Comparison of exponential smoothing method and autoregressive integrated moving average (ARIMA) method in predicting dengue fever cases in the city of Palembang

E Munarsih* and I Saluza

1Institute of pharmacy Bhakti Pertiwi Palembang, Jl. Ariodilla III No. 22A Palembang
2Indo Global Mandiri University, Jl. Jend. Sudirman KM 4,5 Palembang

*Corresponding author’s email: ensiwi.munarsih@gmail.com

Abstract. Forecasting is the process of making a statement about an event where the event has not been known or observed. For pharmaceutical students, learning about forecasting techniques can help determine the treatment of various diseases, one of which is dengue fever. Dengue fever is an acute disease caused by the dengue virus. Dengue fever is still a public health issue in major cities in Indonesia, one of which is Palembang. Based on the profile done by Palembang City’s Public Health Office in 2017, dengue fever cases in the area from year to year tend to fluctuates. To get the overview of the number of dengue fever cases in the upcoming years, time series forecasting methods are used, namely the Exponential Smoothing method and the Autoregressive Integrated Moving Average (ARIMA) method. Afterward, the results of predictions from the two methods are compared. Forecasting using the ARIMA method gives the smallest MSE and MAE results of 108077.877 and 172.424, respectively, compared to the Exponential Smoothing method. This means that the ARIMA method is better at predicting the number of dengue fever cases in Palembang in the coming years.

1. Introduction
Dengue fever is an acute disease caused by the dengue virus. This disease is commonly found in tropical and sub-tropical regions, and is widespread in many Southeast Asian countries, including Indonesia. Dengue fever in Indonesia was first reported in Surabaya in 1968, and reported in Jakarta in 1969. In 1994, cases of dengue fever have spread to 27 provinces in Indonesia. Since 1968, the number of dengue fever cases in Indonesia continues to increase. In 1968, there were 53 cases of dengue fever with 24 deaths. Whereas in 1988, there were an increase to 47,573 reports with 1,527 deaths [1].

There are three factors that play a role in the transmission of dengue virus infections, these factors are humans, viruses, and intermediary vectors. The dengue virus is transmitted to humans through the bite of the *Aedes aegypti* mosquito [2]. Dengue fever is still a public health issue in major cities in Indonesia, one of which is Palembang. Based on the profile done by Palembang City’s Public Health Office in 2017, dengue fever cases in the area from year to year tend to fluctuates. The government is expected to be able to maximize prevention efforts in order to reduce the number of dengue fever cases in the coming years. To get the overview of the number of dengue fever cases in the upcoming years, time series forecasting methods are used. Time series forecasting method is a quantitative method used to analyze past data that has been collected regularly using the right techniques. The results can be used as a reference for forecasting data in the future [3].
In recent years, time series forecasting methods are known to be divided into two parts. One of them is the forecasting method based on traditional mathematical models such as exponential smoothing, ARIMA, non-parametric regression methods, etc.

The exponential smoothing method is one method that can be used to forecast data with a moving average. The exponential smoothing method consists of single, multiple, and more complicated methods. All of them have the same properties, where newer values are given relatively greater weight compared to previous observations. Munarsih [4] has conducted research on the application of exponential smoothing methods to predict coffee exports in Indonesia. ARIMA (Autoregressive Integrated Moving Average) is a forecasting model that generates predictions based on the synthesis of historical data patterns.

The ARIMA method was used by Hermawan [5] to predict OLR pented anomalies in the western region of Indonesia. Compared with the Holt-Winter method, the ARIMA method has better results. Rahayu, Istionah, and Sari [6] tested the effectiveness of the Box-Jenkins and Exponential Smoothing methods to forecast the retribution of motor vehicle testing in the Department of Transportation in Klaten, and the results of the Box-Jenkins method were better at predicting data compared to the Exponential Smoothing method. In addition, study of the comparison of forecasting using the exponential smoothing holt-winter and ARIMA methods showed that the exponential smoothing holt-winter method produced a smaller error compared to the ARIMA method, or in other words the method was better than the ARIMA method [7]. The purpose of this study is to predict the number of dengue fever cases in the city of Palembang by using the exponential smoothing and ARIMA methods to further compare the results of the predictions from the two methods.

