

SAR Journal Artikel

by Sri Listia Rosa

Submission date: 26-Dec-2022 05:12PM (UTC+0700)

Submission ID: 1986680480

File name: J-5_SARJournal_Prediction_of_Hotspots_in_Riau_Province-min.pdf (418.77K)

Word count: 5760

Character count: 29174

5 Prediction of Hotspots in Riau Province, Indonesia Using the Autoregressive Integrated Moving Average (ARIMA) Model

Evizal Abdul Kadir¹, Nur Ezzati Dayana², Sri Listia Rosa¹
Mahmod Othman³, Rizauddin Saian²

¹Department of Informatics Engineering, Faculty of Engineering, Universitas Islam Riau Marpoyan, Pekanbaru, Riau, 28284 Indonesia

²Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA Arau, Perlis, 02600 Malaysia

³Department of Fundamental and Applied Science, Universiti Teknologi PETRONAS Seri Iskandar, Perak, 32610, Malaysia

3 **Abstract** – Various forms of disasters occur worldwide, one of which is fire. Indonesia has been suffering from frequent land and forest fires. These events are not a new phenomenon and seem to be an annual tradition, especially in the dry season. This nation was most affected by an excessively disastrous forest fire in 2015. The misfortunes suffered were massive and resulted in land and forest damage that may have great economic and environmental costs. One solution to reduce the impacts of such events is to predict the emergence of hotspots. Therefore, in this work, a modeling method using time series produced by the Box-Jenkins' Autoregressive Integrated Moving Average (ARIMA) model was used to predict the appearance of hotspots. Since the forecasting system does not expect any detailed form to be predicted in terms of the time series of historical data, the data demonstrated in the proposed model were different from data from other models used for prediction. The study was conducted based on monthly hotspot occurrence data from January 2014 through June 2019 in Riau Province, Indonesia.

5 The data were downloaded from the collection of the "LAPAN-MODIS-Catalog". Based on the results shown, the Autoregressive Integrated Moving Average (ARIMA) model (2,1,2) produced good predictions based on its lowest value of Mean Squared Error (MSE), 9540.088. Moreover, the proposed model has produced highly accurate forecasts of hotspots for time periods of up to five months using the fitting model of ARIMA (2,1,2), and the values forecasted for 5 months ahead were 25, 31, 26, 30 and 27.

Keywords – hotspot, prediction, ARIMA, Riau province, Indonesia.

1. Introduction

Fire in general can be defined as a situation where a building, such as a market, factory or house, is subject to fire that causes casualties and losses. Forest and land fires can be defined as conditions where forests and land are subject to fires that result in land and forest damage that may cause great economic and environmental costs. These events are usually caused by natural factors such as hot weather in the dry season, lightning and volcanic eruptions. In addition, forest and land fires can also happen due to human activities such as logging, illegal burning and land clearing. The consequences of these events are often haze disasters that can affect the activities and health of surrounding communities.

Indonesia is a nation in which land and forest fire disasters are highly possible. These events are not new phenomena but seem to be an annual tradition, especially in the dry season. This nation was affected by its worst land and forest fire disaster in the year 2015. Riau is one of the regions in Indonesia that contribute smoke every year because of peatland conditions and irresponsible parties who tend to clear land by burning. Fires on peatlands are said to be more dangerous than fires on dry land. The peatlands are dry and flammable during the dry season, with

DOI: 10.18421/SAR33-03

<https://doi.org/10.18421/SAR33-03>


Corresponding author: Evizal Abdul Kadir, Department of Informatics Engineering, Faculty of Engineering, Universitas Islam Riau Marpoyan, Pekanbaru, Riau, 28284 Indonesia.
Email: evizal@eng.uir.ac.id

Received: 12 July 2020.

Revised: 01 September 2020.

Accepted: 09 September 2020.

Published: 29 September 2020.

14  © 2020 Evizal Abdul Kadir et al; published by UIKTEN. This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivs 3.0 License.

The article is published with Open Access at www.sarjournal.com

fuel in the form of plant residues. Therefore, small fires can enter under the surface of the soil and slowly burn the peatlands in a way that is difficult to detect them and that presents the danger of thick smoke. Such burning is hard to extinguish and can probably last for weeks.

The main districts that have peatland distributions with very deep peatland categories are the Bengkalis and Dumai Regencies (Cities) of 470,985.5 ha, the Rokan Hilir District (303,639 ha), the Pelalawan District (275,428.1 ha), the Siak and Pekanbaru Regencies (Cities) (258,231.9 Ha), the Indragiri Hilir Regency (181,864.4 ha), the Meranti Regency (84,114.9 ha), the Indragiri Hulu Regency (71,976.3 ha), the Rokan Hulu Regency (28,499.3 ha), and the Kampar Regency (18,419.7 ha). The Kuantan Singingi District is considered to have no peatland distribution [1],[2],[3].

