven their applicability in various epidemiological contexts. For instance, ARIMA and SARIMA models have provided predictive insights into dengue transmission in Bangladesh

(Hasan et al., 2024). Furthermore, researchers have employed SARIMA and its climate-augmented variant (SARIMAX) to develop dengue prediction models tailored to Tawau, Malaysia (Jayaraj et al., 2019). Building upon this body of work, this study aims to develop a reliable predictive model for DHF by applying these time-series techniques to historical dengue data in Malaysia.

LITERATURE REVIEW

Dengue Haemorrhagic Fever (DHF) remains a significant public health challenge in Malaysia. The country's tropical climate facilitates the proliferation of *Aedes aegypti* and *Aedes albopictus* mosquitoes, which serve as the primary vectors for the virus. The incidence of dengue in Malaysia has increased substantially over the past few decades, with official statistics indicating a rise from 44.3 cases per 100,000 population in 1999 to 181 cases in 2008 (MoH, 2015). Consequently, accurate forecasting of DHF cases is essential for proactive early intervention and optimized resource allocation. Exponential Smoothing (ES) represents one of the foundational time-series techniques for short-term forecasting. Valued for its computational simplicity and efficiency in capturing immediate patterns, it is effective at smoothing noisy data and adapting quickly to recent trends (Kumar et al., 2024; Cantarella et al., 2025). However, its limited capability to explicitly model seasonality restricts its utility for diseases like dengue, which typically exhibit strong seasonal cycles. The Autoregressive Integrated Moving Average (ARIMA) model, introduced by Box and Jenkins, is widely utilized in time-series forecasting for its capacity to model linear trends in stationary data. It has been applied successfully across numerous epidemiological studies, proving effective for short-term projections (ArunKumar et al., 2022; Siddique et al., 2025). Nevertheless, its performance diminishes when applied to datasets characterized by strong seasonality.

To address the limitations of ES and ARIMA, the Seasonal ARIMA (SARIMA) model was selected for this study. SARIMA is particularly effective for analysing time-series data with periodic behaviour, a feature widely recognized in healthcare, logistics, and energy forecasting (Li et al., 2025; Feng et al., 2025). In addition to these time-series models, this study employs nonlinear regression via piecewise curve fitting to detect structural changes in the data. By segmenting the data into distinct intervals with varying slopes, this method can identify critical turning points in disease trends that often correspond to major outbreaks or shifts in public health intervention efforts (Elshaar et al., 2025). Model performance is evaluated using the Mean Absolute Percentage Error (MAPE). While MAPE is a popular and intuitive metric in time-series analysis, it can be biased in the presence of small actual values (McKenzie et al., 2011). Despite these limitations, it remains a standard choice for benchmarking accuracy in disease forecasting. Ultimately, the combined application of ES, ARIMA, SARIMA, and piecewise curve fitting enables a comprehensive approach to DHF trend analysis, providing a comparative evaluation to inform the selection of effective tools for public health planning and response in Malaysia.

METHODOLOGY

The flowchart outlines a sequential research methodology for evaluating and selecting an optimal forecasting model or analytical technique based on accuracy.

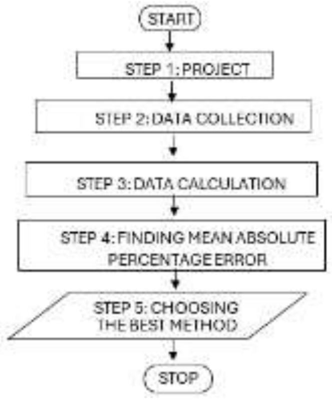

Figure 1. Research flow of Methodology

ISSN No. 2454-6186 | DOI: 10.47772/IJRISS | Volume X Issue VIII August 2026

Model and Data

This study forecasts annual Dengue Hemorrhagic Fever (DHF) cases in Malaysia using three time series models: Double Exponential Smoothing, Autoregressive Integrated Moving Average (ARIMA), and Seasonal ARIMA (SARIMA). Annual DHF case data for the period 2011–2021 was obtained from the official Malaysia Open Data Portal (archive.data.gov.my) on order to ensuring data reliability.

