RESEARCH ARTICLE • OPEN ACCESS

Forecasting Fire Hotspots in Indonesia: A Comparative Performance Analysis of Sarima and Pulse Intervention Models

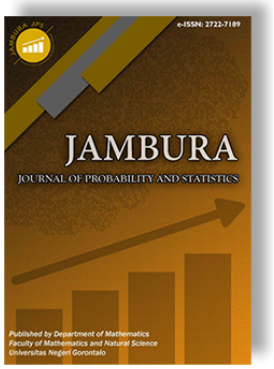
Bustami, Gustriza Erda, and Putri Soraya Tampubolon

Volume 7, Issue 1, Pages 1–9, May 2026

Received 08 April 2026, Revised 07 Mei 2026, Accepted 1 Juni 2026, Published Online 30 Mei 2026

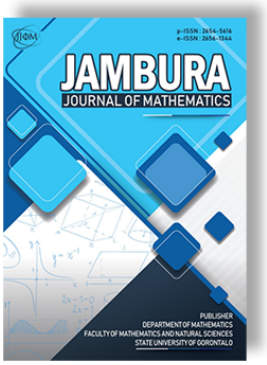
To Cite this Article : Bustami, G. Erda, and P. S. Tampubolon ,“ Forecasting Fire Hotspots in Indonesia: A Comparative Performance Analysis of Sarima and Pulse Intervention Models ”, *Jambura J. Probab. Stat.*, vol. 7, no. 1, pp. 1–9, 2026, *https://doi.org/10.34312/jjps.v7i2.38133*

JOURNAL INFO • JAMBURA JOURNAL OF PROBABILITY AND STATISTICS



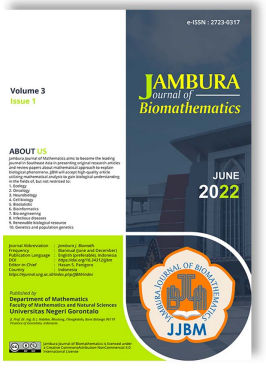
Homepage	:	https://ejurnal.ung.ac.id/index.php/jps/index
Journal Abbreviation	:	Jambura J. Probab. Stat.
Frequency	:	Biannual (May and November)
Publication Language	:	English (preferable), Indonesia
DOI	:	https://doi.org/10.34312/jjps
Online ISSN	:	2722-7189
Editor-in-Chief	:	Ismail Djakaria
Publisher	:	Department of Mathematics, Universitas Negeri Gorontalo
Country	:	Indonesia
OAI Address	:	http://ejurnal.ung.ac.id/index.php/jps/oai
Google Scholar ID	:	kWdujzMAAAAJ
Email	:	redaksi.jjps@ung.ac.id

JAMBURA JOURNAL • FIND OUR OTHER JOURNALS

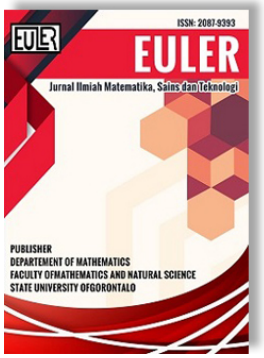


Jambura Journal of Mathematics

Jambura Journal of Mathematics Education



Jambura Journal of Biomathematics



EULER : Jurnal Ilmiah Matematika, Sains, dan Teknologi

Jambura Journal of Probability and Statistics, Volume 7, Issue 1, Pages 1–9, May 2026 https://doi.org/10.34312/jjps.v7i2.38133

Research Article

Check for updates

Forecasting Fire Hotspots in Indonesia: A Comparative Performance Analysis of Sarima and Pulse Intervention Models

Bustami[1], **Gustriza Erda**[1], **Putri Soraya Tampubolon**[1]

[1] *Study Program of Statistics, Department of Mathematics, Faculty of Mathematics and Natural Science, Universitas Riau, Indonesia*

ARTICLE HISTORY
Received 08 April 2026
Revised 07 Mei 2026
Accepted 1 Juni 2026
Published 30 Mei 2026

KEYWORDS
forecasting
hotspot
intervention analysis
pulse intervention
Density Method.

