



Article Model Forecasting Development for Dengue Fever Incidence in Surabaya City Using Time Series Analysis

Mahmod Othman ¹, Rachmah Indawati ²,*, Ahmad Abubakar Suleiman ^{1,3}, Mochammad Bagus Qomaruddin ² and Rajalingam Sokkalingam ¹

- ¹ Department of Fundamental and Applied Sciences, Universiti Teknologi Petronas, Seri Iskandar 32610, Perak Darul Ridzuan, Malaysia
- ² Faculty of Public Health, Universitas Airlangga, Surabaya 60115, Indonesia
- ³ Department of Statistics, Kano University of Science and Technology, Wudil 713281, Nigeria
- * Correspondence: rachmah.indawati@fkm.unair.ac.id

Abstract: Dengue hemorrhagic fever (DHF) is one of the most widespread and deadly diseases in several parts of Indonesia. An accurate forecast-based model is required to reduce the incidence rate of this disease. Time-series methods such as autoregressive integrated moving average (ARIMA) models are used in epidemiology as statistical tools to study and forecast DHF and other infectious diseases. The present study attempted to forecast the monthly confirmed DHF cases via a time-series approach. The ARIMA, seasonal ARIMA (SARIMA), and long short-term memory (LSTM) models were compared to select the most accurate forecasting method for the deadly disease. The data were obtained from the Surabaya Health Office covering January 2014 to December 2016. The data were partitioned into the training and testing sets. The best forecasting model was selected based on the lowest values of accuracy metrics such as the root mean square error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE). The findings demonstrated that the SARIMA (2,1,1) (1,0,0) model was able to forecast the DHF outbreaks in Surabaya City compared to the ARIMA (2,1,1) and LSTM models. We further forecasted the DHF cases for 12 month horizons starting from January 2017 to December 2017 using the SARIMA (2,1,1) (1,0,0), ARIMA (2,1,1), and LSTM models. The results revealed that the SARIMA (2,1,1) (1,0,0) model outperformed the ARIMA (2,1,1) and LSTM models based on the goodness-of-fit measure. The results showed significant seasonal outbreaks of DHF, particularly from March to September. The highest cases observed in May suggested a significant seasonal correlation between DHF and air temperature. This research is the first attempt to analyze the time-series model for DHF cases in Surabaya City and forecast future outbreaks. The findings could help policymakers and public health specialists develop efficient public health strategies to detect and control the disease, especially in the early phases of outbreaks.

Keywords: ARIMA model; SARIMA model; forecasting; dengue fever; time-series analysis

1. Introduction

Dengue is a mosquito-borne disease caused by the dengue virus, affecting most tropical regions worldwide [1]. According to [2], dengue infection has been defined as dengue without warning signs, dengue with warning signs, and severe dengue. Therefore, dengue fever is more likely to be dengue without warning signs, however, a few cases of dengue fever may present warning signs. DHF will include dengue with warning signs, especially plasma leakage and severe dengue. It is estimated that between 50 and 500 million people worldwide are infected with dengue each year [3,4]. Between 10,000 and 20,000 people die each year, and about two and a half billion people are in danger of infection [5–7]. Recent figures have shown that 60% of the world's population would be susceptible to dengue disease by 2080 [8]. According to this estimation, 10,000 people have died from dengue in more than 125 countries worldwide. Even though dengue deaths are 99% preventable, case fatality rates significantly greater than 1% have been recorded globally [9].



Citation: Othman, M.; Indawati, R.; Suleiman, A.A.; Qomaruddin, M.B.; Sokkalingam, R. Model Forecasting Development for Dengue Fever Incidence in Surabaya City Using Time Series Analysis. *Processes* 2022, 10, 2454. https://doi.org/10.3390/ pr10112454

Academic Editor: Philippe Bogaerts

Received: 11 October 2022 Accepted: 17 November 2022 Published: 19 November 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/).

Dengue fever is one of the most severe and common health problems in Indonesia. Since 1968, the number of cases and the transmission of dengue fever have been rising [10]. The growing population, rapid urbanization, and modern transportation have significantly contributed to the spread of the disease. Indonesia is a tropical country with a high population density, especially in urban areas, which could serve as a habitat for dengue viruses. Dengue viruses are transmitted through the bite of the Aedes aegypti and Ae. Albopictus causes a high fever, red spots on the skin, and pain in the muscles. There is no vaccine available for preventing DHF. The current disease prevention plan is not effective as the only available treatments for DHF patients are supportive symptomatic care such as antipyretics, antiemetics, and IV fluids [11]. According to information from the Surabaya Health Office, Surabaya City is a dengue-endemic area. In 2010, there were 3379 cases of dengue, with an incidence rate (IR) of 116.03 per 100,000 people and a case fatality rate (CFR) of 0.8%. In 2011, there were 1008 dengue cases with 36.22 IR and 0.3% CFR. In 2012, there were 1091 dengue cases with 38.60 IR and 0.55% CFR. In 2013, there were 2207 dengue cases with 78.35 IR and 0.86% CFR. The government has implemented a variety of initiatives through a variety of programs to prevent the rise in incidence including preventive and promotive efforts. Preventive efforts can be described by instilling clean living habits (such as not littering, hoarding junks, and not allowing any containers to be the breeding ground for larvae). This activity is known as mosquito nest eradication (MNE). This approach, however, cannot appropriately recognize changes in prevalence [12].