Model Calibration

The Double Exponential Smoothing method was implemented in Microsoft Excel. By using a smoothing constant α=0.8. That is selected based on trial performance. The ARIMA and SARIMA models were developed in RStudio using the forecast and fable packages.

For ARIMA, the data was first tested for stationarity using the Augmented Dickey-Fuller test. Differencing was applied where necessary. The (p, d, q) parameters were selected using Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots.

For SARIMA, the dataset was converted into a time-aware tribble structure with both non-seasonal (p, d, q) and seasonal (P, D, Q, s) parameters were determined automatically by the fitting procedure. That is with manual validation against residual diagnostics.

Equation

The general equation for Exponential Smoothing is:

$$F_{t+m} = \alpha y_t + (1-\alpha)F_t \quad ,0 < a \leq 1, \ \ t > 0$$

Where, y_t is actual value at time t. α is smoothing constant. F_t is forecast demand for time *t*. F_{t+m} is forecast demand for time *t* plus period *m*.

The ARIMA model is expressed:

$$y_t = c + \emptyset_1 y_{t+1} + \cdots + \emptyset_p y_{t-p} + \theta_1 y_{t-1} + \cdots + \theta_q \varepsilon_{t-q} + \varepsilon_t$$

Where, y_t is actual value at time *t*. *c* is constant or intercept. $\emptyset_p$ is an autoregressive (AR) coefficient. θ_q is moving average (MA) coefficients. ε_t is random error (white noise) at time t. *p* is order of the autoregressive (AR) terms. *d* is degree of differencing to make the series stationary. *q* is the order of the moving average (MA) terms.

The SARIMA model extend ARIMA as:

$$(1-\emptyset_1 B)(1-B)^d(1-\Phi_1 B^m)(1-B^m)^D y_t = (1+\theta_1 B)(1+\Theta_1 B^m)\varepsilon_t$$

Where, y_t is actual value at time *t*. *m* represent seasonal period. *B* is a backshift operator. $(1-\emptyset_1 B)$ is a non-seasonal autoregressive (AR). $(1-\Phi_1 B^m)$ is seasonal autoregressive (AR). $(1-B)^d$ is non-seasonal difference. $(1-B^m)^D$ is seasonal difference. $(1+\theta_1 B)$ is non-seasonal moving average (MA). $(1+\Theta_1 B^m)$ is seasonal moving average (MA). ε_t is random error (white noise) at time t. *p* is order of the autoregressive (AR) terms $(1-\emptyset_1 B)$. *d* is degree of differencing to make the series stationary. *q* is the order of the moving average (MA) terms. P is seasonal autoregressive (AR). D is seasonal differencing and Q seasonal moving average (MA) terms.

Model Evaluation

The Mean Absolute Percentage Error (MAPE) was used to compare forecasting across models:

ISSN No. 2454-6186 | DOI: 10.47772/IJRISS | Volume X Issue VIII August 2026

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{\left|Y_i - \hat{Y}_i\right|}{Y_i} \times 100\%$$

where Y_i and $\hat{Y}$ represent actual and forecasted values. The model with the lowest MAPE was selected for final prediction.

Additionally, nonlinear regression via piecewise curve fitting (three-, four-, and five-segment) was conducted using Sigma Plot to detect structural changes in DHF trends. The best-fitting model was selected based on R^2 and error statistics.

RESULTS AND DISCUSSION

Exponential Smoothing

Table 1 presents the actual and forecasted annual DHF cases obtained using the Exponential Smoothing method. Whereas Figure 2 illustrates the comparison between observed and predicted values.