For more than a decade, hotspot information has been one of the references for monitoring the appearance of land and forest fires worldwide, especially in Indonesia using remote sensing data. Hotspots are hot points on the surface of the earth indicative of forest and land fires. Forecasting the appearance or occurrence of hotspots has become a vital issue for forecasters because it is one of the indications that can lead to reductions in the occurrence of forest and land fires and can provide early warning information to communities. Therefore, this study implemented and used the Autoregressive Integrated Moving Average (ARIMA) method to predict the appearance of hotspots several months in advance. The research aims are to implement the Autoregressive Integrated Moving Average (ARIMA) method in modeling the time series hotspot data in Riau Province, Indonesia, as well as to evaluate the model generated as a useful method in the prediction of the appearance of fire hotspots several months in advance of their occurrence.

2. Related Work

Land and forest fires have disadvantages to human life because of the damage they produce. Many ways to observe the occurrence of land and forest fires, such as satellite imagery used to identify the hotspots in certain map areas, remote sensing systems, and ground sensing by the installation of sensor nodes in areas of potential forest fires. Hotspots can also mean zones with surface temperatures above normal levels, which can be determined by the techniques mentioned above. Normally, when a forest fire of sufficient size and temperature occurs in an area, it is then detected by remote sensing or satellite imagery as a hotspot. For example, a fire of 1 km x 1 km are is likely to be discovered as a hotspot, and

[4] has claimed that if an area surpasses the 46.85°C limit, a hotspot will be found.

As was mentioned above, hotspots are one of the fire indicators, and it is thus very important to know the appearance of this indicator to reduce the incidence of forest and land fires. Various studies have been conducted on the prediction of hotspot occurrences. In [4], research on classification rules for the occurrence numbers of hotspots in the Bengkalis area was reported. The classification rules were generated using a spatial entropy-based decision tree algorithm based on fire data, including data of temperature, city centers, and wind speed. The result of this research using the fivefold cross-validation test was decision trees with the normal accuracy of 89.04% on the training set and 52.05% on the testing set. From the decision tree, 255 out of 560 rules were obtained to classify hotspot occurrences. In addition, most hotspots occur in regions where the community's source of revenue is planting. Users can evaluate the cause of the hotspots in the presence of socioeconomic circumstances by integrating this socioeconomic dimension [5].

Additionally, in [6] a study was conducted to calculate the appearances and causes of land and forest fires in the Central Kapuas Area in Kalimantan province. Spatial analysis was used to identify the number of hotspots with rainfall to determine the temporal distribution of hotspots, and causative variables were derived to recognize the spatial characteristics of fires. The authors claimed that whenever the monthly rainfall distribution decreases, the number of hotspots will increase. However, they also claimed that a high number of hotspots does not continually occur in the region of the Kapuas district since fire events there are caused by the presence of human activities, land cover and peatlands.

On the other hand, an analysis of the driving factors, the pressures that arise and the responses to the fires was reported in [7],[8]. The research was conducted in the Pelalawan area of Riau province using two approaches: the analysis of biophysical factors such as soil properties and semi structured interviews for social and policy aspects. Analysis of variance (ANOVA) values and the statistics simulation software were used to analyze the soil properties and estimate the general linear model, respectively. It was shown that the relation between forest functions and hotspot density is positive, with an R-squared value of 0.9868, whereas the nearest distance to roads has the smallest relationship, with a hotspot density R-squared value of 0.1612.

Finally, a temporal prediction of the appearance of hotspots in the Rokan Hilir district, Riau province, Indonesia was produced by [9] using a Long Short Term Memory Recurrent Neural Network (LSTM RNN). This research was implemented using a model

developed to predict the emergence of hotspots by considering the pattern of training data inferred from time-shaped data. The result achieved was considered good at forecasting the actual data but not sufficiently good at very high values. The results indicated that the RNN LSTM model had a correlation value of 0.99 and that the RMSE value was 88.54. It was stated that this method is better in forecasting the appearance predictions of hotspots even though the error is still quite high compared with ERNN. Moreover, [10] also performed the same research in predicting the appearance of hotspots. In this work, data modeling was conducted using the Autoregressive Integrated Moving Average (ARIMA) model based on monthly data of hotspots in the period from year 2001 to 2012. At the end of the study, the conclusion was that the ARIMA (2,0,0) model is one of the best models to use to forecast the numbers of hotspots, with a MAPE value of 40.974.

Times series data are the values of variables that change over a period of time, which can be periods of weeks, months, years and so on. Currently, forecasting is a crucial and important approach for efficient and effective planning. There are two strategies for time series forecasting: smoothing and modeling. Smoothing is isolated and based on its data type. For constant data types, forecasting is done using Single Moving Average and Single Exponential Smoothing. For trend data types, forecasting uses Double Moving Average and Double Exponential Smoothing. ARIMA is used for modeling time series forecasting.

In [11],[12] much research is reported on the prediction of daily and monthly average global solar radiation in Seoul, South Korea using the time series ARIMA model. Their target was to develop a seasonal autoregressive integrated moving average based on hourly solar radiation data from the year 1981 to the year 2017. The results show that ARIMA (1,1,2) and the Seasonal Autoregressive Integrated Moving Average (SARIMA) (4,1,1) of 12 lags can be used as forecasting models to represent daily and monthly solar radiation, respectively, and in [13],[14],[15],[16] implemented an ARIMA model to forecast Bitcoin exchange rates in a high volatility environment. From the results, ARIMA (2,1,2) was demonstrated to be a perfect forecasting technique, with an R-squared value of 0.444432, and was considered a model with good fitness because it produced a small MAPE value (5.36%) in the error analysis of forecasting compared with actual data.

