ORIGINAL ARTICLE

DEVELOPMENT OF A UNIVARIATE TIME SERIES MODEL FOR FORECASTING DENGUE HEMORRHAGIC FEVER CASES IN NAKHON SI THAMMARAT

Onuma Ruksachol[1], Siriluk Sintupachee[1*], Manit Pollar[2] and Suppawan Promprao[2]

[1] Program in Creative Innovation in Science and Technology, Faculty of Science and Technology, Nakhon Si Thammarat Rajabhat University, Nakhon Si Thammarat 80280, Thailand
[2] Program in Mathematics, Faculty of Science and Technology, Nakhon Si Thammarat Rajabhat University, Nakhon Si Thammarat 80280, Thailand

***Corresponding author: Siriluk Sintupachee**
Email: siriluk_sint@nstru.ac.th

ABSTRACT

Dengue hemorrhagic fever (DHF) has been associated with numerous hospitalizations and a high mortality rate in Thailand, especially in the southern parts. In the present study, an epidemiological realistic model was constructed to estimate the trend of outbreaks of DHF using epidemic data from 2010 to 2020. The SARIMA model with the Box-Jenkins approach was employed. Bayesian Information Criteria (BIC), root mean square error (RMSE), and mean absolute percentage error (MAPE) were used to determine their accuracy. The results showed that the SARIMA (2,0,1)(1,0,0)$_{12}$ model fitted the Nakhon Si Thammarat endemic data. Their accuracy had the smallest BIC, MAPE, and RMSE values, yielding 9.637, 848.743, and 214.661, respectively. To summarize, the DHF ARIMA model could be used to forecast the incidence of DHF in other locations and develop public health initiatives to prevent and treat the condition.

Keywords: *Univariate, Box-Jenkins, SARIMA, dengue hemorrhagic fever, database*

INTRODUCTION

Dengue hemorrhagic fever (DHF) is caused by a virus that is transmitted to humans via female *Aedes* mosquitoes[1-2]. DHF is common in tropical and subtropical locations, largely affecting Asian and Latin American countries. In these countries, DHF is becoming a major cause of hospitalizations and fatalities, with the World Health Organization (WHO) predicting that 500,000 people require hospitalization annually, with roughly 2.5% of them dying[3-4]. DHF became more common in Thailand over the preceding decade, with the first big epidemic being documented in Bangkok in 1958. There were 2,158 cases and 300 fatalities because of this pandemic[5].

Every Thai province currently keeps an annual database of DHF cases. Nakhon Si Thammarat is one of the areas hit by the outbreak. The pattern is every 3 to 5 years during the rainy season, when a major dengue-endemic outbreak develops, with the top 3 months being in the decreasing order of July, June, and August. If the pandemic is untreated, the number of DHF patients is expected to increase to 10,000 to 20,000 annually.

Furthermore, 1,660 DHF episodes between January 1 and July 8, 2019, caused six deaths. However, certain areas could witness outbreaks at the end of the year or throughout the year, albeit insignificantly, and epidemic months may vary slightly depending on the year and region[6-8].

It is highly crucial to understand the factors contributing to dengue incidence and identify the ideal period for vector management. However, researchers should be very well aware of the dengue seasonal cycle. Mathematical and statistical models are highly valuable tools to analyze transmission patterns and variations. Previous endemic distributions of DHF are used to forecast epidemiological modeling and anticipate future trends and outbreaks. For instance, the widespread use of DHF monitoring played a crucial role in determining the underlying epidemiological reasons and therapy effectiveness. The Seasonal Autoregressive Integrated Moving Average (SARIMA) model, consisting of a collection of time-series data with a single variable, is popular for forecasting and explaining the occurrence of a disease.

SARIMA models, which have previously been employed in a range of prognostic studies, can efficiently forecast the number of cases in several illnesses, including DHF. In Brazil (Ribeiro Preto, Sao Paulo State), a SARIMA model was developed to estimate the number of DHF patients in 2009, and the SARIMA (2,1,3)(1,1,1)$_{12}$ was demonstrated to be a feasible model[1]. Dengue fever data from January 1995 to July 2017 in Fiji were analyzed to predict the incidence from August 2017 to December 2018, using ARIMA as the best-fitting model(3,0,4), which has a lower Bayesian information criterion (BIC) and a lower mean absolute percentage error (MAPE). In addition, it displayed excellent performance in predicting the future incidence of dengue cases[9].

The SARIMA $(6,0,3)(0,1,1)_{52}$ model was selected to estimate the incidence of dengue fever in Thailand, and its accuracy was evaluated using Akaike's information criteria (AIC), BIC, and root mean square error (RMSE)[10]. In Rayong, Thailand, a seasonal SARIMA model$(1,0,2)(0,0,3)_{12}$ was developed from DHF case data (January 2014–December 2015)[11]. Therefore, a unique forecasting approach is used depending on data dispersion in a particular area. We used the secondary data from 2010 to 2019 to develop an appropriate univariate time series analysis model for predicting monthly cases in Nakhon Si Thammarat and forecasting DHF cases in 2020. BIC was used to select the best model, and model performance was measured using MAPE and RMSE.

METHODS

Data

The monthly DHF data in Nakhon Si Thammarat were obtained from the Department of Disease Control, Ministry of Public Health, Thailand. There were 132 cases of monthly data (case report data) for analysis, divided into two parts, namely, part 1 consisting of 120 datasets from January 2010 to December 2019 to construct the suitable model, and part 2, comprising 12 datasets from January 2020 to December 2020 to compare with the forecast values obtained from the final SARIMA model. Statistical Packages for Social Science (SPSS) and Microsoft Office Excel were used to develop the model.