Table 1. Result of Exponential Smoothing

Year (X)	Total Cases (Y)	Forecast
2011	1418	1418.00
2012	719	1418.00
2013	774	858.80
2014	1101	790.96
2015	1040	1038.99
2016	635	1039.80
2017	406	715.96
2018	406	467.99
2019	523	418.40
2020	344	502.08
2021	64	375.62

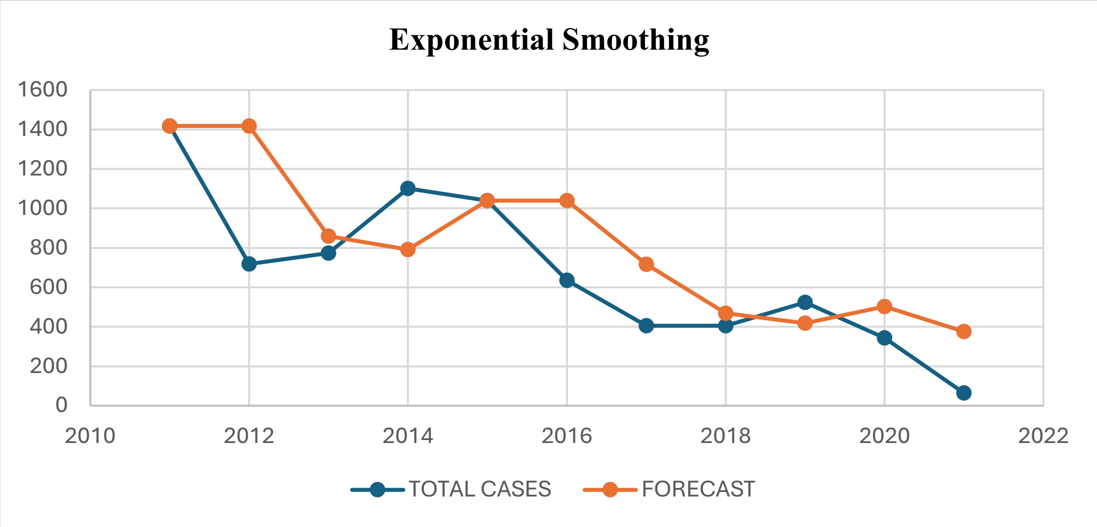

Figure 2. Result of Exponential Smoothing

Figure 2 illustrates the comparative performance between actual recorded cases and the predicted values generated using Single Exponential Smoothing over an 11-year period (2011–2021). Overall, both actual cases and forecasted values demonstrate a significant downward trajectory, declining from approximately 1,420 cases in 2011 to fewer than 400 cases by 2021.

Arima

The table below displays the actual data together with the forecasted data from the Autoregressive Integrated Moving Average (ARIMA). While the figure below the table displayed the chart graph of Autoregressive Integrated Moving Average (ARIMA).

Table 2. Result of ARIMA

Year (X)	**Total Cases (Y)**	**Forecast**
2011	1418	913.24
2012	719	998.31
2013	774	516.02
2014	1101	945.89
2015	1040	860.07
2016	635	851.10
2017	406	497.76
2018	406	606.19
2019	523	515.82
2020	344	693.91
2021	64	357.50