ABSTRACT. *Wildfires in Indonesia have a widespread impact on health, the environment, society, and the economy. The number of hotspots, detected through satellite imagery, is a key indicator in monitoring the severity of fires. Because hotspot data is seasonal and prone to spikes due to extraordinary events such as El Niño, an adaptive forecasting method is needed. The SARIMA model is effective for capturing seasonal patterns, but it is less responsive to extreme spikes. Therefore, intervention analysis with pulse functions is used as an alternative to model sudden and temporary changes in time series data. This study aims to compare the performance of the SARIMA model and an intervention model using a pulse function in forecasting the number of hotspots in Indonesia. The data used in this study were obtained from the Ministry of Environment and Forestry through the SiPongi platform, consisting of monthly data from January 2014 to December 2022. The modeling results show that the SARIMA model produced a MAPE value of 36.93%, an RMSE of 66.27, and an MAE of 47.83. In contrast, the intervention model with a pulse function at order $b = 0$, $s = 0$, and $r = 1$ achieved a MAPE of 8.06%, an RMSE of 8.45, and an MAE of 6.67, substantially outperforming the SARIMA model across all metrics. These findings indicate that the intervention model provides much more accurate forecasts of hotspot occurrences in Indonesia. Furthermore, forecasts up to 2025 indicate a declining trend in the number of hotspots over time. However, seasonal patterns remain evident, with expected increases in hotspot activity during the months of February, August, and October. These results are expected to contribute valuable insights for developing more effective forest fire mitigation strategies in Indonesia.*

1. Intoduction

Indonesia possesses one of the world's largest tropical forest ecosystems, earning it recognition as "The Lungs of the Earth" due to its critical role in global carbon sequestration and biodiversity conservation. However, Indonesia's forest area continues to experience significant deforestation, largely driven by land clearing practices that employ burning for agricultural expansion and plantation development [1]. Forest and land fires are not only triggered by human activities but also by natural conditions including the El Niño phenomenon [2],[3] which leads to prolonged dry seasons due to increased sea surface temperatures and reduced rainfall [4],[5]. This phenomenon makes Indonesia vulnerable to forest fires [6], especially in areas with extensive peat cover [7] and during the dry season [8]. Forest and land fires are a global environmental problem. Beyond causing ecological destruction and severe peatland degradation [9], forest fires contribute to increased carbon dioxide (CO_2) emissions [10], public health problems [11], and a decline in social and economic well-being [12]. The severity of this issue in Indonesia is evidenced by the 2024 National Disaster Management Agency (BNPB) report, which documented 1,478 forest fire incidents that caused 363 fatalities and 783 injuries, highlighting the immediate human cost of this environmental disaster. Furthermore, the 2023 Global Forest Watch report revealed that Indonesia lost 292,000 hectares of primary forest in a single year, equivalent to 218 megatonnes of CO_2 emissions—approximately 0.4% of global annual greenhouse gas emissions.

For forest fire prevention and control, the Indonesian Ministry of Environment and Forestry (KLHK) operates a satellite-based monitoring system that detects thermal anomalies or "hotspots", geographic locations exhibiting significantly elevated surface temperatures (typically $> 50°$C above ambient) that indicate active fires or high fire risk. Hotspot data is seasonal time series data because it shows recurring patterns on a monthly basis [13]. Time series forecasting methods have been increasingly applied to environmental and atmospheric data in Indonesia, including air pollution monitoring [14]. A commonly used forecasting model for seasonal data is *Seasonal Autoregressive Integrated Moving Average* (SARIMA), which is capable of capturing seasonal patterns in time series data. However, SARIMA models exhibit limitations when confronted with time series containing extreme outliers or structural breaks caused by exogenous events such as El Niño-induced drought conditions, policy interventions, or catastrophic natural disasters. These models assume stationarity and may fail to accurately capture sudden, transient shocks that deviate substantially from established patterns. To address this limitation, intervention analysis incorporating pulse functions provides a robust framework for explicitly modeling

*Corresponding Author.

Email : *gustrizaerda@lecturer.unri.ac.id* (G.Erda)
Homepage : https://ejurnal.ung.ac.id/index.php/jps/index / E-ISSN : 2722-7189

sudden, short-duration shocks in time series data. Pulse interventions quantify the magnitude and duration of transient impacts, enabling more accurate forecasting in the presence of exceptional events [15].