Time series analysis

SARIMA$(p, d, q)(P, D, Q)_s$ is an application of the Autoregressive Integrated Moving Average model, where p, d, and q are the same as in the ARIMA model; P, D, and Q are the seasonal autoregressive orders, number of seasonal differences, and seasonal moving average orders, respectively; and s is the seasonal period length. SARIMA generic model is depicted in equation (1)[12-13].

$$\phi_p(B)\Phi_P(B^s)(1-B)^d(1-B^s)^D Y_t = \theta_q(B)\Theta_Q(B^s)\varepsilon_t \quad (1),$$

where

Y_t is appropriately transformed in the period t.,

B is the backshift operator where $B^sY_t = Y_{t-s}$.

$\phi_p(B) = 1-\phi_1B-\phi_2B^2-\ldots-\phi_pB^p$ is a Non-seasonal autoregressive operator of order p: AR(p).

$\Phi_P(B^s) = 1-\Phi_1B^s-\Phi_2B^{2s}-\ldots-\Phi_PB^{Ps}$ is the seasonal autoregressive operator of order P: SAR(P).

$\theta_q(B) = 1-\theta_1B-\theta_2B^2-\ldots-\theta_qB^q$ is the Non-seasonal moving average operator of order q: MA(q).

$\Theta_Q(B^s) = 1-\Theta_1B^s-\Theta_2B^{2s}-\ldots-\Theta_QB^{Qs}$ is a seasonal moving average operator of order Q: SMA(Q).

d is the order of differencing, D is the order of seasonal differencing, s is the number of seasons per year, and ε_t is the random process.

It was applied using the following steps:

1. The time series data were graphed to confirm the consistency of the quality profile, with a constant mean and variation. If the time series data had non-stationary characteristics, it was converted using difference or seasonal difference, natural logarithms, square root transformation, or square transforming.
2. The calibration curves for the autocorrelation function (ACF) and partial autocorrelation function (PACF) of data were used to calculate the values of p, q, P, and Q, where p and P were defined by the PACF graph and q and Q were determined using the ACF graph. The values of p and q were the number of the initial correlation bars whose value exceeded the given borders, whereas the values of P and Q were the number of seasonal correlation bars whose value surpassed the set bounds.
3. The parameters of each viable model were estimated and decided whether to include each parameter in the model equation. Afterward, the forecast model was redefined, and the parameters were re-estimated until a forecast model with all key parameters was established by removing one insignificant parameter. Several models may exist.
4. The model with the lowest BIC value and no statistically significant Ljung-Box Q was selected as the forecast model. The forecast error $\{e_t\}$ did not show a normal distribution, as confirmed by the Kolmogorov-Smirnov test. The error levels were validated using ACF and PACF graphs or from the error movement curve concerning time. The (et, t) mean average was zero when computed using the *t*-test. Levene's test verified the same variance at all time intervals. If one of the error time series conditions was found to be untrue, the prediction model was considered inapplicable and not used for predictions
5. The forecast values and prediction discrepancies were calculated using MAPE and RMSE values as per the following formula:

$$MAPE = \frac{100}{n}\sum_{t=1}^{n}\left|\frac{e_t}{Y_t}\right| \quad (2)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}e_t^2} \quad (3),$$

Where $e_t = Y_t - \hat{Y}_t$ is the random at period time t.

Y_t and $\hat{Y}_t$ are actual values and forecasting values at a period t, respectively and n is the amount of data

6. Finally, the most suitable model was applied for forecasting[9-11,14-18].

RESULTS

The SPSS software was used to estimate the trend for retrospective data on DHF cases and a graph was obtained. The graph depicted in Figure 1(a) was used to evaluate the early data characterization. The natural logarithm transformation method and differencing were employed to stabilize the variance and eliminate the seasonal trend, respectively. Natural logarithm transformation suppressed the fluctuations, thereby enhancing the normality of the data (Figure 1 (b)). The PACF plot of the natural logarithm-transformed DHF case data, shown in Figure 1(d), depicted seasonality, which reduced slightly, whereas the ACF plot of the DHF case data, shown in Figure 1(c), tailed off after lag 4. In an initial attempt to resolve the non-stationarity of the time series (depicted in Figure 1(b)) and eliminate the trend and seasonality, non-seasonal differencing was employed.

The ACF and PACF plots were used to define the initial model. The ACF plot was terminated at lag 4, which had been parameterized to MA(4), with no seasonal lag (SMA(0)), as shown in Figure 1(c). Furthermore, as shown in Figure 1(d), the PACF figure was terminated at lag 3, with a seasonal lag (lag 12), which was parameterized to AR(3) and SAR (1) and SARIMA $(3,0,4)(1,0,0)_{12}$ was the outcome. In addition, Figure 1(e) ACF plots were trimmed out at lag 2 and had three seasonal lags (lag 12, lag 24, and lag 36), as shown in Figure 1(f). Because the PACF plot was cut off at lags 1 and 12, we adjusted the parameters to AR(1), SAR(1), MA(2), and SMA(3) with d = 1, yielding SARIMA$(1,1,2)(1,0,3)_{12}$, which demonstrated that the Ljung box of the models SARIMA$(3,0,4)(1,0,0)_{12}$ and SARIMA$(1,1,2)(1,0,3)_{12}$ was insignificant (*p*-value > 0.05), with R-squared values of 0.738 and 0.633, respectively. However, considering the parameters of the model, the *p*-value was greater than the 0.05 level of significance, as shown in Table 1, indicating that the parameter could be omitted from the predictable model. We removed extraneous parameters from the model and found three models that satisfied the assumption: SARIMA$(2,0,1)(1,0,0)_{12}$, SARIMA$(1,1,0)(1,0,0)_{12}$, and SARIMA$(0,1,2)(1,0,0)_{12}$. The Ljung box of the three models was insignificant (*p*-value > 0.05), with the value of R-squared and BIC shown in Table 1. All parameter estimations had significance at the level of 0.05.