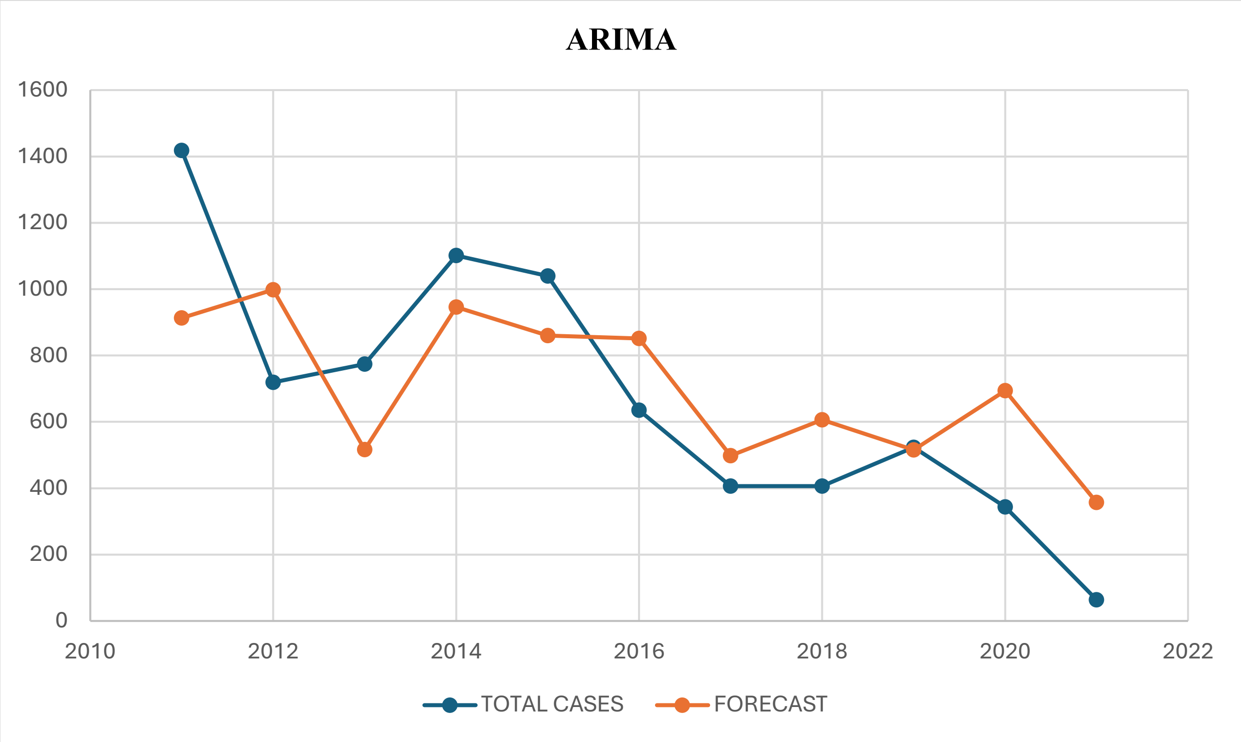

Figure 3. Result of ARIMA

Figure 3 illustrates the forecasting performance of the ARIMA model compared against actual recorded cases over an 11-year period (2011–2021). The model captures the overall downward trend alongside key directional shifts, such as synchronously forecasting the 2014 peak (~950 predicted vs. ~1,100 actual) and intersecting accurately at the 2019 data point (~520 cases). However, ARIMA exhibits significant forecast error during sudden extreme fluctuations, most notably underpredicting the 2011 baseline (~910 predicted vs. ~1,420 actual) and overpredicting the sharp final drop in 2021 (~360 predicted vs. ~65 actual).

Sarima

The table below displayed the actual data together with the forecasted data from the Seasonal Autoregressive Integrated Moving Average (SARIMA). While the diagram below the table displays the chart graph of Seasonal Autoregressive Integrated Moving Average (SARIMA).

Table 3. Result of SARIMA

Year (X)	**Total Cases (Y)**	**Forecast**
2011	1418	913.24
2012	719	998.31
2013	774	516.02
2014	1101	945.89
2015	1040	860.07
2016	635	851.10
2017	406	497.76