Intervention analysis is a time series methodology designed to quantify and model the effects of specific exogenous or endogenous events on temporal data patterns. Two primary intervention functions are distinguished based on impact duration: step functions for permanent regime changes and pulse functions for transient shocks. A step intervention models a long-term or permanent change in the time series level and is suitable for policy changes, corporate decisions, regime shifts, or travel warnings. In contrast, a pulse intervention captures temporary shocks or short-term disruptions, such as natural disasters, bombings, wars, promotional discounts, or protests [15],[16]

The use of intervention models has been widely applied in various fields, such as the research by [17] assessing the impact of policies to reduce antipsychotic drug doses in Australia, and by [18] analyzing the impact of rebellions on tourist arrivals in Thailand. In Indonesia, [19] used a pulse function intervention model to model hotspot spikes in East Kalimantan. While this provincial-level study demonstrated that pulse intervention modeling can be applied to Indonesian hotspot data, several important limitations remain. The focus on a single province reduces the relevance of the findings for national policymaking, since forest fire dynamics in Indonesia vary considerably across provinces due to differences in land cover, peatland extent, and regional climate conditions. Previous studies have also not examined whether the intervention modeling approach can be effectively applied to aggregated national hotspot data, where fire events from multiple provinces interact and may create more complex time series patterns than those observed at the provincial level. In addition, there has been no comprehensive comparison between SARIMA and pulse intervention models using rigorous out-of-sample forecasting evaluation at the national scale. Addressing these gaps is essential, particularly because Indonesia requires accurate national early warning systems to support evidence-based forest fire mitigation strategies and improve centralized decision-making for resource allocation.

This study addresses this research gap by comparing the forecasting performance of standard SARIMA and pulse intervention models for national-scale monthly hotspot forecasting in Indonesia. The novelty of this research lies in three aspects: (1) extending the intervention modeling framework from provincial to national scale, thereby capturing aggregate fire dynamics across Indonesia's diverse ecosystems; (2) providing rigorous model validation using out-of-sample testing on 2022 data with multiple error metrics; and (3) generating multi-year forecasts (2023–2025) to support proactive national fire prevention planning. Specifically, the research objectives are to: (1) develop and validate optimal SARIMA models for pre-intervention hotspot data, (2) incorporate pulse intervention functions to model the September 2019 El Ni no event, (3) quantitatively compare model performance using MAPE, RMSE, and MAE as evaluation metrics, and (4) generate forecasts for January 2023 to December 2025 to support proactive fire prevention strategies.

2. RESEARCH METHOD

2.1. Data Sources

This study utilized secondary data on monthly hotspot occurrences obtained from the Indonesian Ministry of Environment and Forestry's SiPongi (*Sistem Informasi Pemantauan Kebakaran Hutan dan Lahan*) Monitoring System (accessible at `https://sipongi.menlhk.go.id`). The dataset comprised 108 monthly observations spanning January 2014 to December 2022. Hotspots were detected using MODIS (*Moderate Resolution Imaging Spectroradiometer*) and VIIRS (*Visible Infrared Imaging Radiometer Suite*) satellite sensors with a confidence level threshold of $\geq 80\%$.

2.2. Data Partitioning

The dataset was partitioned into training and testing subsets to enable model validation. The training set consisted of 96 observations (January 2014 to December 2021), which was used for all model estimation stages: (1) the SARIMA baseline model was estimated on pre-intervention data (January 2014 to August 2019, $t = 1$ to 68) to identify the noise model structure, and (2) the complete intervention model, incorporating both the SARIMA noise component and the pulse intervention parameters, was fitted using the full training data spanning January 2014 to December 2021 (all 96 observations), ensuring that the intervention effects at $t = 69$ and the post-intervention dynamics were captured in the final model estimation. The testing set comprised 12 observations (January 2022 to December 2022), reserved exclusively for out-of-sample forecast evaluation using MAPE, RMSE, and MAE metrics. This partitioning strategy ensures that model fitting leverages the complete available training history while maintaining a hold-out period for unbiased performance assessment.