A forecast-based early warning system is required to reduce the incidence rate of this disease. There are several other forecasting models in use globally and even in Indonesia. However, these models are not efficient enough to accommodate all the characteristics of the DHF data. Hence, there is a dire need to develop more flexible forecasting models that can provide better results than the existing ones. This will assist policymakers to make better decisions. Time-series forecasting is a statistical method that has been used extensively in numerous fields, particularly in the study of infectious disease epidemiology [13,14]. Several studies have used statistical models developed with the aim of forecasting dengue in various settings [15–23]. Due to the time-varying behavior, seasonal pattern, secular trend, and rapid fluctuations in time-series exhibited by DHF data, it is feasible to forecast the incidence of DHF with time-series methods to enable an early response to the disease [24]. Autoregressive integrated moving average (ARIMA) is a popular time-series forecasting technique in health science research [25]. It enables us to identify hidden behaviors in the data. However, the ARIMA model is inappropriate for time-series data containing seasonality [10]. The ARIMA model is also a tedious method requiring computational skill, despite producing efficient results [7]. Consequently, the ARIMA method is modified for seasonal data and is known as SARIMA. The SARIMA model combines seasonal and non-seasonal autoregressive and moving average models [3,5]. The SARIMA is a method for identifying the patterns from seasonal time-series data for forecasting future values of the time-series and has received the most attention in recent years [26,27]. DHF data exhibit both linear and nonlinear behavior [28]. However, the SARIMA model is only used for modeling linear time-series data and cannot handle nonlinear behavior. A modern method of deep learning algorithms has been developed for prediction applications. This method can handle nonlinearity and complexity in time-series forecasting. LSTM is one deep learning method that allows for the processing of longer temporal sequences.

Time-series analysis can be used to examine past trends of DHF outbreaks and improve the present prevention and control measures. This approach is one of the most accurate statistical models that can be developed using the data to predict future DHF epidemics. The time-series forecasting approach has not previously been considered to forecast DHF prevalence in Surabaya City. Hence, this research aimed to propose time-series models that can forecast future values of DHF outbreaks in Surabaya City.

2. Materials and Methods

In this section, we highlighted the study area where the data on DHF cases were collected. Next, the concepts and main definitions of the ARIMA, SARIMA, and LSTM models are introduced. In addition, the non-stationary models test is presented and followed by the algorithm of the proposed method.

2.1. Study Area

Surabaya is the capital city of the Indonesian province of East Java and the secondlargest city in Indonesia after Jakarta, located at 7°14′45″ S, 112°44′16″ E, and covers an area of 911/km². It has a total population of 2,874,314 (2020) and a density of 7134/km². Surabaya is an endemic area for DHF cases with the highest incidence of dengue in the country. Figure 1 shows a map of the study area.



Figure 1. A map of the study area showing Surabaya City, East Java, Indonesia.

2.2. Data Collection

This study used monthly DHF cases from January 2014 to December 2016 obtained from the Surabaya Health Office. The monthly dataset contains a total of 36 monthly observations. We used the DHF cases covering 2014–2016 due to the availability of these datasets recorded from the Surabaya Health Office at the time of this project and their sufficient size for model testing to validate this research. The data were processed, coded, and entered into R-studio version 4.2.1 and evaluated for normality. The R software was created by Ross Ihaka and Robert Gentleman at the University of Auckland, New Zealand, and is currently developed by the R Development Core Team. The data were partitioned into training and testing sets. The training set was applied for the model development, while the testing set was applied for the validation of the developed model. The best model was selected and used for out-of-sample forecasting.

2.3. ARIMA Model

The ARIMA model is suitable for modeling stationary time-series data, although most time-series data from real-life phenomena exhibit non-stationary patterns. However, the model assumes that a non-stationary series could become stationary through a differencing approach [29–31]. The generic version of the ARIMA model is given in Equation (1):

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

This model is denoted by

$$ARIMA(p,d,q) \tag{2}$$

where Y_t is a dependent variable on t that might have been differenced once or more; p is the order of the autoregressive part; d is the degree of differencing; q is the order of the moving average part; ε_t is the error term; C is the constant number of the model.

2.4. SARIMA Model

SARIMA means the seasonal ARIMA. In time-series analysis, seasonality is a regular pattern of variations that repeats over *s* periods, where *s* denotes the number of periods until the pattern repeats. For instance, there is seasonality in the monthly data for which high values tend to occur in some months, while low values tend to occur in other months. The general form of the SARIMA model is given in Equation (3):

$$\Phi_P(B^s)\varphi(B)\nabla^D_s\nabla^d x_t = \Theta_O(B^s)\theta(B)w_t \tag{3}$$

This model can be denoted by

SARIMA
$$(p, d, q)(P, D, Q)_s$$
 (4)

where x_t is the nonstationary time-series; w_t is the usual Gaussian white noise process; and s is the period of the time-series. The autoregressive and moving average components are represented by polynomials Φ_P and Θ_Q of orders p and q. The seasonal autoregressive and moving average components are $\Phi_P(B^s)$ and $\Theta_Q(B^s)$, where P and Q are their orders. ∇^d and ∇_s^D are the ordinary and seasonal difference components. B is the backshift operator. The expressions are shown as follows:

$$\begin{split} \varphi(B) &= 1 - \varphi_1 B - \varphi_2 B^2 - \ldots - \varphi_p B^p \\ \Phi_P(B^s) &= 1 - \Phi_1 B^s - \Phi_2 B^{2s} - \ldots - \Phi_P B^{Ps} \\ \theta(B) &= 1 + \theta_1 B + \theta_2 B^2 + \ldots + \theta_q B^q \\ \Theta_Q(B^s) &= 1 + \Theta_1 B^s + \Theta_2 B^{2s} + \Theta_Q B^{Qs} \\ \nabla^d &= (1 - B)^d \\ \nabla^D_s &= (1 - B^s)^D \\ B^k x_t &= x_{t-1} \end{split}$$