ISSN No. 2454-6186 | DOI: 10.47772/IJRISS | Volume X Issue VIII August 2026

2018	406	606.19
2019	523	515.82
2020	344	693.91
2021	64	357.50

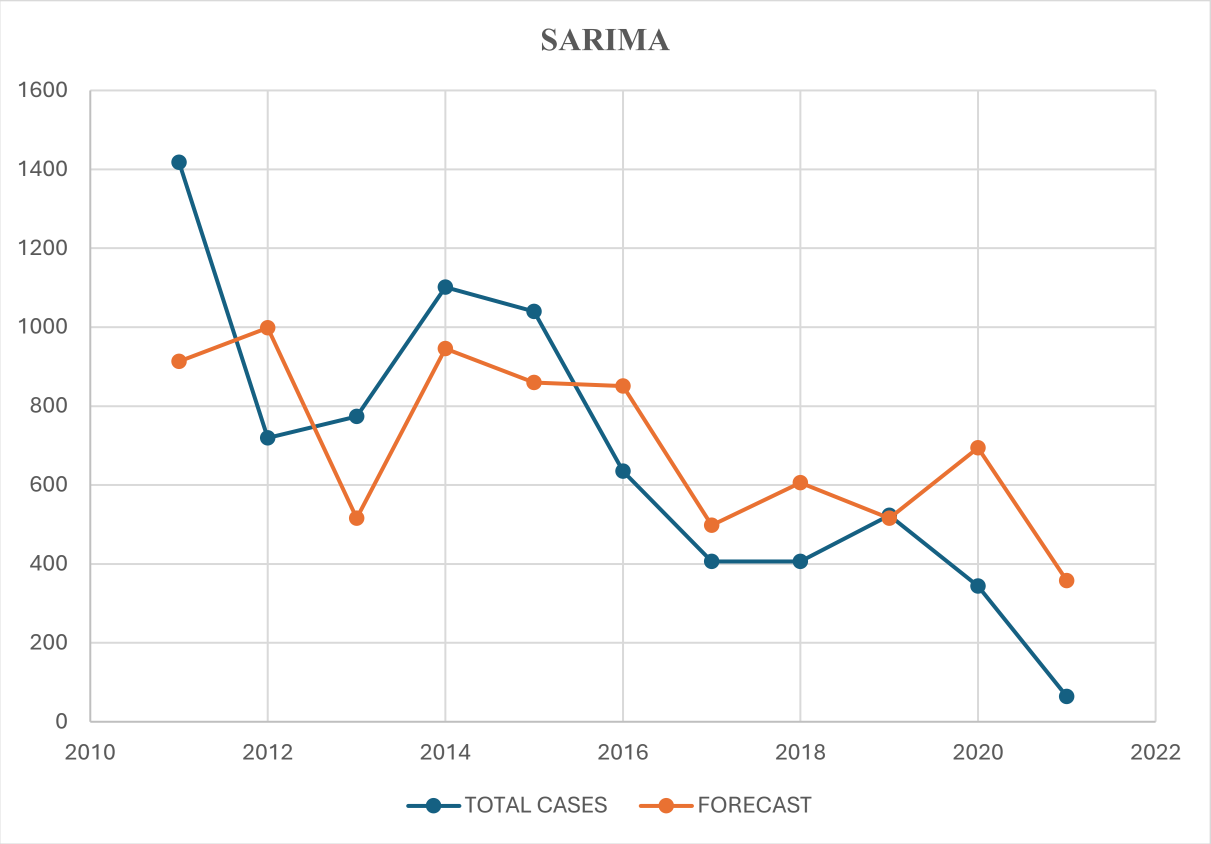

Figure 4. Result of SARIMA

Figure 4 illustrates the empirical and predicted case values generated by the Seasonal Autoregressive Integrated Moving Average (SARIMA) model from 2011 to 2021. Overall, SARIMA captures the downward trajectory of the series while correctly signalling directional inflections, such as the peak observed in 2014 and the convergence point in 2019.

However, because the dataset consists of annual observations without multi-period seasonal dynamics within each year, the seasonal parameters provide minimal distinct adjustments over standard ARIMA modelling. Consequently, the model encounters substantial tracking error at the boundaries, severely underestimating the initial 2011 peak (~910 predicted vs. ~1,420 actual) and overestimating the trailing edge in 2021 (~360 predicted vs. ~65 actual).

Error Analysis

The graph bellow represents the actual data, Exponential Smoothing, Seasonal Autoregressive Integrated Moving Average (SARIMA) and Autoregressive Integrated Moving Average (SARIMA). In order to visualize which method is the closes to achieve the best fit.