2.3. Seasonal Autoregressive Integrated Moving Average (SARIMA) Model

The *Seasonal Autoregressive Integrated Moving Average* (SARIMA) model extends the classical ARIMA framework by incorporating seasonal components. This model is denoted as SARIMA$(p, d, q)(P, D, Q)^S$, where lowercase letters represent non-seasonal orders and uppercase letters denote seasonal orders, with S indicating the seasonal period ($S = 12$ for monthly data). The general SARIMA equation is [20]:

$$Z_t = \frac{\theta_q(B)\Theta_Q(B^S)a_t}{\phi_p(B)\Phi_P(B^S)(1-B)^d(1-B^S)^D}$$

where ϕ_p and θ_q are the non-seasonal AR and MA parameters of orders p and q, respectively; Φ_P and Θ_Q are the seasonal AR and MA parameters of orders P and Q, respectively; $(1-B)^d$ and $(1-B^S)^D$ denote the non-seasonal and seasonal differencing operators of orders d and D; a_t is a white-noise error term; and B is the backshift operator.

The SARIMA modeling procedure follows the Box–Jenkins methodology and begins with stationarity diagnostics. Variance stationarity is assessed through *Box–Cox* transformation analysis, where λ values significantly different from 1 indicate the need for variance stabilization. Mean stationarity is evaluated using the *Augmented Dickey–Fuller* (ADF) unit root test, where rejection

of the null hypothesis ($p < 0.05$) confirms stationarity. Non-stationary series undergo differencing transformations until the stationarity criteria are satisfied. If the data are stationary, the model order is identified by analyzing the *Autocorrelation Function* (ACF) and *Partial Autocorrelation Function* (PACF) plots. The seasonal and non-seasonal orders are selected based on the correspondence between the empirical patterns observed in the ACF and PACF plots and the theoretical patterns of the SARIMA components.

After the initial model has been identified, the next step is parameter estimation. Subsequently, parameter significance testing is conducted to determine whether the estimated coefficients are statistically significant. Model diagnostics are then performed using the *Kolmogorov–Smirnov* test to assess residual normality and the *Ljung–Box* test to evaluate whether the residuals constitute a random process (*white noise*).

2.4. Intervention Model

Time series data can be influenced by specific events or occurrences, such as government policies, advertising promotions, natural disasters, human behavior, and other external factors known as interventions. Changes in the patterns of time series data may arise due to such interventions, and their effects can be analyzed using intervention models. The objective of intervention analysis is to determine the magnitude and duration of the intervention effects on the observed time series [15].

Intervention analysis incorporates two types of intervention functions, namely the *step function* and the *pulse function*. These functions are distinguished based on the duration of the intervention effect. The *step function* represents an intervention that begins at time T and produces a sustained impact over a relatively long period. In contrast, the *pulse function* represents an intervention that occurs only at time T and does not persist in subsequent periods [21].

This study employs the *pulse intervention function* rather than the *step function* due to the transient nature of fire events and their impact on hotspot time series data. Although forest fires may generate long-term ecological consequences, the hotspot count response typically exhibits a sharp and temporary increase during the fire event and subsequently returns to its normal pattern. Therefore, the *pulse function* is considered more appropriate for capturing these short-duration anomalies than the *step function*, which is designed to model permanent shifts in the level of a time series.

The pulse intervention function is denoted by $P_t^{(T)}$ and is defined as

$$P_t^{(T)} = \begin{cases} 1, & t = T, \\ 0, & t \neq T, \end{cases}$$

Based on figure 1 , there are two possible responses to an intervention, namely direct permanent intervention and gradual permanent intervention. A direct intervention occurs when the intervention effect begins immediately at time $t = T$ and remains permanent thereafter. In contrast, a gradual permanent intervention occurs when the intervention effect starts at time $t = T$ and increases progressively over time, while its impact remains permanent.

The general form of the intervention model is given by

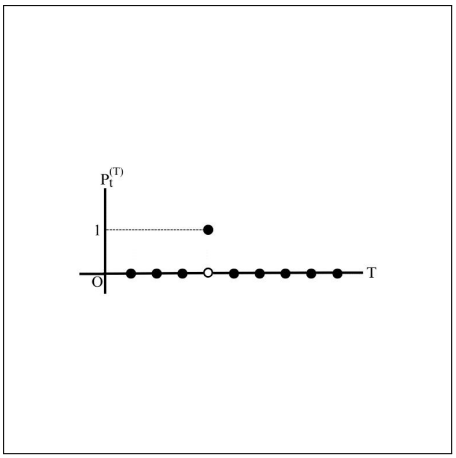


Figure 1. The Pulse Function in Intervention Analysis

$$Z_t = \frac{\omega_s(B)B^b}{\delta_r(B)} I_t + N_t$$

where:

- I_t : intervention variable in the form of a step or pulse function,
- $\omega_s(B)$: moving average operator of order s,
- $\delta_r(B)$: autoregressive operator of order r,
- b : delay time before the intervention effect appears,
- s : duration of the intervention effect on the series after time b,
- r : pattern of the intervention effect after orders b and s,
- N_t : noise model.