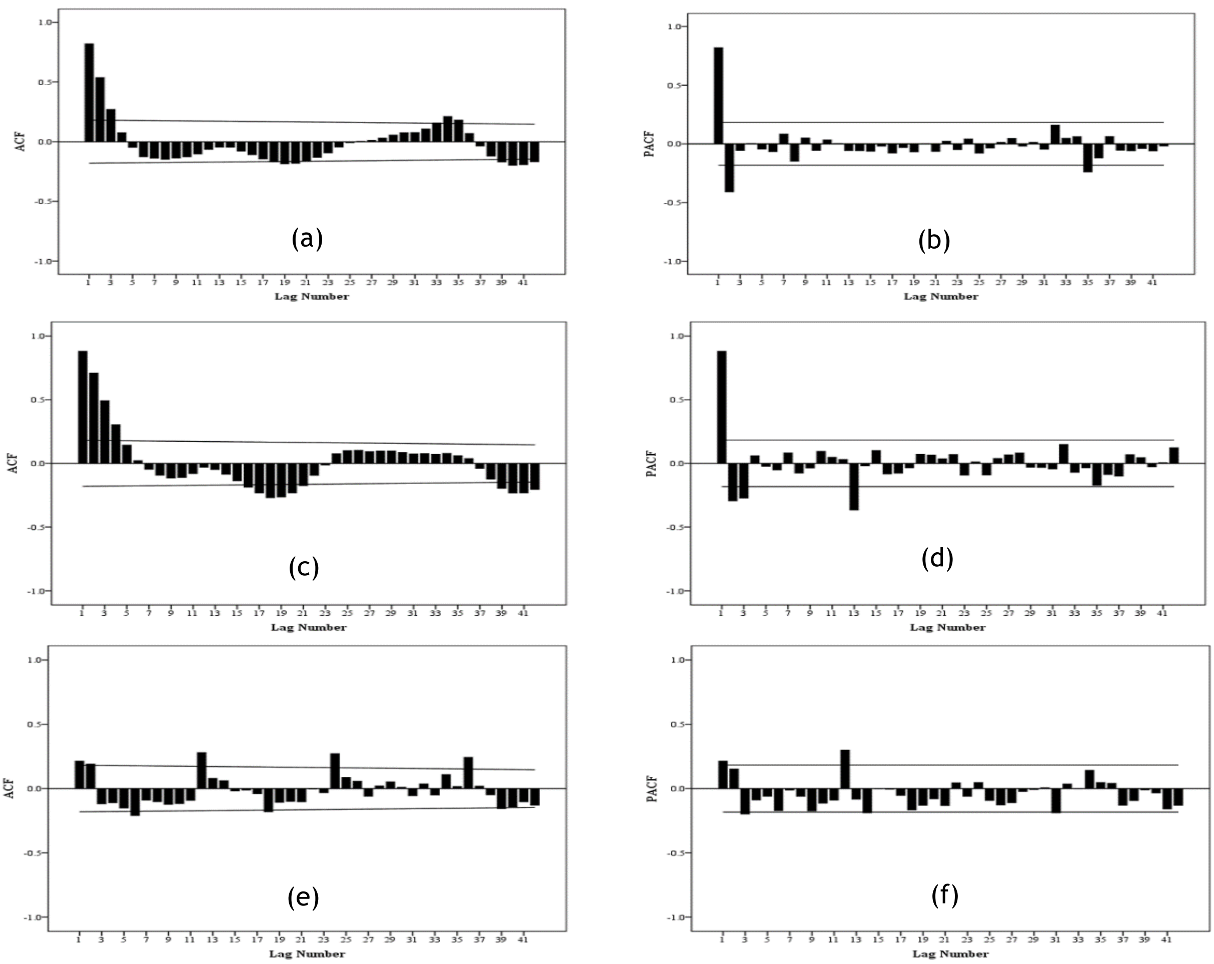


Figure 1: ACF and PACF plot of DHF cases in Nakhon Si Thammarat (a), (b) the original data (c), (d) natural logarithm, and (e), (f) the first differencing of the natural logarithm.

