Application Seasonal Autoregressive Integrated Moving Average to Forecast the Number of East Kalimantan Hotspots

S Wahyuningsih, R Goejantoro, M Siringoringo, A R Saputra, S Aminah
Statistics Study Program, Department of Mathematics, Faculty of Mathematics and Natural Sciences, Mulawarman University, Samarinda, East Kalimantan, Indonesia

Abstract. The objective of this research is to determine the best time series model for forecasting the number of hotspots in East Kalimantan. Seasonal time series model is applied to the data. The results of this research is the best model for the number of hotspots in East Kalimantan is SARIMA (0,1,1)(0,1,1)12.

1. Introduction
Hotspots are highly recommended for early detection of forest and land fires because it is an indicator of forest or land fires in an area. The number of hotspots in East Kalimantan Province observed increase dramatically in early 2016 [1]. Most of it is the heat of the fire recorded by the Terra-Aqua satellite. Based on this, it is important to know the number of hotspots in the future as an anticipation of forest or land fire, so it is necessary to forecast hotspots using one of time series model that is the Seasonal Autoregressive Integrated Moving Average (SARIMA).

SARIMA is a forecasting model used for data that has a seasonal pattern. In hotspot data of East Kalimantan, it is known that the amount from one period to another period are far enough and still has the same hotspot’s range. Forecasting with the SARIMA model will produce estimation of the number of hotspots over the coming time. There are still some difficulties in forecasting although the SARIMA model is more applicable and accurate in modeling seasonal data compared to Autoregressive Integrated Moving Average (ARIMA) model.

There are several preliminary studies related to forecasting hotspots used correlation analysis, linear regression and nonlinear regression method [2], used smoothing method and multiple regression [3], and used Elman recurrent neural network [4].

2. Methodology
Time series analysis comprises methods for analyzing time series data in order to extract meaningful statistics and other characteristics of the data. Time series forecasting is the use of a model to predict future values based on previously observed values. One of the time series model is Seasonal ARIMA or SARIMA \((p,d,q)(P,D,Q)\) \(^S\) [5]:

\[
\phi_p (B) \Phi_P (B^S)(1-B) (1-B^S) D Z_i = \theta_q (B) \Theta_Q (B^S) a_i
\]

where

\((1-B)^d\) : orde differencing non-seasonal
D differently orde differencing seasonal

The hypothesis of this model is:

\[ H_0 : \beta = 0, \beta = \{\phi_h, \Theta_j, \Theta_k\}; h = 1, ..., p; i = 1, ..., q; j = 1, ..., P; k = 1, ..., Q \]

(The parameters in the model is not significantly)

\[ H_1 : \beta \neq 0, \beta = \{\phi_h, \Theta_j, \Theta_k\}; h = 1, ..., p; i = 1, ..., q; j = 1, ..., P; k = 1, ..., Q \]

(The parameter in the model is significantly)

Statistics test:

\[ t_{hit} = \frac{\hat{\beta}}{SE(\beta)} \]

(2)

\[ H_0 \text{ is rejected if} \quad |t_{hit}| > t_{\alpha/2} \quad df = (n - p), \quad p = \text{the number of parameters} \]

Akaike’s Information Criterion (AIC) is used to determine the best model:

\[ AIC = n \ln(\sum_{t=1}^{n} \hat{\alpha}_t^2 / n) + 2f + n + n \ln(2\pi) \]

(3)

3. Result and Discussion

Based on Figure 1, it can be seen that the highest number of hotspots in East Kalimantan Province occurred in October 2015, during which period there were very large cases of forest fires in East Kalimantan. In this study, the data is divided into 2 parts, namely the data before the intervention which is the data of the number of hotspots of East Kalimantan Province in the period January 2010 – September 2015 (Z1t) and data after the intervention which is the data of the number of hotspots of the Province of East Kalimantan in the period October 2015 – December 2018 (Z2t).

An increase in some periods is quite extreme and data fluctuations tend to be unstable over time, indicating the data is not stationary in variance. Another way that can be done to see the stationarity of
data in variance is to do a Box-Cox transformation. The estimated value of \( \lambda \) obtained from the Box-Cox transformation is -0.114887 which shows that the value is not yet close to 1 so that it is known to be not stationary. Thus the data needs to be transformed using a rank transformation (\( Z_{tr}^2 \)) which then checks the stationarity of variance again and an \( \lambda \) value of 1 is obtained which indicates that the data is stationary in variance. The time series plot and Autocorrelation Function (ACF) plot data on the number of hot spots of East Kalimantan after transformation (\( Z_{tr}^2 \)) are shown in Figure 2 and Figure 3 as follows.

![Time series plot of \( Z_{tr}^2 \)](image)

**Figure 2.** Time series plot of \( Z_{tr}^2 \)

![ACF plot of \( Z_{tr}^2 \)](image)

**Figure 3.** ACF plot of \( Z_{tr}^2 \)

Based on Figure 3, it can be seen that the data is not stationary in the average, because plot of data tends to fluctuate and not be around a constant average value over time. Stationary checks in the mean can also be done by looking at the ACF plot which tends to slow down and sinusoidal patterned, so it can be concluded that the data is not stationary in the average. In addition, it can be seen that the data has a repeating pattern every 12 lag or it can be seen the data has a seasonal pattern. Thus the data needs to be done differencing in non-seasonal first order and seasonal first order differencing (\( Z_{tr}^d(d = 1)(D = 1) \)), and obtained a time series plot in Figure 4.