Research was reported in [17],[18],[19] on a time series ARIMA application used to produce a forecasting model to predict the production of sugarcane in India using secondary data of a 62-year period of sugarcane production. The research showed that ARIMA (2,1,0) was the best candidate model for forecasting sugarcane production 5 years ahead and

that the ARIMA distribution appeared to be normal, with constant variance and a mean of zero. Moreover, it was concluded that this model seemed to produce a suitable prediction model for the production of sugarcane in India, with an overall increase in production predicted. Since the conditions of the peatland distributions are one of the contributors to the occurrence of forest and land fires, in this research the ARIMA model was used to forecast the number of fire hotspots in Riau province.

3. Research Method

The data on the number of fire hotspots in Riau were collected from the "LAPAN-MODIS-Catalog", in which the hotspot information and guides are given. This study only focuses on the data collected in Riau province, Indonesia. The monthly fire hotspot data were collected from year 2014 through year 2019. Riau province was selected as one of the main provinces that have very deep peatland distributions, totalling 470,985.50 ha in size.

In the arrangement of observed data characterized by the time series over the time period, the ARIMA model is one of the models that have capabilities to present stationary data as well as nonstationary time series data. ARIMA models can thus deliver precise predictions based on the historical data of single variables [20],[21]. The approach of the Box-Jenkins methodology is used to construct ARIMA models based on initial data investigation, model identification, model validation and checking and selection of the model's use. Data pertaining to the number of fire hotspots in Riau province for months from January 2014 through June 2019 as shown in Table 1. were used.

The autocorrelation coefficient is the key function of the analysis of time series data; it can produce the correlation of time series data by itself. Equation (1) defines the autocorrelation function.

$$r_k = \frac{\sum_{t=k+1}^n (y_t - \bar{y})(y_{t-k} - \bar{y})}{\sum_{t=1}^n (y_t - \bar{y})^2} \quad (1)$$

r_1 represents the correlation of values of Y that are one time period apart, r_2 represents the correlation of values of Y that are 2 periods apart, and so on. The autocorrelations at lag numbers 1, 2, ..., produce the Autocorrelation Function (ACF).

The ACF is a respected function in investigations of the properties of time series empirical data. Rather than study the ACF model in a time but the results shows is not significantly better repeats in time series. For example, in the first number of 10 (r_1 through r_{10}), a common test is the Box-Pierce test, which is based on the Box-Pierce Q statistic as shown in (2).

$$Q = n \sum_{k=1}^h r_k^2 \tag{2}$$

13 An alternative portmanteau test is the Ljung-Box test:

$$Q^* = n(n+2) \sum_{k=1}^h (n-k)^{-1} r_k^2 \tag{3}$$

where:

- 15 Q* is a Chi-Square distribution with (h-m) degrees of freedom.

- 7 The data are not white noise when the values of Q or Q* are larger than the given value in a chi-square table with $\alpha = 5\%$.

The higher order of the Moving Average (MA) model as the general model of order q can be written as in (4):

$$y_t = C + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} + \dots + \theta_q e_{t-q} \tag{4}$$

where:

- 7 C is the static value
- θ_j is the MA parameter.
- e_{t-k} is the fault term in time t-k

Table 1. Hotspot data year 2014 through year 2019

8 Year	Month	No. of fire hotspots in Bengkalis District	Year	Month	No. of fire hotspots in Bengkalis District
2014	January	15	2017	January	25
	February	66		February	11
	March	122		March	18
	April	21		April	20
	May	5		May	24
	June	15		June	15
	July	11		July	55
	August	20		August	16
	September	40		September	30
	October	10		October	19
	November	10		November	10
	December	10		December	25
2015	January	11	2018	January	24
	February	75		February	48
	March	21		March	28
	April	65		April	30
	May	15		May	11
	June	44		June	16
	July	16		July	43
	August	26		August	13
	September	52		September	12
	October	16		October	10
	November	30		November	10
	December	28		December	20
2016	January	41	2019	January	35
	February	74		February	35
	March	22		March	27
	April	20		April	33
	May	26		May	88
	June	36		June	40
	July	32		July	-
	August	11		August	-
	September	42		September	-
	October	13		October	-
	November	20		November	-
	December	12		December	-

3.1. Initial Data Investigation

Initial data to conduct an investigation with simple data were chosen to recognize the basic parameters of the time series method and to identify any abnormal characteristic in the existing cases. The investigation was conducted by creating a simple time series data plot as shown in Fig. 1. and then fitting a linear trend. A cursory observation indicates that the time series of monthly hotspots in the Bengkalis district, Riau, Indonesia is not stationary since its trend sometimes decreases and sometimes increases. However, the series do not indicate the presence of seasonal effects that could be construed as irregular effects, although significantly large values at the time points March 2014 and May 2019 were observed. Since the data showed that the monthly hotspot data obtained were not stationary in terms of variable values over time, the data transformation was performed using the Box-Cox transformation in Minitab software.