Table 1: Estimated parameters and model description

Model	Type	Lag	Estimate	Sig.	Model fit Statistics		Ljung BoxQ (18)	
					R-squared	*BIC*	*Statistics*	*Sig.*
SARIMA $(3,0,4)(1,0,0)_{12}$	Constant		5.502	0.000	0.738	9.902	9.382	0.496
	AR	1	1.540	0.000				
		2	-1.387	0.000				
		3	0.589	0.000				
	MA	1	0.420	0.954				
		2	-1.017	0.943				
		3	-0.104	0.981				
		4	-0.369	0.954				
	SAR	1	0.335	0.001				
SARIMA $(1,1,2)(1,0,3)_{12}$	Constant		0.225	0.179	0.633	10.200	15.800	0.148
	AR	1	-0.149	0.631				
	Diff	1	-	-				
	MA	1	-0.365	0.213				
		2	-0.329	0.001				
	SAR	1	0.981	0.000				
	SMA	1	0.841	0.012				
		2	0.065	0.667				
		3	-0.029	0.841				
SARIMA $(2,0,1)(1,0,0)_{12}$	Constant		5.320	0.000	0.744	9.637	19.344	0.152
	AR	1	1.704	0.000				
		2	-0.784	0.000				
	MA	1	0.596	0.000				
	SAR	1	0.353	0.000				
SARIMA $(1,1,0)(1,0,1)_{12}$	AR	1	0.207	0.027	0.629	9.916	25.378	0.050
	Diff	1	-	-				
	SAR	1	0.989	0.000				
	SMA	1	0.898	0.001				
SARIMA $(0,1,2)(1,0,0)_{12}$	Diff	1	-	-	0.570	10.065	14.490	0.489
	MA	1	-0.258	0.004				
		2	-0.341	0.000				
	SAR	1	0.393	0.000				

The forecast error parameters of all three models were examined, which showed that the error followed a normal distribution (Kolmogorov-Smirnov, *p*-value > 0.05). Consistent variance across all intervals (Levene's test, *p*-value > 0.05), a mean of zero (*t*-test, *p*-value > 0.05), and residual ACF and PACF values were displayed within the 95% confidence interval (Table 2).

Table 2: Diagnostic Test

Models	Kolmogorov-Smirnov		Levene Test		One sample t-test	
	Z	*Sig*	*Z*	*Sig.*	*T*	*Sig.*
SARIMA$(2,0,1)(1,0,0)_{12}$	0.056	0.200	1.190	0.310	-0.067	0.947
SARIMA$(1,1,0)(1,0,1)_{12}$	0.040	0.200	1.340	0.230	0.097	0.923
SARIMA$(0,1,2)(1,0,0)_{12}$	0.052	0.200	1.300	0.250	-0.045	0.965

The MAPE and RMSE values of the three models indicate that the SARIMA$(2,0,1)(1,0,0)_{12}$ model was the most accurate forecasting because it delivered the anticipated number that differs the least from the actual data (Table 3) and had the lowest MAPE and RMSE values.

Table 3: Model accuracy of SARIMA$(2,0,1)(1,0,0)_{12}$, SARIMA$(1,1,0)(1,0,1)_{12}$ and SARIMA$(0,1,2)(1,0,0)_{12}$

Models	Fitting performance		Forecasting performance	
	MAPE	RMSE	MAPE	RMSE
SARIMA$(2,0,1)(1,0,0)_{12}$	29.333	112.029	848.743	214.661
SARIMA$(1,1,0)(1,0,1)_{12}$	30.104	135.029	1819.076	515.330
SARIMA$(0,1,2)(1,0,0)_{12}$	30.462	144.627	1782.310	428.841

The Box-Jenkins equation of the SARIMA$(2,0,1)(1,0,0)_{12}$ model could be written as follows:

$$\begin{aligned} Z_t &= 5.32 + \varepsilon_t - 0.596\varepsilon_{t-1} + 1.704Z_{t-1} \\ &- 0.784Z_{t-2} + 0.353Z_{t-12} - 0.602Z_{t-13} \\ &+ 0.277Z_{t-14} \end{aligned} \quad (4)$$

when $\log(Y_t) = Z_t$

The time-series parameters were created using lag intervals based on the dengue incidence over a year (s = 12). However, the model could vary based on the location, and different versions will be deployed in other regions. The SARIMA$(2,0,1)(1,0,0)_{12}$ model was used to forecast DHF cases, and the results are presented in Figure 2. The graph indicates that the anticipated model matched the real data well.