This study focused on the monthly confirmed DHF cases. If the seasonal period of the series s = 12, then we can rewrite Equation (3) as:

$$\Phi_P(B^{12})\varphi(B)\nabla^D_{12}\nabla^d x_t = \Theta_Q(B^{12})\theta(B)w_t$$
(5)

2.5. LSTM Model

One of the advancements in neural networks that can learn long-term dependence is known as LSTM. The architecture of an LSTM is composed of three gates: forget gates, input gates, and output gates [32].

The forget gate G_t determines the specific information that is deleted from the memory cells (cell state). Forget gates use a sigmoid activation function as their activation function, where the result is between 0 and 1. If the output is 1, all the information will be retained. If it is 0, all the information will be discarded. It is given by

$$G_t = \sigma \left(W_g P_{t-1} + W_g X_t \right) \tag{6}$$

where W_g is the forget gate weight; P_{t-1} represents the previous state or state at time t - 1; X_t denotes the input at time t; and σ denotes the sigmoid activation function.

The input gate V_t is responsible for determining the information that is added to the cell state (s_t). This process is broken down into two distinct processes. In the first step of

the process, the candidate value Ψ , which can be added to the cell states, is calculated. The activation values V_t of the input and gates are determined in the second step in the process, where W_v denotes the input gate weight and W_s denotes the cell state weight. The two processes are, respectively, given by

$$V_t = \sigma(W_v P_{t-1} + W_v X_t) \tag{7}$$

$$\psi = tanh(W_s P_{t-1} + W_s X_t) \tag{8}$$

The new cell states S_{t-1} are determined based on the outcomes of the previous processes. The formula can be broken down as follows:

$$s_t = (G_t * s_{t-1} + v_t * \psi)$$
(9)

After the memory cell has passed the input gate and the forget gate, the output gate m_t will generate output (n_t) . Two gates will be implemented at the output gates. The first gates will use a sigmoid layer to determine which parts of the cell state will be output. Utilizing the *tanh* activation function will result in the storage of a value within the memory cell. Finally, the two gates are multiplied together to produce a value that will be distributed (n_t) . Applying the following formula, we have

$$m_t = \sigma(W_m P_{t-1} + W_m X_t) \tag{10}$$

$$n_t = m_t * tanh(s_t) \tag{11}$$

2.6. Non-Stationary Test

The statistical check was performed using the Augmented Dickey–Fuller (ADF) test and the Kwiatkowski–Phillips-Perron Unit Root (KPSS) test to assess stationarity in both the original dataset and the differenced time-series. The null hypothesis of both ADF and KPSS assumed that the time-series was non-stationary. In the ADF and KPSS tests, if the *p*-value was less than 5% or there was a 0.05 level of significance for a time-series, then we rejected the null hypothesis and inferred that the series was assumed to be stationary.

The proposed model consists of five steps, as shown in Figure 2. The first step is data processing, the second step is model identification, the third step is model development, the fourth step is evaluating models, and the last step is forecasting.



Figure 2. Framework for the model development and forecasting.

3. The Proposed Model

This section discusses the descriptive statistics of the DHF cases, model identification, stationarity testing as well as training and testing models.

3.1. Descriptive Statistics

Table 1 displays the DHF reported cases from 2014 to 2016. The maximum (164) number of cases was reported in April 2016, while the minimum (4) number of cases were obtained each in November 2015 and 2016, respectively. Table 2 provides a descriptive summary of the Surabaya City dengue time data. From Table 2, there was a rise in the incidence in 2016. Table 2 also shows that the dataset was positively and negatively skewed, indicating non-normality. Graphical inspection of the disease indicated time-varying patterns in Figure 3. The series tended to rise and fall gradually, with no noticeable outliers. Higher numbers showed a trend, especially in April, when the rate was at its maximum in 2016. Regular fluctuations and the time-varying behavior of the trend revealed that the ARIMA models were appropriate [7].

Month	2016	2015	2014	Total	Mean
January	60	46	36	142	12
February	114	109	46	269	23
March	134	107	71	312	26
April	164	119	94	377	33
May	141	95	127	363	31
June	119	78	110	307	26
July	89	40	82	211	18
August	66	21	71	158	14
September	24	9	71	104	9
Ôctober	18	5	42	65	6
November	4	4	37	45	4
December	5	7	29	41	4
Grand Total	640	640	816	2394	200

Table 1. Average of DHF cases in Surabaya City from 2014 to 2016.

Table 2. Descriptive statistics of the annual confirmed DHF cases in Surabaya from 2014 to 2016.

Years	Min.	Max.	Q1	Q2	Q3	Mean	S.D.	Skewness	Kurtosis
2016	4.00	164.00	22.50	77.50	122.75	78.17	56.71914	-0.00655	-1.6732
2015	4.00	119.00	8.50	43.00	98.00	53.33	45.52189	0.213275	-1.8190
2014	29.00	127.00	40.75	71.00	85.00	68.00	31.34848	0.4051189	-1.2137



Figure 3. Plot of the monthly DHF case data from 2014–2016, showing maximum peaks from April to May each year.