ISSN No. 2454-6186 | DOI: 10.47772/IJRISS | Volume X Issue VIII August 2026

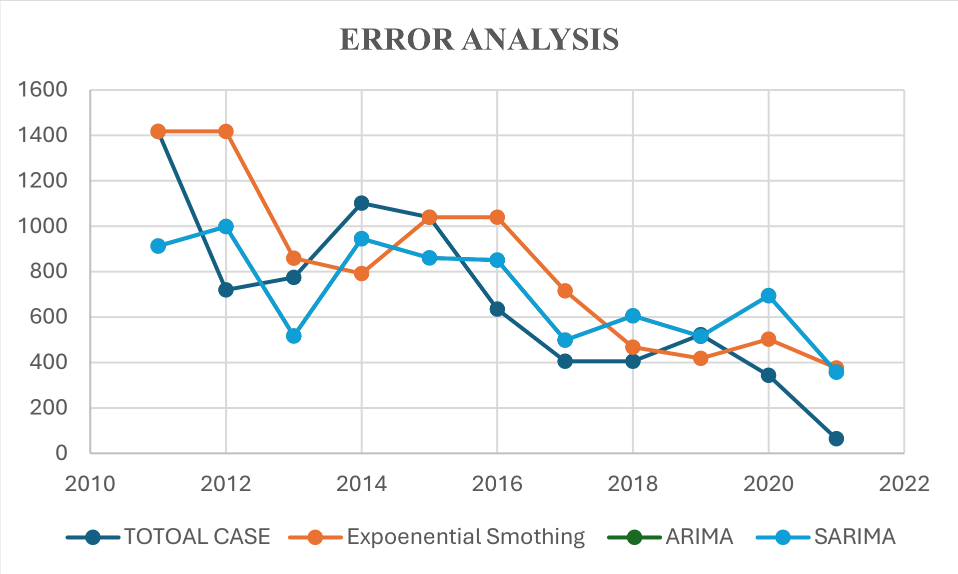

Figure 5. Error Analysis

The table below present Mean Absolute Percentage Error (MAPE) of each method. To evaluate which methods has the least amount of error.

Table 4. Error Analysis

Methods	**MAPE**
Exponential Smoothing	76.79%
ARIMA	73.34%
SARIMA	73.34%

Exponential Smoothing produced the highest error among the evaluated models, yielding a Mean Absolute Percentage Error (MAPE) of 76.79%. In contrast, both ARIMA and SARIMA demonstrated superior predictive performance with an identical MAPE of 73.34%, exhibiting lower deviation from the actual recorded cases compared to Exponential Smoothing. Given their identical accuracy metrics, ARIMA is preferred for operational deployment due to its lower computational complexity and structural simplicity.

Overall, the empirical findings confirm that ARIMA and SARIMA outperform Exponential Smoothing in predicting Dengue Haemorrhagic Fever (DHF) cases. Although SARIMA incorporates seasonal parameters, it failed to provide an accuracy advantage over standard ARIMA for this dataset. This lack of improvement is primarily attributable to the absence of strong, multi-period seasonal dynamics in annual case counts.

Despite outperforming Exponential Smoothing, both ARIMA and SARIMA exhibited exceptionally high MAPE values exceeding 70%, indicating limited overall predictive precision. This substantial forecast error is likely driven by unmodeled exogenous drivers, such as climatic variables, vector control interventions, and population mobility. Future research could enhance predictive accuracy by integrating these external covariates or using higher-frequency data (e.g., weekly or monthly case counts) to capture underlying temporal dynamics more effectively.

CONCLUSION

The empirical evaluation of the forecasting models based on Mean Absolute Percentage Error (MAPE) revealed that both ARIMA and SARIMA achieved the lowest overall error rate of 73.34%. In contrast, the Exponential Smoothing method recorded the highest error at 76.79%. These outcomes demonstrate that

ARIMA and SARIMA yield predictions that align more closely with actual Dengue Haemorrhagic Fever (DHF) case data than Exponential Smoothing.

Although SARIMA extends standard ARIMA modelling to capture seasonal dynamics, the identical error rates between the two models indicate that seasonal variations are weak or negligible within this dataset. Conversely, Exponential Smoothing—which lacks the structural capacity to account for temporal autocorrelation—exhibited the lowest overall accuracy among all evaluated methods.

Consequently, the core objectives of this study were successfully achieved through the implementation and comparative assessment of Exponential Smoothing, ARIMA, and SARIMA models for forecasting DHF incidence in Malaysia. Balancing predictive precision with structural simplicity, ARIMA emerged as the optimal forecasting approach. Utilizing the ARIMA framework, the projected number of DHF cases for 2022 is estimated at 423. These findings underscore ARIMA's utility as a robust tool for short-term epidemiological trend forecasting in contexts where seasonal effects are minimal.

ACKNOWLEDGEMENT

The contributions by all are acknowledged.