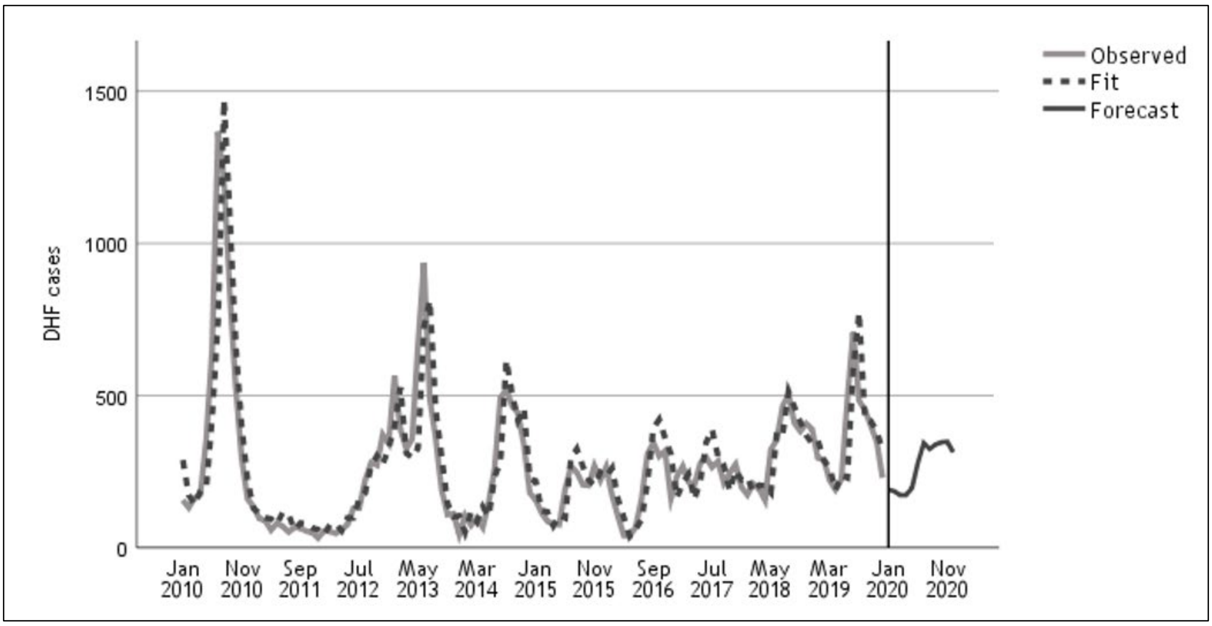
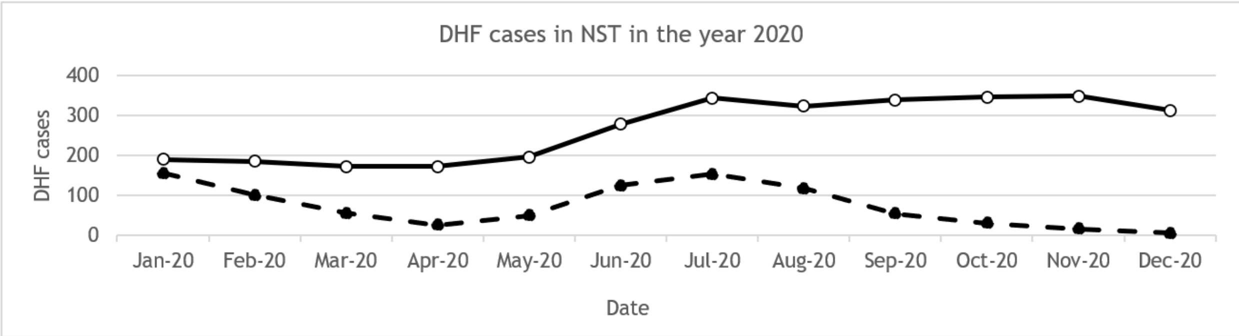


Figure 2: Comparative dengue incidence in 2010-2020 between actual data and forecast data from SARIMA$(2,0,1)(1,0,0)_{12}$.

Afterward, the model was used to forecast a 12-month dataset from January 2020 to December 2020 while accounting for seasonality in a year, as illustrated in Figure 2. Data predictions, according to the graph, were more useful than the actual data. This could be attributable to a more targeted and comprehensive dengue illness education program. The statistics, in contrast, demonstrated similar seasonal trends. In addition, it was demonstrated that DHF has the propensity to increase the number of patients beginning in April and continuing through June. The disparity between these figures could be related to a variety of environmental variables as well as healthcare professionals' treatments.



	Jan-20	Feb-20	Mar-20	Apr-20	May-20	Jun-20	Jul-20	Aug-20	Sep-20	Oct-20	Nov-20	Dec-20
Actual	157	102	56	27	51	125	154	118	55	31	17	6
Forecast	191	186	173	173	197	279	344	324	340	347	349	314
UCL	350	440	487	546	665	978	1231	1167	1222	1247	1257	1138
LCL	93	61	42	33	33	42	50	47	49	50	50	44

Figure 3: Actual and anticipated DHF cases for 2020

DISCUSSION

The best SARIMA model developed in this study for predicting dengue cases was SARIMA(2,0,1)(1,0,0)$_{12}$, with the lowest MAPE and RMSE values and the highest R^2 value. In addition, it displayed appropriate model requirements, which were confirmed by inspecting the autocorrelation function (ACF) and partial autocorrelation function (PACF) graphs, which can be described as ACF and PACF in the first and second months (p = 2)˝ of patients with dengue fever in the following month and the previous season (P = 1, s = 12).˝

The prediction made using the model in the present study were similar to those done using Rio de Janeiro's SARIMA(2,0,0)(1,0,0)$_{12}$ and Sao Paulo's SARIMA(2,1,2)(1,1,1)$_{12}$, where predicting the number of dengue cases implied that the number of dengue cases in a month can be estimated using the number of dengue cases occurring 1, 2, and 12 months earlier[19-20]. However, certain differences existed in the steps followed to evaluate the data. For instance, the Sao Paulo model used stationary data based on non-seasonal and seasonal differentiation, whereas our model stabilized the data using a natural logarithm.

A seasonal SARIMA(1,0,0)(0,1,1)$_{12}$ model, confirmed by the mean square error (MSE) value, was constructed and reported as the best in predicting the dengue cases in Indonesia[15]. Similar to India, the subject SARIMA(1,0,0)(0,1,1)$_{12}$ was found to be the best for predicting the number of dengue fever patients[14]. In contrast to the Fiji model, the ARIMA(3,0,4)˝ model displayed no seasonal trend for DHF cases[9]. The p, d, q and P, D, Q values of the seasonal ARIMA model could differ due to environmental factors and the availability of healthcare services in different areas[14].