3.2. Model Identification

To identify a suitable ARIMA model for time-series forecasting, it must be free of trend and seasonality. A time-series may be influenced by the trend and seasonality components at distinct periods [33]. The time-series data must be stationary to develop a model that is effective in forecasting future values. Several tests can be used to determine whether a series is stationary [34]. These tests include the Augmented Dickey–Fuller unit root test (ADF test), the Kwiatkowski–Phillips–Schmidt–Shin (KPSS) test, the partial correlation function (PACF), and the autocorrelation function (ACF) [32]. A time-series is assumed to be stationary in the ADF test if the *p*-value is less than a 5% level of significance. On the other hand, if the time-series is non-stationary, then it is necessary to look at the time-series graph and differentiate the data appropriately [35]. The time-series had a seasonality effect, as shown by the repeating cycles in Figure 4. The ACF and PACF of the original dataset are shown in Figure 5. Observing the ACF in Figure 5, it is reasonable to see that the worst outbreaks occurred every six months, which indicates a seasonal pattern. This suggests that the incidence of DHF in Surabaya City was strongly affected by the seasons. Using a differencing approach, non-stationarity in the time-series data can be corrected [36]. A similar seasonal effect was seen in the original dataset after the first differencing, necessitating the second differencing. The ACF and PACF plots of second-order differenced time-series are shown in Figure 6. It can be seen that the ACF decreased to zero exponentially, indicating stationary behavior [37]. Therefore, the SARIMA model could be used to fit the deseasonalized data [36].

Decomposing the time-series data assists in discovering various hidden behavior within the time-series. In general, a time-series consists of four components: random, seasonal, trend, and cyclic. Three of these components are indicated in Figure 7. It can be observed that the seasonal component exhibited seasonal fluctuations in the dataset. However, the trend and random components seemed to be stationary.



Figure 4. Monthly collected DHF cases in Surabaya City, 2014–2016, showing the time-series pertaining to the time-varying patterns and seasonality across the study period.



Figure 5. (a) Autocorrelation function. (b) Partial autocorrelation for the original time-series of monthly confirmed DHF cases in Surabaya City.



Figure 6. (a) Autocorrelation function. (b) Partial autocorrelation for second-ordered differencing time-series of monthly confirmed DHF cases in Surabaya City.



Figure 7. Decomposition of Surabaya City monthly DHF cases time-series data.

3.3. Non-Stationary Test

Table 3 presents the results of the ADF for both the original and second-order differenced datasets. It can be observed that the *p*-value for the original dataset was greater than 0.05, and we can say that we failed to reject the null hypothesis and concluded that the original dataset was nonstationary. In contrast, the *p*-value for the second differenced timeseries was less than 0.05, so we can say that the time-series is stationary. Similarly, it can be observed in Table 4 that the *p*-value obtained from the KPSS test for the second differenced dataset was less than 0.05, which means that the time-series is stationary. Interestingly, the time-series data under consideration is non-stationary in its original form and becomes stationary when it is second differenced. The level of significance of the ADF and KPSS tests for the second differenced series were 5% and 1%, respectively. Therefore, the data can be suitable for time-series analysis.

Dataset	Critical Value	<i>p</i> -Value	
Original	-3.2892	0.08936	
Second-order differencing	-3.7968	0.03287	

Table 3. Augmented Dickey–Fuller test for the original and second differenced of the Surabaya City monthly DHF case time-series data.

Table 4. Kwiatkowski–Phillips–Perron unit root test for the original and second differenced of the Surabaya City monthly DHF case time-series data.

Dataset	Critical Value	<i>p</i> -Value
Original	-12.042	0.3605
Second-order differencing	-43.200	0.0100

3.4. Training and Test Models

It is important to determine that the models are adequate to forecast l future values with high accuracy. Therefore, our first step is to split the sample into training and test sets. We selected the data points from January 2014 to December 2015 for model training, consisting of 24 data points. The data points from January 2016 to December 2016 were for model testing.

Fitting the ARIMA model to the training data point will enable the model to learn from the time-series dataset. We used an ARIMA diagnosis plot on the training dataset to determine what lag to use in the model. We examined the properties of the data using residuals plots of the ACF and PACF. The statistical criteria used to evaluate these models are the root mean square error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE), which are expressed in Equations (12)–(14), respectively. More details on these measures can be found in [38,39].

$$RMSE = \sqrt{\frac{\sum\limits_{i=1}^{N} (Y_i - F_i)^2}{N}}$$
(12)

$$MAE = \frac{\sum_{i=1}^{N} |Y_i - F_i|}{N}$$
(13)

$$MAPE = \frac{1}{N} \sum_{i=1}^{N} \left| \frac{Y_i - F_i}{Y_i} \right| \times 100\%$$
(14)

where Y_i is the actual data; F_i is the forecasted values; and N is the total number of observations in the data.

It can be observed from Figure 8 that the ACF geometrically declined, indicating a signal of the error process. The PACF in the plot had two significant lags. As a result, we developed an ARIMA (2,1,1) model, which means that we included two AR lags, one difference, and one MA lag. Thus, ARIMA (2,1,1) passed this test. Next, we tested the residuals of the model. The residuals plot in Figure 9 indicates that there was one lag at the 12 periods that was significant. This suggests that there is seasonality in the model.



Figure 8. The ARIMA (2,1,1) diagnosis plot in the training dataset.