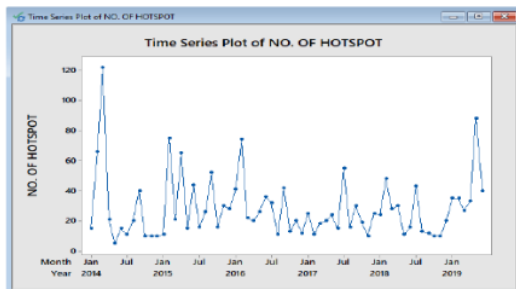


Figure 1. Time plot of number of fire hotspots and trend line

The transformation was repeatedly performed until the hotspot data were stationary over time. The transformation was conducted whenever the rounded value achieved in the Box-Cox Plot was equal to 1, as demonstrated in Fig. 2., the transformation for the Box-Cox Plot of C1 when the value of λ is in the range of -5 to 2.

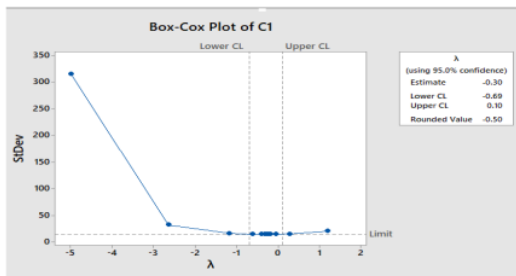


Figure 2. First transformation of data hotspots in a box-cox plot

The next step is transformation for the second round of the Box-Cox Plot of C2 over the time variation. Fig. 3. shows the results with values of λ

between -5.0 and 5.0. This procedure was performed several times to obtain better simulation results of the transformation data of the fire hotspots, with the estimation value and lower Confidence Limit (CL) as well as the upper CL set to achieve the number of prediction fire hotspots for the future time.

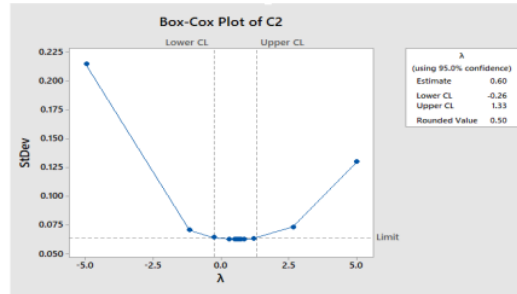


Figure 3. Second transformation of data hotspots in a box-cox plot

The next step is transformation for the third round of the Box-Cox Plot C3 over the time variation. Fig. 4. shows the results with values of λ between -5.0 and 5.0.

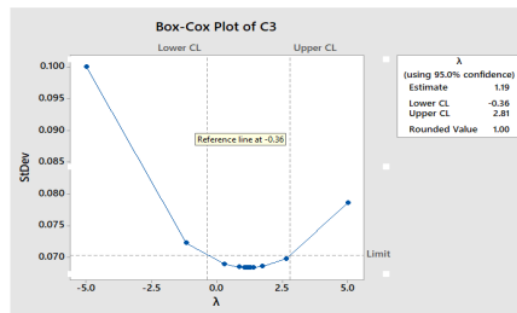


Figure 4. Third transformation of data hotspots in a box-cox plot

The Autocorrelation Function (ACF) and the Partial Autocorrelation Function (PACF) were also plotted to collect more conclusive evidence on stationarity. Fig. 5. shows the pattern of the autocorrelations, which demonstrates that the spike for MA exceeded the significance limit at two points on the upper and lower lines.

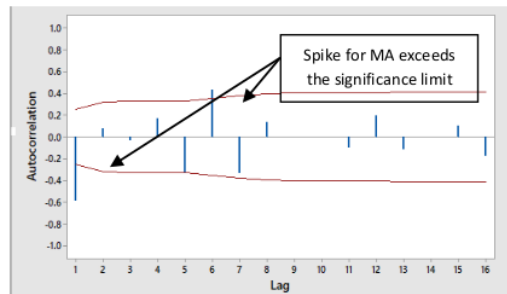


Figure 5. ACF plot of monthly hotspots

The partial autocorrelations are shown in Fig. 6. Two values of the autocorrelations exceed the significance limit. Therefore, to find a stationary time series plot, first-order differencing was performed, which shows the spike for the Auto Regression (AR) at the upper line.

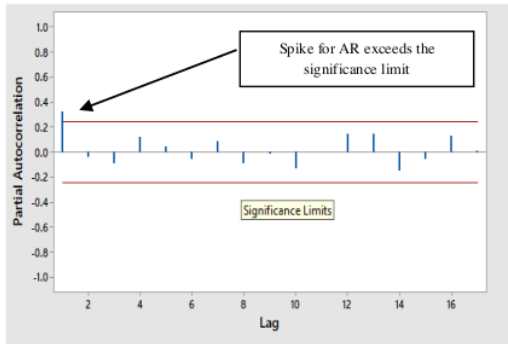


Figure 6. PACF plot of monthly hotspots

3.2. Performing the First-Order Differencing

First-order differencing was performed to render the original series stationary. The evidence of the stationarity condition is obtained by plotting the ACF and PACF as shown in Fig. 7., in which the spike of MA at the lower line with two points of the spike is shown.