![Time series plot of \( Z_{tr}^d(d = 1)(D = 1) \)](image)

**Figure 4.** Time series plot of \( Z_{tr}^d(d = 1)(D = 1) \)
Based on Figure 4, it can be seen that the data has been stationary in the average, because the data tends to be around a constant average value from time to time. Checking stationarity in the average can be done further by looking at the ACF plot of the data shown in Figure 5. Visually it can be seen that the ACF plot has been stationary in the average. This is indicated by the FOK graph cut off after lag 1. This explains that the data has been stationary in the non-seasonal average. In addition, in Figure 5 shows that the ACF plot is cut off after lag 12, it can be seen that the data has been stationary in the seasonal average.

![ACF plot](image1)

**Figure 5.** ACF plot $Z_{it}(d = 1)(D = 1)$

![PACF plot](image2)

**Figure 6.** PACF plot $Z_{it}(d = 1)(D = 1)$

Temporary identification of the model is formed by looking at the ACF plot in Figure 5 and the Partial Autocorrelation Function (PACF) plot in Figure 6 to determine the order of the AR and MA models both seasonal and non-seasonal. Based on Figure 5 it can be seen that the order for non-seasonal MA is 1 (cut off after lag 1) and based on Figure 6 the order for non-seasonal AR is 2 (cut off after lag 2), with non-seasonal differencing (d) 1 time. In addition, based on Figure 5 it can be seen that the ACF cut-off after lag 12, so the order for seasonal MA is 1. It can be seen that the PACF cut-off after lag 12, so that the order for seasonal AR is 1 , and the seasonal differencing order (D) is 1. In order to obtain a temporary combination of models that can be seen in Table 1 as follows:

| Model                      | Model                      | Model                      |
|----------------------------|----------------------------|----------------------------|
| SARIMA(1,1,0)(1,1,0)       | SARIMA(1,1,0)(0,1,1)       | SARIMA(1,1,0)(1,1,1)       |
| SARIMA(0,1,1)(1,1,0)       | SARIMA(0,1,1)(0,1,1)       | SARIMA(0,1,1)(1,1,1)       |
| SARIMA(1,1,1)(1,1,0)       | SARIMA(1,1,1)(0,1,1)       | SARIMA(1,1,1)(1,1,1)       |
| SARIMA(2,1,0)(1,1,0)       | SARIMA(2,1,0)(0,1,1)       | SARIMA(2,1,0)(1,1,1)       |
| SARIMA(2,1,1)(1,1,0)       | SARIMA(2,1,1)(0,1,1)       | SARIMA(2,1,1)(1,1,1)       |

Table 1. ARIMA Temporary Model
Parameter estimation is done using the Maximum Likelihood method with the help of Rstudio Software. After that a diagnostic check is performed, and the best model is obtained SARIMA(0,1,1)(0,1,1)$^{12}$ with an AIC value of -72.05.

\[
(1-B)(1-B^{12})Z_{t,i}^* = \theta_1(B)\Theta_1(B^{12})a_t
\]

\[
(1-B)(1-B^{12})Z_{t,i}^* = (1-\theta_1 B)(1-\Theta_1 B^{12})a_t
\]

\[
Z_{t,i}^* = Z_{t,i-1}^* + Z_{t,i-12}^* - \theta_1 a_{t-1} - \Theta_1 a_{t-12} + \theta_1 \Theta_1 a_{t-13} + a_t
\]

\[
Z_{t,i}^* = Z_{t,i-1}^* + Z_{t,i-12}^* - Z_{t,i-13}^* + 0.6634a_{t-1} + a_{t-12} + 0.6634a_{t-13} + a_t
\]

(4)

4. Conclusion
Forecasting hotspots can be used seasonal time series model. The best model for hotspots in East Kalimantan is SARIMA(0,1,1)(0,1,1)$^{12}$ with an AIC value of -72.05. Forecasting the number of East Kalimantan hotspots did not occur in 2019.

References
[1] LAPAN 2016 Panduan Teknis (V.01) Informasi Titik Panas (Hotspot) Kebakaran Hutan/Lahan. (Jakarta : Pusat Pemanfaatan Penginderaan Jauh Deputi Bidang Penginderaan Jauh)
[2] Spessa A C, Field R D, Pappenberger F, Langner A, Englhart S, Weber U, Stokdale T, Slegert F, Kaiser J W and Moore J 2015 Seasonal Forecasting of Fire Over Kalimantan Indonesia Natural Hazards Earth System Sciences 15 p429-442
[3] Halide H 2016 Applying Seasonal Climate Information for Wildland Fires as Eearly Warning System in Southern ASEAN Proceeding Seminar Nasional Geofisika
[4] Amaranagana T T and Agmalaro M A 2016 Temporal Prediction for Hotspots Occurrence in Riau Province using Elman Recurrent Neural Network (IPB)
[5] Wei W W S 2006 Time Series Analysis: Univariate and Multivariate Methods. (Kanada: Addison Wesley Publishing Company)
[6] Dewa G A P I, Haryanto E, Sopaheluwakan A, Pradana R P and Haryoko U 2019 Sebaran Spasial dan Temporal Titik Panas (Hotspot) di Indonesia dari Satelit Modis dengan Metode Gridding Prosiding Seminar Nasional Geomatika 2018 p 1123-1128