Figure 9. The residual plots of ARIMA (2,1,1) on the training dataset.

4. Results and Discussion

In this section, we discuss the results obtained from the model development. The results of the ARIMA (2,1,1), SARIMA (2,1,1) (1,0,0), and LSTM models were compared. The out-of-sample forecast for the twelve month horizons were obtained based on the three competitive models.

After identifying and evaluating each parameter, the ARIMA (2,1,1), SARIMA (2,1,1) (1,0,0), and LSTM models can be used to predict future values for both the training and test datasets. We found that the SARIMA (2,1,1) (1,0,0) model outperformed the ARIMA (2,1,1) and LSTM models in terms of the accuracy metrics reported in Tables 5 and 6, respectively. Figures 10–12 depict the predicted dengue outbreaks based on the training datasets. From these figures, we can note that the SERIMA (2,1,1) (1,0,0) model predicts the dengue outbreak values more accurately than the ARIMA (2,1,1) and LSTM models.

Model.	RMSE	MAE	MAPE
ARIMA (2,1,1)	19.66198	15.10929	60.09833
SARIMA (2,1,1) (1.0.0)	13.07250	10.27440	53.47647
LSTM	15.35412	12.54872	56.67457

Table 5. Comparison of ARIMA (2,1,1), SARIMA (2,1,1) (1,0,0), and LSTM models on the training dataset using accuracy metrics.

Table 6. Comparison of the ARIMA (2,1,1), SARIMA (2,1,1) (1,0,0), and LSTM models on the testingdataset using accuracy metrics.

Model	RMSE	MAE	MAPE
ARIMA (2,1,1)	77.42119	71.02259	104.95288
SARIMA (2,1,1) (1,0,0) LSTM	38.00313	33.54329	54.565850
	45.37822	41.45632	76.983421



Figure 10. Plot of the actual, fitted, and forecasted values using ARIMA (2,1,1) on the training dataset.



Figure 11. Plot of the actual, fitted, and forecasted values using SARIMA (2,1,1) (1,0,0) on the training dataset.



Figure 12. Plot of the actual, fitted, and forecasted values using SARIMA (2,1,1) (1,0,0) on the training dataset.

The results of the training set indicate that the SARIMA (2,1,1) (1,0,0) is more appropriate to perform the out-of-sample forecasting of dengue outbreaks in Surabaya City in the next 12 months.

Out-of-Sample Forecasting

Finally, a 12-month out-of-sample forecasting of dengue cases was conducted from January 2017 to December 2017. According to the findings, SARIMA (2,1,1) (1,0,0) performed well in forecasting future outbreaks. Figure 13 shows that this model is stable since all the data points lie within the unit root circle. The findings of the monthly future values of the outbreaks from January 2017 to December 2017 are presented in Table 7. From this table, the SARIMA (2,1,1) (1,0,0) model revealed the smallest value of the RMSE metric in comparison to the ARIMA (2,1,1) and LSTM models. Thus, the SARIMA (2,1,1) (1,0,0) model can be selected as the best forecasting model for the monthly DHF cases in Surabaya City.

From Table 7, the forecasted results showed that the number of monthly dengue outbreaks began to increase in March and continued through September, with the maximum incidence occurring in May. This time-dynamic pattern suggested that there was a significant seasonal effect in the forecast. It is noteworthy that the highest forecasted incidence in May was like the outbreaks that occurred in 2014, which had the highest incidence in May. Figure 14 confirms the findings of Table 7 to show the trend pattern in the number of DHF cases across the forecasting horizons. Therefore, this forecast using the SARIMA (2,1,1) (1,0,0) model provides the future values of dengue outbreaks and could be used to assist authorities and public health professionals in designing effective public health measures to prevent and control the disease, particularly during the early stages of the outbreaks.

Table 7. The out-of-sample forecasted value for monthly DHF outbreaks in Surabaya City from January 2017 to December 2017 using ARIMA (2,1,1), SARIMA (2,1,1) (1,0,0), and LSTM models.

Model	Month	January	February	March	April	May	June	July	August	September	October	November	December	RMSE
ARIMA	Forecast	6	9	26	18	37	33	25	16	13	8	6	4	31.21
SARIMA	Forecast	47	46	63	71	91	80	67	59	61	44	43	38	11.35
LSTM	Forecast	13	14	46	101	99	109	89	74	41	26	16	13	20.13



Figure 13. Unit root circle plot of the out-of-sample forecasting analysis.



Figure 14. Out-of-sample forecasted values of the monthly dengue outbreak in Surabaya City from January 2017 to December 2017 using the SARIMA (2,1,1) (1,0,0) model.