The fitting performance of MAPE = 29.333 and RMSE = 112.029 were the best for our model, with a MAPE value less than 50, which was suitable for forecasting[21]. This is consistent with the SARIMA(1,1,1)(0,1,1)$_{12}$ model, with MAPE = 40.591, resulting in 39.840 accuracies in forecasting the incidence of mumps in China[22]. The 14 models had MAPE values less than 50, with the lowest MAPE (17.81) being SARIMA(0,1,0)(3,0,2)$_{12}$.˝ The forecasting model was used, and the predicted value was within 95% of the confidence interval of the monthly DF cases predicted in 2019[23]. Our findings showed that the number of dengue cases predicted in 2020 had a 95% confidence interval (Figure 3),˝

Furthermore, when the number of cases was considered, the pattern of predicted DHF cases was intricately matched with the pattern of reported DHF cases. There were 3,272 DHF cases, with the highest and the lowest numbers in November (349 cases) and March (173 cases), respectively, of the forecast data, whereas the highest and the lowest numbers of the actual data were in January (157 cases) and December (6 cases), respectively. Based on the number of patients with dengue fever in the Nakhon Si Thammarat Province between 2010 and 2020[24], with a tendency to increase in the middle of the year until the end of the year before gradually declining, certain years could witness an increase until the beginning of the next year before gradually declining. This research was similar to that conducted by Abdulsalam *et al.* (2020), who investigated the impact of climate variables on the incidence of dengue fever[25-26].

In Thailand's Nakhon Si Thammarat Province, the number of patients increased in April, peaked in July, and gradually declined in October, whereas the number of cases was low from mid-February to mid-May[25]. A study on the relationship between climate variables and dengue incidence reported that the incidence of dengue infection spiked from April to October in 2021[26], implying that several factors influenced the pattern of DHF occurrence. An understanding of the incidence trend of DHF could assist the agencies involved in outbreak prevention preparedness by focusing on early disease prevention, such as eliminating mosquito breeding sites during periods of low disease activity. The information in this section is useful to relevant agencies because it can be used as guidelines for pre-prevention planning, such as mosquito breeding site elimination campaigns and avoiding mosquito bites. The ideal time would be at the start of the year because of the fewer chances of the disease to occur[27].

Furthermore, if the confidence interval obtained is considered, if the actual DHF cases in the following year are within the confidence range. That is, the severity of the dengue epidemic is moderate, and the spread of the disease is under control. If the monthly incidence exceeds the upper confidence interval value, the government or other relevant agencies should pay closer attention and identify causes to mitigate the further spread of the disease[28].

Forecasting using the SARIMA model could be inaccurate. Figure 3 shows the forecasting performance with MAPE = 848.734 and RMSE = 214.661. The value of the movement was insignificant. However, a considerable disparity is reported in the second half of the year, which could be ascribed to different factors, including an underestimation of dengue fever infection because people with only a few symptoms or no symptoms could not be tested for dengue fever, which impacts the reported number of patients, such that the number of cases could be four to six times higher than the reported data[27], or it could be due to the COVID-19 outbreak during the restricted period[29-30]. The epidemic has subsided, affecting the number of cases reported, as has the SARIMA model, which is suitable for short-term forecasting[31]. A Chinese study reported that the model SARIMA$(1,1,2)(0,1,1)_{12}$ was used to forecast the number of hand, foot, and mouth disease (HFMD) cases from September to December 2018,˝

The mean error rate (MER) and determination coefficient (R^2' values were, respectively, 16.86% and 94.27%[32]. A study in Brazil reported that 1-month dengue incidence forecasts were significantly more accurate than the 12-month dengue incidence forecasts (*p*-value = 0.002, Wilcoxon signed-ranks test).˝ Thus, forecasting should be repeated every 1 month, 2 months, 3 months, or according to the nature of the season with the updated data to improve the forecast's accuracy and efficiency of the applied model.[19]

The performance of the model depends on the data and the purpose of the forecast. The SARIMA model is one of the most efficient linear models to display seasonal trends with only one variable[31]. Therefore, it is possible to analyze the data using a statistical package[33]. A previous study demonstrate that the SARIMA model outperformed time-series regression and decomposition models, as well as the exponential smoothing method and the combined forecasting method[34-35]. Moreover, the dataset could be stationary data and should be sufficiently large to observe the data's trend[31]. Other models, such as error trend seasonality (ETS), neural network autoregression˝ (NNAR), and hybrid models, could be used for non-stationary data[31,33,36]. Other variables influencing or correlating with dengue cases that could improve the model's accuracy but are not used in this research analysis include environment, climate, water content, and socioeconomic status. Other models, such as multiple linear regression or Poisson's regression, should be used to incorporate additional variables into the analysis[25-27,37-38].

Our model was best suited for forecasting the number of dengue patients in the Nakhon Si Thammarat Province. Its application to other areas or diseases could depend on the characteristics of the data and related factors in each area. Therefore, future studies should focus on the associated dengue fever variables affecting the cases to improve the forecasting accuracy of the applied models. We believe the findings of the present study will be useful in developing and implementing effective dengue surveillance and prevention programs.

CONCLUSIONS

In this work, we used the SARIMA model to anticipate the occurrence of DHF in Thailand's Nakhon Si Thammarat area. We employed the RMSE and MAPE values to check the model's validity. We found that the SARIMA$(2,0,1)(1,0,0)_{12}$ model had the highest coefficient of determination (R^2) and the lowest forecasting error (MAPE = 848.734 and RMSE = 214.661), indicating its appropriateness for the data.