2.5. Analysis Method

The steps undertaken in this study are as follows:

1. Describing and grouping data based on the time of intervention
2. Model the data prior to intervention using the ARIMA modeling procedure
   (a) Testing stationarity in variance using the *Box–Cox* transformation, assessed through the rounded value of λ. If $\lambda \neq 1$, a transformation is required. Stationarity in mean is examined using the *Augmented Dickey–Fuller* (ADF) test with the following hypotheses:

   H_0 : The series contains a unit root (non-stationary)

   H_1 : The series is stationary

   The null hypothesis is rejected when the ADF test statistic is less than the critical value at $\alpha = 0.05$.
   (b) Identifying the SARIMA model order (p, d, q, P, D, Q) through the ACF and PACF plots of the differenced series. The non-seasonal MA order q is identified from the cutoff lag in the ACF, while the non-seasonal AR order p is identified from the cutoff lag in the PACF. The seasonal orders P and Q are determined from significant lags at multiples of the seasonal period $s = 12$. All candidate models are subsequently evaluated for stationarity and invertibility.

(c) Performing parameter significance tests using the t-test (H_0: parameter $= 0$, rejected if $|t_0| > t_{0.025}$) and verifying model adequacy through several diagnostic checks: (a) The Ljung–Box test is used to assess residual independence with the null hypothesis that the residuals are *white noise*, which is not rejected if $Q < \chi^2_{5\%}$, (b) the Kolmogorov-Smirnov test for residual normality with the null hypothesis that the residuals are normally distributed, which is not rejected if $D_0 < D_\alpha$, (c) residual ACF plot to visually confirm absence of autocorrelation, (d) residual histogram to assess approximate normality, and (e) Q-Q plot to evaluate the normality of residuals graphically.

3. Forecasting the post-intervention period using the selected SARIMA intervention model. The intervention orders (b, s, r) are determined based on the residual response plot. Forecast performance is evaluated using the following accuracy measures:

$$\text{MAPE} = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{Y_t - \hat{Y}_t}{Y_t}\right| \times 100\%$$

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{t=1}^{n}\left(Y_t - \hat{Y}_t\right)^2}$$

$$\text{MAE} = \frac{1}{n}\sum_{t=1}^{n}\left|Y_t - \hat{Y}_t\right|$$

where Y_t denotes the actual observation and $\hat{Y}_t$ denotes the forecasted value at time t.

3. RESULT AND DISCUSSION

3.1. Data Exploration

Data on the number of hotspots from January 2014 to December 2021, totaling 96 data points, was used in the modeling process. Meanwhile, data on the number of hotspots from January 2022 to December 2022 was used to evaluate the accuracy of the model.

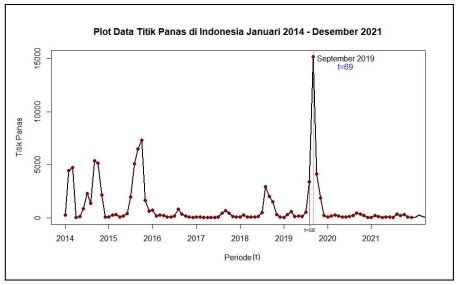


Figure 2. The Number of Hotspots from January 2014 to December 2021

Figure 2 reveals a pronounced spike in hotspot occurrences during September 2019, with counts reaching approximately 15,000. This extreme anomaly coincided with a strong El Niño event characterized by the *Indian Ocean Dipole* (IOD) in its positive phase, which induced severe drought conditions across the Indonesian archipelago, particularly in Sumatra and Kalimantan. The identification of $t = 69$ was determined through visual inspection of the time series plot and corroborating climatological records. While formal structural break tests such as the Chow test could further validate this, the use of domain knowledge and graphical evidence for intervention point identification is a well-established practice in intervention analysis, as noted by Wei [15]. The magnitude of the spike exceeding 15,000 hotspots compared to the typical range of fewer than 500 provides unambiguous empirical evidence for this breakpoint.