2. Methods

2.1. Forecasting with the Exponential Smoothing Method
The exponential smoothing method is one of the forecasting methods consisting of single, multiple, and more complicated methods. All of them have the same properties, where newer values are given relatively greater weight compared to previous observations. The exponential smoothing method is a type of forecasting technique for a moving average that weighs past data in an exponential manner so that the latest data has a greater weight in the moving average [8].

\[ S_{t+1} = \alpha X_t + (1 - \alpha)S_t \]  

where,
- \( S_{t+1} \) = forecasting value to t+1,
- \( X_t \) = actual data to t,
- \( \alpha \) = parameter values between 0 to 1,
- \( S_t \) = forecasting value to t.

2.2. Forecasting with the Exponential Smoothing Method
The single exponential smoothing method is actually a development of a simple moving average method. If there is data from the t-observation, the forecast value at time \( t + 1 \) is:

\[ S_{t+1} = \alpha X_t + (1 - \alpha)S_t \]  

2.3. Time Series Model
Autoregressive/ Integrated/ Moving Average (ARIMA) models have been studied in depth by George Box and Gwilym Jenkins (1976). The Autoregressive (AR) model was first introduced by Yule (1926) and later developed by Walker (1931), while the Moving Average (MA) model was first used by Sluzky (1973). But it was Word (1938) which produced the theoretical foundations of the ARMA combination process [8].

2.4. ARIMA Process
AR, MA, and ARMA models are models that use the assumption that the time series data analyzed is stationary. The mean and variance of time series data is constant and the covariance is not affected. In its application, more time series data are non-stationary, or in other words, integrated. To station it, differencing is done as much as d, so that the stationary process is expressed as ARIMA (p, d, q) as follows:

$$
(1 - \phi_1 B + \phi_2 B^2 + \cdots + \phi_p B^p)X_t = (1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q)e_t
$$

and

ARIMA (p,d,q)(P,D,Q)\(^S\)

where,

(p,d,q) = parts that are not seasonal from the model
(P,D,Q) = parts that are seasonal from the model
S = periods per season

There are several stages in carrying out time series analysis, namely identification of models, estimation of models, and diagnosis of models [9]. At the model identification stage, what must be considered is data stationarity. If the data is not stationary, then differencing is done several times until the data is stationary. To stationatethe data, the following equation can be used:

$$
X'_t = X_t - X_{t-1}
$$

The Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) are used in the second stage. ACF and PACF are used to determine the temporary model. After the temporary model is obtained, then the estimation of good model is carried out. In this case, estimating the model is done by the least square method.

2.5. Mean Square Error (MSE)

The MSE approach is used as a measure of forecasting accuracy, as well as determining the magnitude of the average squared error of a forecasting method [8]. With equation as follows:

$$
MSE = \frac{1}{n} \sum_{t=1}^{n} e_t^2 = \frac{1}{n} \sum_{t=1}^{n} (X_t - \hat{X}_t)^2
$$

where,

n = forecasting value

X\(_t\) = observation value at t

\(\hat{X}_t\) = forecasting result’s value at t

2.6. Mean Absolute Error (MAE)

The MAE approach is used to assess the achievements of the ARIMA used, because MAE is able to recognize the significance of the relationship between the observation data and forecasting data with how much error is generated [7]. With equation as follows:

$$
MAE = \frac{1}{n} \sum_{t=1}^{n} |e_t| = \frac{1}{n} \sum_{t=1}^{n} |X_t - \hat{X}_t|
$$

where,

n = forecasting value

X\(_t\) = observation value at t

\(\hat{X}_t\) = forecasting result’s value at t

After getting a good model, the model cannot be directly used as the next data forecast, but a model diagnosis must be done. This is because the model that is made is not necessarily in accordance with the data owned. Therefore, it is necessary to diagnose the model by testing whether the residual estimation results are white noise.

2.7. Data Used
The data used in this study is data on dengue fever cases in the city of Palembang from 2003 to 2016 obtained from the health profile of the city of Palembang [10].