DHF is a serious public health issue in Surabaya. This is mostly due to an unprecedented rise in the number of cases recorded each year. Consequently, only a few preventative measures have been implemented to prevent outbreaks. Instead of responding in advance to safeguard against disease, the health sector responds after it emerges. The present research attempted to forecast the DHF incidence using an effective forecasting method to assist the public and authorities in properly adjusting to outbreak and provide enough preparations in creating public awareness. The objective was to forecast the DHF outbreaks using a time-series model from 2014 to 2016 for twelve months covering January 2017 to December 2017 using time-series analysis consisting of monthly confirmed DHF cases from 2014 to 2016. The series showed that the disease had a strong seasonal effect with the maximum rates occurring between March and September. The forecast showed peaks of incidence in May with the highest incidence forecasted to be 91 cases in May 2017. Similarly, researchers from Rajasthan designed a forecasting model for DHF using timeseries data from the past decade to forecast monthly dengue fever/dengue hemorrhagic incidence for 2011. The SARIMA model was employed for statistical modeling. Dengue fever/dengue hemorrhagic cases that were reported between January 2001 and December 2010 displayed a cyclical pattern with seasonal variation. The forecast for 2011 indicated a seasonal peak in October with a predicted 546 cases [12]. Another study conducted in Ribeirao Preto, Sao Paulo State, Brazil also applied SARIMA to fit in a model of monthly reported cases of DHF from 2000 to 2008. The forecasted values for the incidence for 2009 were obtained and compared with the results of the original number of cases. The researchers found that the SARIMA model effectively forecast the number of DHF cases and is a reliable approach for disease control and prevention [40]. In a study recently conducted in Jeddah, Saudi Arabia, the SARIMA model was used to forecast DHF mortality and morbidity using time-series from 2006 to 2016 for the years 2017 to 2019. According to the study, incidence rates increased from May to September, having the highest rate in 2012, suggesting a strong seasonality [7].

The prevalence of DHF has been associated with climate variations. High temperatures promote mosquito reproduction, while an increase in rainfall contributes to the availability of vector habitat [41,42]. DHF epidemics often occur seasonally and tend to spike in the summer and spring when the weather is hot and humid. However, the association between incidence and climate remains poorly understood and typically varies among locations due to local climate heterogeneity and virus–host interactions, which are all factors in the spread of the disease. Globally, every year, 50–100 million humans are infected by a female Aedes aegypti mosquito that has fed on infected human blood. In many tropical and subtropical countries, dengue disease is seasonal. This is because rainfall provides breeding sites and stimulates egg hatching, and temperature affects the mosquito's survival, development, and reproduction. Temperature enhances the mosquitoes' capacity to spread the dengue virus; higher temperatures boost transmission rates.

5. Conclusions

Dengue is the fastest-spreading vector-borne disease in the world. Monthly confirmed DHF cases in Surabaya City were obtained from 2014 to 2016 for this study to forecast disease outbreaks in the early phases and enable quick response. To develop effective forecasting models, monthly DHF occurrence patterns were studied. The prevalence of DHF in 2017 was then forecast using the best models. The results of DHF incidence revealed a significant seasonal effect. These findings showed an increase in DHF cases from March to September. Additionally, the dengue outbreaks appeared to spike more in May every three years, in 2014 and 2017. The air temperature is suspected to be the significant factor associated with DHF cases in Surabaya City.

6. Recommendations

To help with health care planning, public health officials want a means to forecast when epidemics will occur. To develop such a system, they need to understand the factors that lead to epidemics. The presence of mosquito larvae is the most important risk factor for dengue fever in several parts of Surabaya City. Hence, it is necessary for the government to exercise early efforts to eradicate and minimize DHF cases by conducting dengue surveillance, which aims to monitor the trends. In future studies, we will incorporate correlation studies of DHF cases with meteorological data, explore more current data for DHF cases, and consider several models for dengue incidence forecasting. **Author Contributions:** Conceptualization, M.O., R.I. and M.B.Q.; Methodology, R.I., M.O., R.S., M.B.Q., A.A.S. and R.S.; Software, M.O. and A.A.S.; Validation, M.O., R.I. and A.A.S.; Formal analysis, M.O., R.I., A.A.S. and R.S.; Writing—original draft preparation, M.O., A.A.S., R.I., M.B.Q. and R.S.; Discussion about behavior data, M.O., R.I. and M.B.Q.; Writing—and editing, M.O., R.I., A.A.S. and R.S. All authors have read and agreed to the published version of the manuscript.

Funding: This project is supported by Universitas Airlangga.

Data Availability Statement: Not applicable.

Acknowledgments: The authors would like to thank the Universiti Teknologi Petronas, Universiti Airlangga, and the Surabaya Health Office, who provided support to this project.

Conflicts of Interest: The authors declare no conflict of interest.