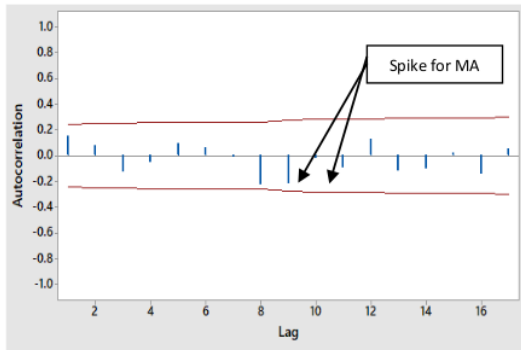


Figure 7. ACF plot of monthly hotspots after first-order differencing

In Fig. 8., the decay patterns in both ACF and PACF have disappeared and none of the autocorrelation values exceed the significance limits. Therefore, from the ACF and PACF results achieved, one can conclude that the series is now stationary. However, time series data are not required to be perfect results due to the fact that series data in such conditions may simply not be able to be achieved because of unexplainable aspects inherent in the datasets. The MA spike is shown at the lower line, and there are two points of the spike.

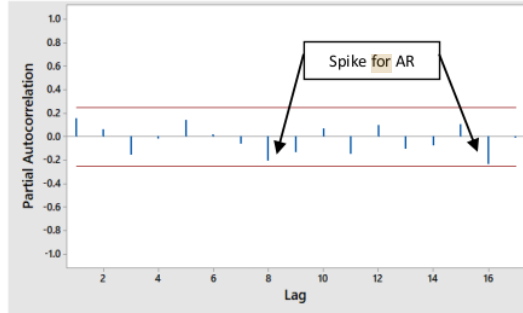


Figure 8. PACF plot of monthly hotspots after first-order differencing

3.3. Model Identification

The next step of analysis in this model is the third step, to identify the model. In this step, the orders (p,d,q) are determined, which involves some components such as AR and MA models. To identify appropriate models to fit the series data involves analysis processes of the PACF and ACF of the stationary series as shown in the Fig. 9, and Fig. 10. representations of the series data.

Model 1: ARIMA (2,1,2)

Final Estimates of Parameters

Type	Coef	SE Coef	T-Value	P-Value
AR 1	0.392	0.131	2.99	0.004
AR 2	-0.299	0.136	-2.20	0.031
MA 1	1.0938	0.0477	22.95	0.000
MA 2	-0.1111	0.0553	-2.01	0.049
Constant	-0.684	0.832	-0.82	0.414

Differencing: 1 regular difference

Number of observations: Original series 66, after differencing 65

Residual Sums of Squares

DF	SS	MS
60	572405	9540.08

Back forecasts excluded

Modified Box-Pierce (Ljung-Box) Chi-Square Statistic

Lag	12	24	36	48
Chi-Square	3.21	9.64	14.75	20.71
DF	7	19	31	43
P-Value	0.865	0.961	0.994	0.998

Figure 9. Final estimation of parameters for ARIMA (2,1,2)

Some challenges exist in specifying a base model in terms of exact results of PACF and ACF since the magnitude impact cannot be summarized on the spikes model. In this predicament there are some models that can possibly be specified, the required tests for validation are performed, and estimates are then produced in the tests. The superior results from one of the models can be seen in the test results.

▪ Model 2: ARIMA (2,1,1)

Final Estimates of Parameters

Type	Coef	SE Coef	T-Value	P-Value
AR 1	-1.404	0.112	-12.51	0.000
AR 2	-0.549	0.104	-5.27	0.000
MA 1	-0.9770	0.0371	-26.31	0.000
Constant	-8.8	25.6	-0.34	0.732

Differencing: 1 regular difference
 Number of observations: Original series 66, after differencing 65

Residual Sums of Squares

DF	SS	MS
61	694000	11377.1

Back forecasts excluded

Modified Box-Pierce (Ljung-Box) Chi-Square Statistic

Lag	12	24	36	48
Chi-Square	23.89	37.42	51.39	61.04
DF	8	20	32	44
P-Value	0.002	0.010	0.016	0.045

Figure 10. Final estimation of parameters for ARIMA (2,1,1)

Based on the model identification in Fig. 11. and Fig. 12. and the number of significant spikes, four models were identified and estimated using Minitab Software.