In data modeling that considers interventions, the dataset is conceptually divided into two parts. First, pre-intervention data at $t < 69$ covering the period from January 2014 to August 2019, which is used to build the baseline SARIMA noise model. Second, post-intervention data at $t \geq 69$ which is used to identify the intervention order (b, s, r) and evaluate the pattern of its impact. It is important to note that while the data is conceptually divided for noise model identification and intervention order determination, the final intervention model parameters are estimated using the complete training dataset (January 2014 to December 2021, $n = 96$ observations). This ensures that all available information, including both pre- and post-intervention dynamics, is utilized in fitting the intervention model.

3.2. SARIMA modelling

The initial stage in SARIMA modeling is to check the stationarity of the data in variance through the Box–Cox plot and in mean through the ACF plot. Figure 3 shows that the $\lambda \neq 1$ indicating that a Box–Cox transformation is necessary to improve the data distribution and stabilize the variance. Then, a Box–Cox transformation is performed with $\lambda = -0.135886$ so that the data is stationary in variance.

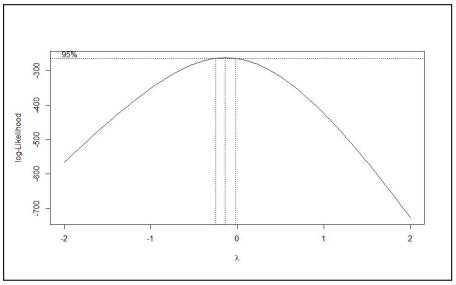


Figure 3. Box-Cox Plot for the Number of Hotspots from January 2014 to December 2021

The next step is to examine the stationarity of the data with respect to the mean using an ACF (*Autocorrelation Function*) plot. Based on Figure 4, the ACF plot shows that the data is not stationary with respect to the mean. This is indicated by the lines on the correlogram that still cross the confidence interval and form a sinusoidal pattern that repeats every 12 lags. This pattern indicates that the data has a seasonal component. To address this non-stationarity, a non-seasonal first-order differencing is performed, followed by a seasonal first-order differencing on the transformed data. This step aims to remove trends and seasonal patterns so that the data becomes stationary and can be used in further time series modeling. Data that is stationary in both variance and mean can then be used to identify the model.

Data that is stationary in variance and mean can then be used to identify the model through ACF and PACF plots in Figure

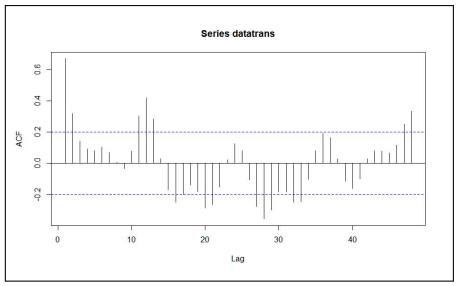


Figure 4. ACF Plot for Data on The Number of Hotspots from January 2014 to December 2021

5 . Based on Figure 5, a seasonal ARIMA model can be identified, which is then referred to as SARIMA. The ACF plot shows a cutoff correlation coefficient after lag 1 for non-seasonal data, while no cutoff occurs at seasonal lags. For visualization, the PACF plot shows a cutoff in the correlation coefficients after lag 2 for non-seasonal data and a cutoff after lag 12 for seasonal data.

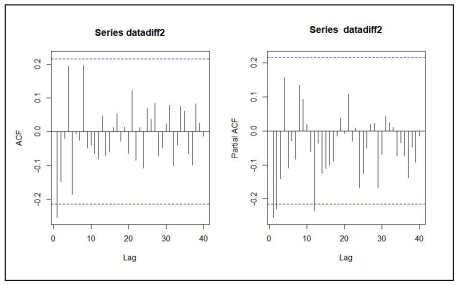


Figure 5. ACF and PACF Plot for The Number of Hotspot From January 2014 to December 2021 (Stationary Data)

The models obtained from the identification of the ACF and PACF plots are AR(1), AR(2), MA(1), and SAR(1) models with one-step differencing for non-seasonal and seasonal data, as summarized in Table 1.