To lessen volatility and measure seasonality, we modified the data using natural logarithms. The projection projected seasonal variations for 2020 matched the actual data. Changes in the p, d, q values, and P, D, Q values of the seasonal ARIMA model in different studies could be attributed to environmental conditions, geographic locations, and the availability of healthcare services in these places. These findings demonstrate the applicability of statistical time series models in the development of public health initiatives and treatments by offering a better understanding of disease processes. A wide range of approaches and models could be used to assist the Ministry of Public Health, which is in charge of monitoring dengue sickness.

ACKNOWLEDGEMENTS

We are grateful to the Department of Disease Control and the Ministry of Public Health, Thailand, for providing the secondary data to construct the model.

REFERENCES

1. Martinez EZ, Silva EA. Predicting the number of cases of dengue infection in Ribeirão Preto, São Paulo State, Brazil, using a SARIMA model. Cad. Saúde Pública 2011; 27(9):1809-18.

2. Somboonsak P. Forecasting Dengue Fever Epidemics using ARIMA Model. In: AICCC2019. Proceedings of the 2019 2nd Artificial Intelligence and Cloud Computing Conference; 2019 December 21 – 23; Kobe, Japan. New York; Association for Computing Machinery;2020. P.144-150.

3. World Health Organization. Dengue and severe dengue [Internet]. Geneva: WHO; 2019 [Cited 2021 July 29]. Available from: http://www.who.int/news-room/fact-sheets/detail/dengue-and-severe-dengue.

4. World Health Organization. Improving data for dengue [Internet]. Geneva: WHO; 2019 [Cited 2021 July 29]. Available from: http://www.who.int/activities/improving-data-for-dengue.

5. Rojanapithayakorn W. Dengue hemorrhagic fever in Thailand. *Dengue Bulletin* 1998;22:60-72.

6. Chumpu R, Khamsemanan N, Nattee C. The association between dengue incidences and provincial-level weather variables in Thailand from 2001 to 2014. *PLoS One* 2019;14(12). doi: 10.1371/journal.pone.0226945.

7. Hinjumpa J, Charerntanyarak L. Changing epidemiology of dengue disease in 1996-2016, Bangkok, Thailand. *Disease Control Journal* 2017;43(4):342-355. doi: 10.14456/dcj.2017.1.

8. Bureau of Vector Borne Disease. Annual Report 2018 Bureau of Vector Borne Disease [Internet]. Bangkok: Department of Disease Control, Ministry of Public Health of Thailand; 2018 [Cited 2019 December 7]. Available from: https://ddc.moph.go.th/uploads/ckeditor/6f4922f45568161. a8cdf4ad2299f6d23/files/Report/Annual%20Report/2561.pdf.

9. Narayan N. Forecast Incidence of Dengue Fever Cases in Fiji Utilizing Autoregressive Integrated Moving Average (ARIMA) Model. *Int J Stat Appl* 2018;8(6):297-304. doi: 10.5923/j.statistics.20180806.02.

10. Phuthomdee S, Soontornpipit P, Viwatwongkasem C, et al. Dengue Forecasting Model using SARIMA Model to predict the Incidence of Dengue in Thailand. *Curr Appl Sci Technol* 2018; 18(2):58-65. doi: 10.14456/cast.2018.1.

11. Mekparyup J, Saithanu K. A Seasonal ARIMA Model For Forecasting The Dengue Hemorrhagic Fever Patients In Rayong, Thailand. *Global Journal of Pure and Applied Mathematics* 2015;11(2):175-181.

12. Wang H, Tian CW, Wang WM, et al. Time-series analysis of tuberculosis from 2005 to 2017 in China. *Epidemiol Infect* 2018;146(8):935–939. doi: 10.1017/S0950268818001115.

13. Cherdchoongam S, Rungreunganun V. Forecasting the Price of Natural Rubber in Thailand Using ARIMA Model. *KMUTNB international journal of applied science and technology* 2016;9(4):271-277. doi: 10.14416/j.ijast.2016.11.005.

14. Nayak M, Narayan KA. Forecasting Dengue Fever Incidence Using ARIMA Analysis. *Int J Collab Res Intern Med Public Health* 2019;11(6):924-932.

15. Siregar FA, Makmur T, Saprin S. Forecasting dengue hemorrhagic fever cases using ARIMA model: a case study in Asahan district. *IOP Conference Series: Materials Science and Engineering IOP Conf Ser Mater Sci Eng* 2018;300. doi: 10.1088/1757-899X/300/1/012032.

16. BERK RA. Encyclopedia of Sociology [Internet]. 2nd ed. USA: Macmillan; 2001 [Cited 2020 April 5]. Available from: https://go.gale.com/ps/retrieve.do?tabID=T003&resultListType=RESULT_LIST&searchResultsType=SingleTab&hitCount=555&searchType=BasicSearchForm&currentPosition=1&docId=GALE%7CCX3404400399&docType=Topic+overview&sort=Relevance&contentSegment=&prodId=GVRL&pageNum=1&contentSet=GALE%7CCX3404400399&searchId=R2&userGroupName=thnstru&inPS=true.

17. Pattranurakyothin T, Kumnungkit K. Forecasting Model for Para Rubber's Export

Sales. *KMITL Science and Tecnology* 2012;12(2):198-202.