References

- 1. World Health Organization. WHO Guidelines Approved by the Guidelines Review Committee. In *Dengue: Guidelines for Diagnosis, Treatment, Prevention and Control: New Edition;* World Health Organization Copyright © 2009; World Health Organization: Geneva, Switzerland, 2009.
- 2. WHO. Dengue: Guidelines for Diagnosis, Treatment, Prevention and Control; WHO Library: Geneva, Switzerland, 2009; pp. 10–12.
- 3. Guha-Sapir, D.; Schimmer, B. Dengue fever: New paradigms for a changing epidemiology. Emerg. Epidemiol. 2005, 2, 1. [CrossRef]
- Sutriyawan, A.; Herdianti, H.; Cakranegara, P.A.; Lolan, Y.P.; Sinaga, Y. Predictive Index Using Receiver Operating Characteristic and Trend Analysis of Dengue Hemorrhagic Fever Incidence. Open Access Maced. J. Med. Sci. 2022, 10, 681–687. [CrossRef]
- 5. Bhatt, S.; Gething, P.W.; Brady, O.J.; Messina, J.P.; Farlow, A.W.; Moyes, C.L.; Drake, J.M.; Brownstein, J.S.; Hoen, A.G.; Sankoh, O. The global distribution and burden of dengue. *Nature* **2013**, *496*, 504–507. [CrossRef]
- Stanaway, J.D.; Shepard, D.S.; Undurraga, E.A.; Halasa, Y.A.; Coffeng, L.E.; Brady, O.J.; Hay, S.I.; Bedi, N.; Bensenor, I.M.; Castañeda-Orjuela, C.A. The global burden of dengue: An analysis from the Global Burden of Disease Study 2013. *Lancet Infect. Dis.* 2016, 16, 712–723. [CrossRef]
- 7. Abualamah, W.A.; Akbar, N.A.; Banni, H.S.; Bafail, M.A. Forecasting the morbidity and mortality of dengue fever in KSA: A time series analysis (2006–2016). *J. Taibah Univ. Med. Sci.* 2021, *16*, 448–455. [CrossRef]
- 8. Messina, J.P.; Brady, O.J.; Golding, N.; Kraemer, M.U.; Wint, G.; Ray, S.E.; Pigott, D.M.; Shearer, F.M.; Johnson, K.; Earl, L. The current and future global distribution and population at risk of dengue. *Nat. Microbiol.* **2019**, *4*, 1508–1515. [CrossRef] [PubMed]
- 9. Aziz, A.T.; Al-Shami, S.A.; Mahyoub, J.A.; Hatabbi, M.; Ahmad, A.H.; Md Rawi, C.S. Promoting health education and public awareness about dengue and its mosquito vector in Saudi Arabia. *Parasites Vectors* **2014**, *7*, 487. [CrossRef] [PubMed]
- 10. Khaira, U.; Utomo, P.E.P.; Aryani, R.; Weni, I. A comparison of SARIMA and LSTM in forecasting dengue hemorrhagic fever incidence in Jambi, Indonesia. J. Phys. Conf. Ser. 2020, 1566, 012054. [CrossRef]
- 11. Riaz, M.M.; Mumtaz, K.; Khan, M.S.; Patel, J.; Tariq, M.; Hilal, H.; Siddiqui, S.A.; Shezad, F. Outbreak of dengue fever in Karachi 2006: A clinical perspective. *J. Pak. Med. Assoc.* 2009, *59*, 339. [PubMed]
- 12. Bhatnagar, S.; Lal, V.; Gupta, S.D.; Gupta, O.P. Forecasting incidence of dengue in Rajasthan, using time series analyses. *Indian J. Public Health* **2012**, *56*, 281.
- 13. Siregar, F.A.; Makmur, T.; Saprin, S. Forecasting dengue hemorrhagic fever cases using ARIMA model: A case study in Asahan district. In Proceedings of the IOP Conference Series: Materials Science and Engineering, Medan, Indonesia, 21–23 August 2017; p. 012032.
- 14. Suleiman, A.A.; Suleiman, A.; Abdullahi, U.A.; Suleiman, S.A. Estimation of the case fatality rate of COVID-19 epidemiological data in Nigeria using statistical regression analysis. *Biosaf. Health* **2021**, *3*, 4–7. [CrossRef]
- Cummings, D.A.; Iamsirithaworn, S.; Lessler, J.T.; McDermott, A.; Prasanthong, R.; Nisalak, A.; Jarman, R.G.; Burke, D.S.; Gibbons, R.V. The impact of the demographic transition on dengue in Thailand: Insights from a statistical analysis and mathematical modeling. *PLoS Med.* 2009, *6*, e1000139. [CrossRef] [PubMed]
- 16. Johansson, M.A.; Reich, N.G.; Hota, A.; Brownstein, J.S.; Santillana, M. Evaluating the performance of infectious disease forecasts: A comparison of climate-driven and seasonal dengue forecasts for Mexico. *Sci. Rep.* **2016**, *6*, 33707. [CrossRef]
- 17. Xu, H.-Y.; Fu, X.; Lee, L.K.H.; Ma, S.; Goh, K.T.; Wong, J.; Habibullah, M.S.; Lee, G.K.K.; Lim, T.K.; Tambyah, P.A. Statistical modeling reveals the effect of absolute humidity on dengue in Singapore. *PLoS Negl. Trop. Dis.* **2014**, *8*, e2805. [CrossRef]
- Honório, N.A.; Nogueira, R.M.R.; Codeço, C.T.; Carvalho, M.S.; Cruz, O.G.; de Avelar Figueiredo Mafra Magalhães, M.; de Araújo, J.M.G.; de Araújo, E.S.M.; Gomes, M.Q.; Pinheiro, L.S. Spatial evaluation and modeling of dengue seroprevalence and vector density in Rio de Janeiro, Brazil. *PLoS Negl. Trop. Dis.* 2009, *3*, e545. [CrossRef] [PubMed]
- 19. Nishiura, H. Mathematical and statistical analyses of the spread of Dengue. *Dengue Bull.* 2006, 30, 51–67.
- Reiner, R.C., Jr.; Perkins, T.A.; Barker, C.M.; Niu, T.; Chaves, L.F.; Ellis, A.M.; George, D.B.; Le Menach, A.; Pulliam, J.R.; Bisanzio, D. A systematic review of mathematical models of mosquito-borne pathogen transmission: 1970–2010. J. R. Soc. Interface 2013, 10, 20120921. [CrossRef]