The formed model was tested for stationarity using the *Augmented Dickey–Fuller* (ADF) test. At a significance level of 5%, all models show test statistic values smaller than the critical table value, so the null hypothesis is rejected. This indicates that all models come from a stationary process. From the ten SARIMA models developed, the model that meets the diagnostic criteria will be selected through parameter significance tests and other diagnostic tests. Based on the test results, it was found that the SARIMA $(0,1,1)(0,1,0)_{12}$, SARIMA $(1,1,0)(0,1,0)_{12}$, and SARIMA $(2,1,0)(0,1,0)_{12}$ models meet the parameter significance test, as the value of $|t_{\text{hit}}|$ for each parameter in the three models is greater than the value of t_{tabel}. This indicates that all parameters used in the three models are statistically significant.

Residual diagnostics were conducted at $\alpha = 0.05$ significance level to verify model adequacy. Formal statistical tests included the Kolmogorov–Smirnov test for residual normality and the Ljung–Box Q-statistic for residual independence (white noise property). Visual diagnostic tools such as residual ACF plots, residual histograms, and Q–Q plots were not produced in this study; however, the formal Kolmogorov–Smirnov and Ljung–Box tests are statistically sufficient to confirm residual normality and independence. Future studies are encouraged to supplement these formal tests with visual diagnostics for a more comprehensive assessment of model adequacy.

Results presented in Table 2 indicate that SARIMA $(0,1,1)(0,1,0)_{12}$ uniquely satisfied both criteria (p-values > 0.05), confirming normally distributed, independently distributed residuals. This model was therefore selected as the optimal pre-intervention specification. The SARIMA $(0,1,1)(0,1,0)_{12}$ model formed is:

$$Z_t = \frac{\theta_1(B)a_t}{(1-B)(1-B^{12})}$$

and obtained:

$$Z_t = Z_{t-1} + Z_{t-12} - Z_{t-13} + a_t + 0.38526\, a_{t-1}$$

3.3. *Modeling with Intervention Methods*

The process of forming a SARIMA model for data before intervention begins with checking the stationarity of the data, both in variance and in mean. The Box–Cox visualization results in Figure 6 show that the value of the transformation parameter λ is not equal to 1, indicating that the data does not meet the assumption of stationarity in variance. Therefore, a transformation is carried out using the Box–Cox function with the parameter $\lambda = -0.08659638$. This transformation aims to stabilize the variance, and after it is done, the data is declared stationary in variance.

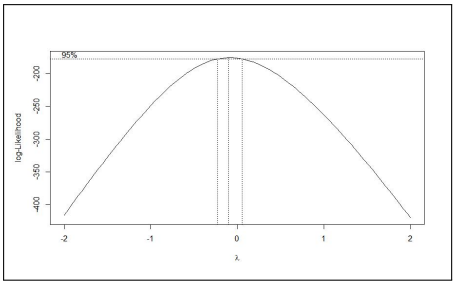


Figure 6. Box-Cox plot for Pre-Intervention Data

Stationarity checks on the average are performed through the Autocorrelation Function (ACF) plot shown in Figure 7. The plot shows that the correlation is still significant at several lags, and forms a sinusoidal pattern that repeats every 12 lags. This pattern indicates the presence of seasonal components and non-stationarity in the average and seasonality. To overcome this, non-seasonal first-order differencing is applied, followed by seasonal first-order differencing. The results of this transformation and differencing produce data that is stationary with respect to the average.

Table 1. Combination of SARIMA models

No.	Model	No.	Model
1	SARIMA $(0,1,1)(0,1,0)^{12}$	6	SARIMA $(0,1,1)(1,1,0)^{12}$
2	SARIMA $(1,1,0)(0,1,0)^{12}$	7	SARIMA $(1,1,0)(1,1,0)^{12}$
3	SARIMA $(1,1,1)(0,1,0)^{12}$	8	SARIMA $(1,1,1)(1,1,0)^{12}$
4	SARIMA $(2,1,0)(0,1,0)^{12}$	9	SARIMA $(2,1,0)(1,1,0)^{12}$
5	SARIMA $(2,1,1)(0,1,0)^{12}$	10	SARIMA $(2,1,1)(1,1,0)^{12}$