3. Result and Discussion
The time series method is a forecasting method analysis of an estimated variable with a variable or time function. An important step in choosing the right forecasting method is consider the type of data pattern. The ARIMA and exponential smoothing methods have similarity, that both of them can analyze the univariate data. Both also assume past values and errors as the basis for forecasting in the future. However, each method has a difference such as the ARIMA method of analyzing stationary data, so the data that is not stationary must be stationary by transformation or differencing. While the exponential smoothing method can analyze a variety of stationary and non-stationary data. The weakness of the ARIMA method is the lack of ability to produce reliable long-term forecasting due to the presumption of the random-walk type model. While the exponential smoothing method requires time in determining alpha parameters by trial and error. However, there is a fairly small difference in accuracy in forecasting between exponential smoothing methods and the ARIMA method.

Figure Data on the number of dengue fever cases from 2003 to 2016 are illustrated through a time series plot in figure 1 as follows.

![Figure 1. Data plot for the number of dengue fever cases in Palembang City in 2003-2016](image)

3.1. Single Exponential Smoothing
The forecasting stage using the single exponential smoothing method begins with creating a data pattern to see the occurrence of trend and seasonal patterns in the data. From Figure 3, it can be seen that the data does not follow the trend or seasonal data patterns. The next step is to determine the estimated initial value that affects the forecasting, which depends on the length of time and the value of the parameter, which is the mean ($\alpha$).

According to Hendikawati [11] trial and error can be done to find parameter values that give the best results. The best model is obtained by entering parameter values between 0 and 1. The iteration starts by selecting $\alpha$ between 0.1 to 0.9. A large $\alpha$ value (0.9) gives a very small smoothing in forecasting, while a small $\alpha$ value (0.1) gives a large smoothing [8]. Furthermore, the exponential smoothing model selection was carried out using the MAE test and MSE test. The model is said to be the best if it has the smallest MAE and MSE results.

Table 1. Search results for values of $\alpha$, MAE and MSE

| $\alpha$ | MAE      | MSE       |
|---------|----------|-----------|
| 0.15    | 317.7624 | 169062.7  |
| 0.30    | 319.1166 | 163919.5  |
| 0.55    | 280.7933 | 147810.9  |
| 0.75    | 275.9293 | 133944.6  |
From Table 1, it can be seen that the search results of $\alpha$ are obtained $\alpha = 0.9$ which gives the smallest error value.

3.2. *Autoregressive Integrated Moving Average Model (ARIMA)*  
The forecasting stage with the Autoregressive Integrated Moving Average Model (ARIMA) method starts with identifying data whether the variable being predicted stationary in time series or not. Stationary is done to see the values of variables over time varying around the average and constant variance. Figure 2 shows that the data has a high and different height / fluctuations each year, this indicates that the data is not stationary in the variance because it is still changing as the time changes. The stationarity of data is needed because the ARIMA model cannot be built to stationary time series data.

![Figure 2. Data autocorrelation function](image)

In Figure 2, it appears that data is declining, this shows the characteristic of autocorrelation. Data that is not stationary in the mean and variance or shows the characteristics of autocorrelation, needs to go through differencing processes.

![Figure 3. ACF plot after differenced once](image)
In the plot of Figure 3 and Figure 4, it appears that the bar is already inside the line, indicating that the data is stationary. The estimation process is then carried out by entering various ARIMA models consisting of parameters p, d, and q. Because the data has been differenced 1 time, then the d value is written in number 1. Then the assessment stage of the ARIMA model is carried out.

Based on the ACF and PACF pair plots, a temporary model was obtained, namely the ARIMA (1,1,1) model; ARIMA (2,1,1); and ARIMA (2,1,2). ACF and PACF plots is a rough description of the relationship statistics between observational data points in a time series data. ACF and PACF give instructions regarding patterns or models from available data. Next step is the models diagnosed to get the best model using the value of Mean Squared Error (MSE) and Mean Absolute Error (MAE).