- Reich, N.G.; Shrestha, S.; King, A.A.; Rohani, P.; Lessler, J.; Kalayanarooj, S.; Yoon, I.-K.; Gibbons, R.V.; Burke, D.S.; Cummings, D.A. Interactions between serotypes of dengue highlight epidemiological impact of cross-immunity. *J. R. Soc. Interface* 2013, 10, 20130414. [CrossRef]
- Lauer, S.A.; Sakrejda, K.; Ray, E.L.; Keegan, L.T.; Bi, Q.; Suangtho, P.; Hinjoy, S.; Iamsirithaworn, S.; Suthachana, S.; Laosiritaworn, Y. Prospective forecasts of annual dengue hemorrhagic fever incidence in Thailand, 2010–2014. *Proc. Natl. Acad. Sci. USA* 2018, 115, E2175–E2182. [CrossRef]
- Buczak, A.L.; Baugher, B.; Moniz, L.J.; Bagley, T.; Babin, S.M.; Guven, E. Ensemble method for dengue prediction. *PLoS ONE* 2018, 13, e0189988. [CrossRef]
- 24. Banditvilai, S.; Anansatitzin, S. Comparative study of three time series methods in forecasting dengue hemorrhagic fever incidence in thailand. In Proceedings of the International Academic Conferences, Sevilla, Spain, 5–8 March 2018.
- Nájera, J. World Health Organiziation Global Partnership to Roll Back Malaria. Malaria Control: Achievements, Problems and Strategies; WHO/CDS/RBM/99.10. 1999. Available online: www.who.int/malaria/publications/atoz (accessed on 25 August 2022).
- Luz, P.M.; Mendes, B.V.; Codeço, C.T.; Struchiner, C.J.; Galvani, A.P. Time series analysis of dengue incidence in Rio de Janeiro, Brazil. Am. J. Trop. Med. Hyg. 2008, 79, 933–939. [CrossRef]
- 27. Wu, P.-C.; Guo, H.-R.; Lung, S.-C.; Lin, C.-Y.; Su, H.-J. Weather as an effective predictor for occurrence of dengue fever in Taiwan. *Acta Trop.* **2007**, *103*, 50–57. [CrossRef]
- Chen, K.-Y.; Wang, C.-H. A hybrid SARIMA and support vector machines in forecasting the production values of the machinery industry in Taiwan. *Expert Syst. Appl.* 2007, 32, 254–264. [CrossRef]
- 29. Conejo, A.J.; Plazas, M.A.; Espinola, R.; Molina, A.B. Day-ahead electricity price forecasting using the wavelet transform and ARIMA models. *IEEE Trans. Power Syst.* 2005, 20, 1035–1042. [CrossRef]
- Suresh, K.; Krishna Priya, S. Forecasting sugarcane yield of Tamilnadu using ARIMA models. Sugar Tech. 2011, 13, 23–26. [CrossRef]
- 31. Idrees, S.M.; Alam, M.A.; Agarwal, P. A prediction approach for stock market volatility based on time series data. *IEEE Access* **2019**, *7*, 17287–17298. [CrossRef]
- Siami-Namini, S.; Tavakoli, N.; Namin, A.S. A Comparison of ARIMA and LSTM in Forecasting Time Series. In Proceedings of the 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA), Orlando, FL, USA, 17–20 December 2018; pp. 1394–1401.
- 33. Narasimha Murthy, K.; Saravana, R.; Vijaya Kumar, K. Modeling and forecasting rainfall patterns of southwest monsoons in North–East India as a SARIMA process. *Meteorol. Atmos. Phys.* **2018**, *130*, 99–106. [CrossRef]
- 34. Thevakumar, P.; Jayathilaka, R. Exchange rate sensitivity influencing the economy: The case of Sri Lanka. *PLoS ONE* **2022**, 17, e0269538. [CrossRef]
- Chang, X.; Gao, M.; Wang, Y.; Hou, X. Seasonal autoregressive integrated moving average model for precipitation time series. J. Math. Stat. 2012, 8, 500–505. [CrossRef]
- 36. Jain, G.; Mallick, B. A study of time series models ARIMA and ETS. SSRN 2017, 2898968. [CrossRef]
- 37. Han, P.; Wang, P.; Wang, Y. Drought forecasting based on the Standardized Precipitation Index at different temporal scales using ARIMA models. *Agr. Res. Arid Areas* 2008, *26*, 212–218.
- 38. Alyousifi, Y.; Othman, M.; Faye, I.; Sokkalingam, R.; Silva, P.C.L. Markov Weighted Fuzzy Time-Series Model Based on an Optimum Partition Method for Forecasting Air Pollution. *Int. J. Fuzzy Syst.* **2020**, *22*, 1468–1486. [CrossRef]
- 39. Dalah, C.M.; Singh, V.; Abdullahi, I.; Suleiman, A. The study of HIV/AIDS trend in Yobe state for the prescribed period (1999–2019). *Int. J. Stat. Appl.* **2020**, *10*, 10–16.
- 40. Martinez, E.Z.; Silva, E.A.S.d.; Fabbro, A.L.D. A SARIMA forecasting model to predict the number of cases of dengue in Campinas, State of São Paulo, Brazil. *Rev. Da Soc. Bras. De Med. Trop.* **2011**, *44*, 436–440. [CrossRef] [PubMed]
- 41. Johansson, M.A.; Cummings, D.A.; Glass, G.E. Multiyear climate variability and dengue—El Nino southern oscillation, weather, and dengue incidence in Puerto Rico, Mexico, and Thailand: A longitudinal data analysis. *PLoS Med.* 2009, 6, e1000168. [CrossRef]
- 42. Morales, I.; Salje, H.; Saha, S.; Gurley, E.S. Seasonal Distribution and Climatic Correlates of Dengue Disease in Dhaka, Bangladesh. *Am. J. Trop. Med. Hyg.* **2016**, *94*, 1359–1361. [CrossRef]