18. Ebhuoma O, Gebreslasie M, Magubane L. A Seasonal Autoregressive Integrated Moving Average (SARIMA) forecasting model to predict monthly malaria cases in KwaZulu-Natal, South Africa. South Africa. *S Afr Med J* 2018;108(7):573-578. doi: 10.7196/SAMJ.2018.v108i7.12885.

19. Luz PM, Mendes BV, Codeço CT, et al. Time series analysis of dengue incidence in Rio de Janeiro, Brazil. *Am J Trop Med Hyg* 2008;79(6):933-9. doi: 10.4269/ajtmh.2008.79.933.

20. Matinez EZ, Silva EA, Fabbro AD. A SARIMA forecasting model to predict the number of cases of dengue in Campinas, State of São Paulo State, Brazil. *Rev Soc Bras Med Trop* 2011;44(4):436-440. doi: 10.1590/S0037-86822011000400007.

21. Singmuang C, Supapakorn T. Forecasting Model for the Export Value of Rubber of Thailand. *Science and Technology Nakhon Sawan Rajabhat University Journal* 2020; 12(15):58-82.

22. Qiu H, Zhao H, Xiang H, et al. Forecasting the incidence of mumps in Chongqing based on a SARIMA model. *BMC Public Health* 2021;21:373. doi: 10.1186/s12889-021-10383-x.

23. Hadi ZA, Dom NC, Ismail R. A Seasonal autoregressive integrated moving average model to predict incidence of dengue cases in Kuantan, Malaysia. *Science, Engineering and Health Studies* 2021;15:21020009. doi: 10.14456/sehs.2021.36.

24. Ministry of Public Health, Department of Disease Control. HDC-Report [Internet]. Thailand: [Cited 2020 January 5]. Available from: https://hdcservice.moph.go.th.

25. Abdulsalam FI, Antunez P, Yimthiang S, et al. Influence of climate variables on dengue fever occurrence in the southern region of Thailand. *PLOS Glob Public Health* 2022;2(4): e0000188.26. doi: 10.1371/journal.pgph.0000188.

26. Abdulsalam FI, Yimthiang S, La-Up A, et al. Association between climate variables and dengue incidence in Nakhon Si Thammarat Province, Thailand. *Geospat Health* 2021;6(2). doi: 10.4081/gh.2021.1012.

27. Polwiang S. The time series seasonal patterns of dengue fever and associated weather variables in Bangkok (2003-2017). *BMC Infect Dis* 2020;208. doi: 10.1186/s12879-020-4902-6.

28. Chen h, Wang X, Zhao J, et al. Application of the ARIMA Model in Forecasting the Incidence of Tuberculosis in Anhui During COVID-19 Pandemic from 2021 to 2022. Infection and Drug Resistance 2022;15:3503-3512. doi: 10.2147/IDR.S367528.

29. Surendran SN, Nagulan R, Sivabalakrishnan K, et al. Reduced dengue incidence during the COVID-19 movement restrictions in Sri Lanka from March 2020 to April 2021. *BMC Public Health* 2022; 388. doi: 10.1186/s12889-022-12726-8.

30. JiangI L, Liu Y, SuID W, et al. Decreased dengue cases attributable to the effect of COVID-19 in Guangzhou in 2020. *PloS Negl Trop Dis* 2021;15(5):e0009441. doi: 10.1371/journal.pntd.0009441.

31. Punyapornwithaya V, Mishra P, Sansamur C, et al. Time-Series Analysis for the Number of Foot and Mouth Disease Outbreak Episodes in Cattle Farms in Thailand Using Data from 2010-2020. *Viruses* 2022;14(7):1367. doi: 10.3390/v14071367.

32. Tian CW, Wang H, Luo XM. Time-series modelling and forecasting of hand, foot and mouth disease cases in China from 2008 to 2018. *Epidemiol Infect* 2019;147:e82. doi: 10.1017/S095026881800362X.

33. Punyapornwithaya V, Jampachaisri K, Klaharn K, et al. Forecasting of Milk Production in Northern Thailand Using Seasonal Autoregressive Integrated Moving Average, Error Trend Seasonality, and Hybrid Models. *Front Vet Sci* 2021;8:775114. doi: 10.3389/fvets.2021.775114.

34. Weisent J, Seaver W, Odoi A, et al. Comparison of three timeseries models for predicting campylobacteriosis risk. *Epidemiol Infect* 2010;138:898–906. doi: 10.1017/S0950268810000154.

35. Riansut W. ELECTION OF FORECASTING MODELS FOR THE MANGO PRICES. *Srinakharinwirot Research and Development (Journal of Humanities and Social Sciences)* 2019;12(23):52-63.

36. Wang Y, Xu C, Zhang S, et al. Temporal trends analysis of tuberculosis morbidity in mainland China from 1997 to 2025 using a new SARIMA-NARNNX hybrid model. BMJ Open 2019;9:e024409. doi: 10.1136/bmjopen-2018-024409.

37. Abolmaali S, Shirzaei S. A comparative study of SIR Model, Linear Regression, Logistic Function and ARIMA Model for forecasting COVID-19 cases. *AIMS Public Health* 2012;8(4):598–613. doi: 10.3934/publichealth.2021048.

38. Wongkoon S, Jaroensutasinee M, Jaroensutasinee K. Weather factors influencing the occurrence of dengue fever in Nakhon Si Thammarat, Thailand. *Trop Biomed* 2013;30(4):631–641. PMID: 24522133.