Table 2. MSE and MAE values comparison

| Model              | MSE         | MAE         |
|--------------------|-------------|-------------|
| ARIMA (1,1,1)      | 122214.566  | 238.085     |
| ARIMA (2,1,1)      | 112046.181  | 220.478     |
| ARIMA (2,1,2)      | 108077.877  | 172.424     |

From the three temporary ARIMA models, the best ARIMA model is obtained based on the smallest MSE and MAE values as shown in table 2. The best ARIMA model based on the smallest MSE and MAE values which indicates that the low forecasting inaccuracies is the ARIMA model (2,1,2).

3.3. Single Exponential Smoothing VS ARIMA

The single exponential smoothing and ARIMA methods are then compared. In forecasting results with the exponential smoothing method at a value of $\alpha = 0.9$ indicates the smallest error value. In the ARIMA Method the smallest error is found in the ARIMA model (2,1,2). The smallest MSE and MAE results with the single exponential smoothing method are $118298.8$ and $272.4295$, while the smallest MSE and MAE results with the ARIMA method are $108077.877$ and $172.424$. From these results, it can be seen that MSE and MAE with the ARIMA method are smaller than the single exponential smoothing method. Single exponential smoothing and ARIMA models were developed to predict the number of dengue cases in Palembang. There are two types of forecasting, namely forecasting the sample period and forecasting the post-sample period. Sample period forecasting is used to develop models and finally used to produce original forecasting that is used for planning and other purposes.

4. Conclusion

The smallest MSE and MAE results with the single exponential smoothing method are $118298.8$ and $272.4295$ while the smallest MSE and MAE results with the ARIMA method are $108077.877$ and $172.424$. From these results, it can be seen that MSE and MAE with the ARIMA method are smaller.
than the single exponential smoothing method. This means that the ARIMA method is better at predicting the number of dengue fever cases in Palembang for the coming years.

5. References

[1] Hadinegoro S R and Satari H I 2005 Demam berdarah dengue : naskah lengkap, pelatihan bagi pelatih dokter spesialis anak dan dokter spesialis penyakit dalam, dalam tatalaksana kasus DBD (Balai Penerbit Fakultas Kedokteran Universitas Indonesia)

[2] Irianto K 2014 Epidemiologi Penyakit Menular dan Tidak Menular: Panduan Klinis (Bandung: Alfabeta)

[3] Makridakis S, Wheelwright S C and McGee V E 1999 Metode dan aplikasi peramalan (Jakarta: Erlangga)

[4] Munarsih E 2017 Penerapan Metode Penghalusan Eksponensial Untuk Meramal Ekspor Kopi Indonesia Konferensi Nasional Teknologi Informasi Dan Aplikasinya 1–4

[5] Hermawan E 2011 Perbandingan metode box-jenkins dan holt-winters dalam prediksi anomaly OLR pentad di kawasan barat Indonesia Jurnal Sains dan Dirgantara 9 1 25–35

[6] Rahayu P, Istiqomah R N and Sari E R 2016.Efektivitas metode Box-Jenkins dan Exponential Smoothing pengujian kendaraan bermotor Dishub Klaten Jurnal Prosiding Konferensi Nasional Penelitian Matematika Dan Pembelajaranry Universitas Muhammadiyah Surakarta 942–951

[7] Safitri, T, Dwidayati N and Sugiman S 2017 Perbandingan Peramalan Menggunakan Metode Eksponensial Smoothing Holt-Winters dan ARIMA Unnes Journal of Mathematics 6 1 48–58

[8] Makridakis S, Wheelwright S C and Hyndman R J 2008 Forecasting: Methods and Applications (John Wiley and Son)

[9] Winarno W W 2002 Analisa Statistik dan Ekonomi dengan EViews (Yogyakarta: YKPN)

[10] Dinas Kesehatan Palembang 2015 Profil Kesehatan kota Palembang (Palembang: Dinkes Kota Palembang)

[11] Hendikawati P 2005 Peramalan Data RuntunWaktu Metode dan Aplikasinya dengan Minitab dan Eviews (Semarang: FMIPA Universitas Negeri Semarang)

Acknowledgments

The author would like to thanks to Palembang’s Public Health Office that has permitted to access the data, thanks to Mrs. Imelda Saluza for the discussion helpful and for the great colaboration and also to LPPM’s Institute of pharmacy Bhakti Pertiwi Palembang for the permission to do this research.