▪ Model 3: ARIMA (1,1,1)

Final Estimates of Parameters

Type	Coef	SE Coef	T-Value	P-Value
AR 1	0.222	0.127	1.75	0.086
MA 1	0.9761	0.0568	17.18	0.000
Constant	-0.651	0.834	-0.78	0.438

Differencing: 1 regular difference
 Number of observations: Original series 66, after differencing 65

Residual Sums of Squares

DF	SS	MS
62	618717	9979.31

Back forecasts excluded

Modified Box-Pierce (Ljung-Box) Chi-Square Statistic

Lag	12	24	36	48
Chi-Square	3.28	10.53	17.13	21.81
DF	9	21	33	45
P-Value	0.952	0.971	0.990	0.999

Figure 11. Final estimation of parameters for ARIMA (1,1,1)

6

Table 2. Summary of portmanteau test

Statistics	ARIMA (2,1,2)	ARIMA (2,1,1)	ARIMA (1,1,1)	ARIMA (1,1,0)
Calculated Q	3.21	28.89	3.28	21.45
DF	7	8	9	10
Tabulated Q	14.06	15.50	16.91	18.30
Decision (5% of sig. level)	Accept H ₀	Accept H ₁	Accept H ₀	Accept H ₁
Conclusion	The error is white noise (model is well specified)	The error is not white noise (model is not well specified)	The error is white noise (model is well specified)	The error is not white noise (model is not well specified)
MSE	9540.088	11377.10	9979.31	12530.90

The identified models for the final estimation of the parameters of fire hotspots using ARIMA models shown in Fig. 9., Fig. 10., Fig. 11., and Fig. 12. are ARIMA (2,1,2), ARIMA (2,1,1), ARIMA (1,1,1) and ARIMA (1,1,0), respectively. The ARIMA models are used because of the common models and the advantages of the method in mathematical modeling of data predictions for future times.

▪ Model 4: ARIMA (1,1,0)

Final Estimates of Parameters

Type	Coef	SE Coef	T-Value	P-Value
AR 1	-0.464	0.109	-4.26	0.000
Constant	-5.1	13.9	-0.37	0.715

Differencing: 1 regular difference
 Number of observations: Original series 66, after differencing 65

Residual Sums of Squares

DF	SS	MS
63	789444	12530.9

Back forecasts excluded

Modified Box-Pierce (Ljung-Box) Chi-Square Statistic

Lag	12	24	36	48
Chi-Square	21.45	35.09	47.46	55.42
DF	10	22	34	46
P-Value	0.018	0.038	0.062	0.161

Figure 12. Final estimation of parameters for ARIMA (1,1,0)

3.4. Validation and Diagnostics Checking of the Model

A good model is fitted to the residuals obtained and is estimated to have the property of “white noise”. Therefore, validation of the model and checking diagnostics in the analysis involves evaluating the residuals for resemblance to white noise characteristics. A simple method is to observe and analyze the PACF and ACF with respect to the residuals (errors). When the residuals are white noise, it is estimated that there are no significant autocorrelation coefficients and that no partial autocorrelation coefficients exist. However, the condition of stationarity of the residuals (errors) is obtained. A more sophisticated method of creating the stationary condition of the residuals is to check the Ljung-Box Q statistic. In this step, a simple process known as the portmanteau test is used to check the occurrence of the correlation among the residuals by calculating the chi-square (calculate Q) values of the mean error terms [22]. This statistic was provided by Minitab Software and the simulation model. The hypotheses are:

H_0 : the errors are random (white noise) / (model is well specified)

H_1 : the errors are nonrandom (not white noise) / (model is not well specified)

The summary of the obtained statistics from correct models using Minitab are tabulated in Table 2.

3.5. Use of Selection Models

In this step, the Chi-square values (calculated Qs) are checked and compared against the tabulated values to identify either that the model accepted the null hypothesis, which means that the model is well specified and adequate, or vice versa. From the summary of the portmanteau test, the ARIMA (2,1,2) and ARIMA (1,1,1) models are well specified since they accept the null hypothesis that the error is white noise, whereas this is not true for ARIMA (2,1,1) and ARIMA (1,1,0), for which the errors are not white noise. Therefore, based on the smallest Q statistic, ARIMA (2,1,2) is the best model.

Nonetheless, to make sure the decision made above is true for further analysis, the particular MSEs are used. According to the results achieved, ARIMA (2,1,2) has the lowest values of MSE (9540.088), followed by ARIMA (1,1,1), ARIMA (2,1,1) and ARIMA (1,1,0), with MSE values of 9979.31, 11377.10 and 12530.90, respectively. Therefore, the result again points toward ARIMA (2,1,2), and this model was selected as the best model for predictions of the occurrence of hotspots in Bengkalis district, Riau province, Indonesia.

4. Results and Discussion

Table 3. show the summary of the portmanteau test for the four ARIMA models ARIMA (2,1,2), ARIMA (2,1,1), ARIMA (1,1,1) and ARIMA (1,1,0). From the table, there are several characteristics that can be measured to select the best ARIMA model for the temporal predictions of the appearance of hotspots in Bengkalis, Riau. In this step, the values of Calculated Qs are checked and compared against the tabulated values, and it is shown that two out of four ARIMA models, ARIMA (2,1,2) and ARIMA (1,1,1), accept the null hypothesis, which indicates that the errors are random because the value of Calculated Q is less than Tabulated Q; values were $3.21 < 14.06$ and $3.28 < 16.91$, respectively.