REFERENCES

1. M. G. Guzman, M. C. Marqueti, E. Martinez, and A. B. Perez, "Dengue, Dengue hemorrhagic fever," in International Encyclopedia of Public Health (Third Edition), S. R. Quah Ed. Oxford: Academic Press, 2025, pp. 275-319.
2. D. S. Shepard, E. A. Undurraga, and Y. A. Halasa, "Economic and disease burden of dengue in Southeast Asia," PLoS neglected tropical diseases, vol. 7, no. 2, p. e2055, 2013.
3. R. J. Hyndman and G. Athanasopoulos, Forecasting: principles and practice. OTexts, 2018.
4. S. Makridakis, R. J. Hyndman, and F. Petropoulos, "Forecasting in social settings: The state of the art," International Journal of Forecasting, vol. 36, no. 1, pp. 15-28, 2020/01/01/ 2020, doi: https://doi.org/10.1016/j.ijforecast.2019.05.011.
5. P. Hasan, T. D. Khan, I. Alam, and M. E. Haque, "Dengue in Tomorrow: Predictive Insights From ARIMA and SARIMA Models in Bangladesh: A Time Series Analysis," Health Science Reports, Article vol. 7, no. 12, 2024, Art no. e70276, doi: 10.1002/hsr2.70276.
6. V. J. Jayaraj, R. Avoi, N. Gopalakrishnan, D. B. Raja, and Y. Umasa, "Developing a dengue prediction model based on climate in Tawau, Malaysia," Acta Tropica, vol. 197, p. 105055, 2019/09/01/ 2019, doi: https://doi.org/10.1016/j.actatropica.2019.105055.
7. M. MoH, "Clinical Practice Guidelines: Management of Dengue Infection in Adults," ed: Putrajaya, Malaysia: Ministry of Health Malaysia, 2015.
8. L. Kumar, K. Sharma, and U. K. Khedlekar, "Dynamic pricing strategies for efficient inventory management with auto-correlative stochastic demand forecasting using exponential smoothing method," Results in Control and Optimization, vol. 15, p. 100432, 2024/06/01/ 2024, doi: https://doi.org/10.1016/j.rico.2024.100432.
9. G. CANTARELLA, C. FIORI, and P. VELONÀ, "Moving average vs. exponential smoothing cost-updating filters for day-to-day dynamic assignment: fixed-point stability and bifurcation theoretical analysis," Transportation Research Part B: Methodological, vol. 199, p. 103253, 2025.
10. K. E. ArunKumar, D. V. Kalaga, C. Mohan Sai Kumar, M. Kawaji, and T. M. Brenza, "Comparative analysis of Gated Recurrent Units (GRU), long Short-Term memory (LSTM) cells, autoregressive Integrated moving average (ARIMA), seasonal autoregressive Integrated moving average (SARIMA) for forecasting COVID-19 trends," Alexandria Engineering Journal, vol. 61, no. 10, pp. 7585-7603, 2022/10/01/ 2022, doi: https://doi.org/10.1016/j.aej.2022.01.011.
11. M. A. B. Siddique, B. Mahalder, M. M. Haque, and A. S. Ahammad, "Forecasting air temperature and rainfall in Mymensingh, Bangladesh with ARIMA: Implications for aquaculture management," Egyptian Journal of Aquatic Research, 2025.

12. J. Li, X. Xu, H. R. Karimi, and Z. Wu, "Optimization of electric vehicle charging strategies in residential integrated energy systems: A SARIMA model approach for dynamic electricity prices," Energy, p. 136600, 2025.
13. W. Feng et al., "Platelet demand forecasting based on the SARIMA model: optimizing blood bank resource allocation and clinical supply," Transfusion Clinique et Biologique, vol. 32, no. 2, pp. 185-194, 2025.
14. M. E. Elshaar and N. A. Qasem, "Enhanced Prediction of Airfoil's Drag Coefficient using Curve Fitting and Artificial Neural Network," Transportation Research Procedia, vol. 84, pp. 641-648, 2025.
15. J. McKenzie, "Mean absolute percentage error and bias in economic forecasting," Economics Letters, vol. 113, no. 3, pp. 259-262, 2011.