Therefore, the conclusions are that these two models are well specified and adequate. In contrast, ARIMA (2,1,1) and ARIMA (1,1,0) are not well specified models since both models have Calculated Q values greater than the Tabulated Q, that is, $28.89 > 15.50$ and $21.45 > 18.30$, respectively. To choose the best model among ARIMA (2,1,2) and ARIMA (1,1,1), measurement of the smallest Q statistics was conducted. Based on this measurement, ARIMA (2,1,2) was chosen as the best model. However, to make sure that the above choice is true for further analysis, MSEs were used for verification. According to the smaller value of MSE, the results again point towards ARIMA (2,1,2). In contrast, by introducing the concept of parsimony, ARIMA (1,1,1), although having marginally greater MSE values and calculated Q than ARIMA (2,1,2), will be the obvious choice.

Table 3. Summary of portmanteau test

Statistics	ARIMA (2,1,2)	ARIMA (2,1,1)	ARIMA (1,1,1)	ARIMA (1,1,0)
Calculated Q	3.21	28.89	3.28	21.45
DF	7	8	9	10
Tabulated Q	14.06	15.50	16.91	18.30
Decision (5% of sig, level)	Accept H_0	Accept H_1	Accept H_0	Accept H_1
Conclusion	The error are white noise	The error are not white noise	The error are white noise	The error are not white noise
MSE	9540.088	11377.10	9979.31	12530.90

10

ARIMA (2,1,2) was selected as the best model with the smallest Mean Square Error (MSE) and was used to predict the appearance of the fire hotspots in Bengkalis, Riau province, Indonesia. Table 4. displays the forecast values of the appearance of fire hotspots for 5 months ahead.

Table 4. Appearance of number fire hotspots for 5 months ahead

Months	Appearance of number fire hotspots
1	25
2	31
3	26
4	30
5	27

5. Conclusion

The results showed that the monthly hotspot data obtained were not stationary because the time varied in the data, and the data transformation was thus done using the Box-Cox transformation in the Minitab software. Model ARIMA (2,1,2) was selected and could be used for monthly data modeling of hotspots for 5 months ahead since it fulfilled all the criteria in the portmanteau test. By looking at the measurement of errors, ARIMA (2,1,2) had the smallest Mean Square Error (MSE) compared with the other models, 9540.088. The results also show that the model was well specified and could be used for predicting the number of hotspots in Bengkalis, Riau province because the model accepted the hypothesis that the error is random or white noise. This is explained by the value of Q being calculated as smaller than Q in the tabulation, that is, 3.21 is less than 14.06. By using the best selected ARIMA (2,1,2) model, the predictions of the appearance of hotspots for 5 months ahead were thus 25, 31, 26, 30 and 27.

Acknowledgements

The authors would like to say thank you very much to KEMENRISTEKDIKTI Indonesia for funding this research and the collaboration research between Universitas Islam Riau and Universiti Teknologi PETRONAS as well as the collaboration with Universiti Teknologi MARA Perlis.

References

- [1]. I. R. Amalia.(2018). Mitigasi Bencana Kebakaran Lahan Gambut di Provinsi Riau Menggunakan Teknologi Modifikasi Cuaca (Studi Kasus : Kabupaten Indragiri Hilir, Kabupaten Kepulauan Meranti, Kabupaten Pelalawan dan Kabupaten Siak), *J. Online Mhs. Fakultas Teknik*, 5(1), 1-8.
- [2]. Turmudi, T., Kardono, P., Hartanto, P., & Ardhisari, Y. (2018). Forest and Land Fire Prevention Through the Hotspot Movement Pattern Approach. *E&ES*, 123(1), 012027.
- [3]. Guha, B., & Bandyopadhyay, G. (2016). Gold price forecasting using ARIMA model. *Journal of Advanced Management Science*, 4(2), 117-121.
- [4]. Nurpratami, I. D., & Sitanggang, I. S. (2015). Classification rules for hotspot occurrences using spatial entropy-based decision tree algorithm. *Procedia Environmental Sciences*, 24(2015), 120-126.
- [5]. Thariqa, P., & Sitanggang, I. S. (2015). Spatial online analytical processing for hotspots distribution based on socio-economic factors in Riau Province Indonesia. *Procedia Environmental Sciences*, 24, 277-284.
- [6]. Thoah, A. S., Saharjo, B. H., Boer, R., & Ardiansyah, M. (2019). Characteristics and causes of forest and land fires in Kapuas District, Central Kalimantan Province, Indonesia. *Biodiversitas Journal of Biological Diversity*, 20(1), 110-117.
- [7]. Tata, H. L., Narendra, B. H., & Mawazin, M. (2018). Forest and land fires in Pelalawan District, Riau, Indonesia: Drivers, pressures, impacts and responses. *Biodiversitas Journal of Biological Diversity*, 19(2), 494-501.
- [8]. Wahyuningsih, S., Goejantoro, R., & Rizki, N. A. (2018). Forecasting hotspots in East Kutai, Kutai Kartanegara, and West Kutai as early warning information. *E&ES*, 144(1), 012022.
- [9]. Nallakaruppan, M. K., & Ilango, H. S. (2017, February). Location Aware Climate Sensing and Real Time Data Analysis. In *2017 World Congress on Computing and Communication Technologies (WCCCT)* (pp. 73-79). IEEE.

- [10]. Robby, I. S. (2014). Prediksi temporal untuk kemunculan titik panas di provinsi riau menggunakan autoregressive integrated moving average (arima) isnan syaiful robby. *Bogor (ID): Institut Pertanian Bogor*.
- [11]. Alsharif, M. H., Younes, M. K., & Kim, J. (2019). Time series ARIMA model for prediction of daily and monthly average global solar radiation: The case study of Seoul, South Korea. *Symmetry*, 11(2), 240.
- [12]. S. Ahmad and H. A. Latif. (2011). Forecasting on the Crude Palm Oil and Kernel Palm Production: Seasonal ARIMA Approach, presented at the IEEE Colloquium on Humanities, Science and Engineering Research, Penang.
- [13]. Bakar, N. A., & Rosbi, S. (2017). Autoregressive integrated moving average (ARIMA) model for forecasting cryptocurrency exchange rate in high volatility environment: A new insight of bitcoin transaction. *International Journal of Advanced Engineering Research and Science*, 4(11), 130-137.
- [14]. Farhath, Z. A., Arputhamary, B., & Arockiam, L. (2016). A Survey on ARIMA Forecasting Using Time Series Model. *Int. J. Comput. Sci. Mobile Comput*, 5, 104-109.
- [15]. Shrivastav, A. K. (2012). Ekata, Applicability of Box Jenkins ARIMA Model in Crime Forecasting: A case study of counterfeiting in Gujarat State. *Int. J. Advanced Res. Comput. Eng. Technol*, 1(4), 494-497.
- [16]. Williams, B. M., & Hoel, L. A. (2003). Modeling and forecasting vehicular traffic flow as a seasonal ARIMA process: Theoretical basis and empirical results. *Journal of transportation engineering*, 129(6), 664-672.
- [17]. Kumar, M., & Anand, M. (2014). An application of time series ARIMA forecasting model for predicting sugarcane production in India. *Studies in Business and Economics*, 9(1), 81-94.
- [18]. Almasarweh, M., & Alwadi, S. (2018). ARIMA model in predicting banking stock market data. *Modern Applied Science*, 12(11).
- [19]. Zhang, G. P. (2003). Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing*, 50, 159-175.
- [20]. Mondal, P., Shit, L., & Goswami, S. (2014). Study of effectiveness of time series modeling (ARIMA) in forecasting stock prices. *International Journal of Computer Science, Engineering and Applications*, 4(2), 13.
- [21]. Yang, X. (2019, January). The Prediction of Gold Price Using ARIMA Model. In *2nd International Conference on Social Science, Public Health and Education (SSPHE 2018)*. Atlantis Press.
- [22]. Shair, S. N., Zolkifi, N. A., Zulkefi, N. F., & Murad, A. (2019). A Functional Data Approach to the Estimation of Mortality and Life Expectancy at Birth in Developing Countries. *Pertanika Journal of Science & Technology*, 27(2).

SAR Journal Artikel

ORIGINALITY REPORT

24%

SIMILARITY INDEX

24%

INTERNET SOURCES

13%

PUBLICATIONS

9%

STUDENT PAPERS

PRIMARY SOURCES

1	www.researchgate.net Internet Source	4%
2	www.termpaperwarehouse.com Internet Source	4%
3	paper.researchbib.com Internet Source	3%
4	Submitted to Universitas Gunadarma Student Paper	3%
5	doaj.org Internet Source	2%
6	S. Sarifah Radiah Shariff, Mohd Azuan Suhaimi, Siti Meriam Zahari, Zuraidah Derasit. "Alternative Methods for Forecasting Variations in Hospital Bed Admission", Indonesian Journal of Electrical Engineering and Computer Science, 2018 Publication	1%
7	www.authorstream.com Internet Source	1%

8	Submitted to Asia Pacific University College of Technology and Innovation (UCTI) Student Paper	1 %
9	core.ac.uk Internet Source	1 %
10	hrmars.com Internet Source	1 %
11	Evizal Abdul Kadir, Mahmud Othman, Sri Listia Rosa. "Smart Sensor System for Detection and Forecasting Forest Fire Hotspot in Riau Province Indonesia", 2021 International Congress of Advanced Technology and Engineering (ICOTEN), 2021 Publication	1 %
12	Submitted to Universiti Teknologi MARA Student Paper	1 %
13	Submitted to Universiti Teknologi Malaysia Student Paper	1 %
14	Rayung Wulan, Achmad Sarwandianto, Nur Alamsyah, Aulia Ar Rakhman Awaludin. "Android-Based Expert System Application to Determine Math Learning Styles of Elementary School Students in Surakarta", SAR Journal - Science and Research, 2020 Publication	1 %
15	dokumen.site Internet Source	

Exclude quotes On

Exclude matches < 1%

Exclude bibliography On