Table 2. Results of SARIMA Model Diagnostic Tests

Model	Diagnostic		Statistical test	Statistical Table	Conclusion
	Normality		0,12583	0,1377	Normality assumed
SARIMA		Lag 12	16,592	19,6751	No autocorrelation
$(0,1,1)(0,1,0)^{12}$	Independence	Lag 24	22,858	35,1725	No autocorrelation
		Lag 36	30,791	49,8019	No autocorrelation
	Normality		0,11981	0,1377	Normality assumed
SARIMA		Lag 12	20,598	19,6751	Autocorrelation detected
$(1,1,0)(0,1,0)^{12}$	Independence	Lag 24	25,897	35,1725	No autocorrelation
		Lag 36	35,556	49,8019	No autocorrelation
	Normality		0,14433	0,1377	Normality violated
SARIMA		Lag 12	16,252	19,6751	No autocorrelation
$(2,1,0)(0,1,0)^{12}$	Independence	Lag 24	23,755	35,1725	No autocorrelation
		Lag 36	30,34	49,8019	No autocorrelation

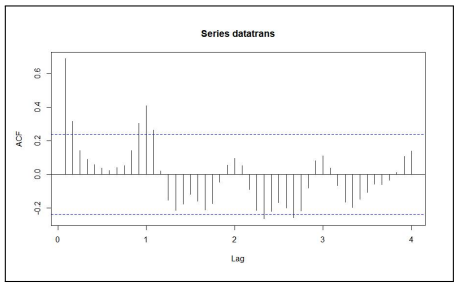


Figure 7. ACF Plot for Pre-Intervention Data

After the data becomes stationary, model identification is performed using ACF and Partial ACF (PACF) analysis as shown in Figure 8. The ACF plot shows a cutoff after the 2nd lag on the non-seasonal component, and a similar pattern is seen in the PACF. However, there is no cutoff on the seasonal lag. Based on this pattern and referring to Table 3, several candidate SARIMA models are proposed as follows:

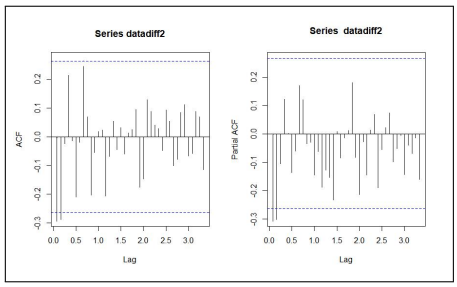


Figure 8. ACF and PACF Plot for pre-intervention data

A combination model was obtained between AR(1), AR(2), MA(1), MA(2) with one-time differencing for non-seasonal and seasonal and is summarized in Table 3.

Then, the stationarity test is performed using the *Augmented Dickey–Fuller* (ADF) test. At a significance level of 5%, all models show test statistic values smaller than the critical table value, so the null hypothesis is rejected. This indicates that all models come from a stationary process. Of the eight candidate models, parameter significance testing was carried out using the t-test with a significance level of 5%. The SARIMA $(2,1,2)(0,1,0)_{12}$ was the only model in which all parameters were statistically significant at the 5% significance level, with $|t\text{-hit}|$ values exceeding the critical value for each parameter. Next, a diagnostic evaluation of the model was carried out using two main tests, namely the Kolmogorov–Smirnov Test to test the normality of the residuals and the Ljung–Box Test to test the independence of the residuals at lags 12, 24, and 36, and the results are obtained in Table 4. Table 4 shows that the initial hypothesis cannot be rejected in the Kolmogorov–Smirnov test and the Ljung–Box test because the test statistic value is smaller than the table statistic value. It is concluded that the residuals for the model are normally distributed and independent. SARIMA $(2,1,2)(0,1,0)_{12}$ is then written as follows (noise model):

$$Z_t = \frac{\theta_2(B)a_t}{\phi_2(B)(1-B)(1-B^{12})}$$

After obtaining the noise model, the next step is to determine the order of b, s, and r through the analysis of the residual response diagram. This diagram is obtained from the difference between the actual data and the forecasted data using the noise model. Based on Figure 9, it can be seen that the residual value is outside the significance limit at the time $t = 69$. This indicates that the intervention has an order of $b = 0$ because the impact of the intervention is immediately visible when the intervention occurs, without any delay.

In addition, the value of the order $s = 0$ is obtained be-

Table 3. Combination of SARIMA Models on Pre-Intervention Data

No.	Model	No.	Model
1	SARIMA$(0,1,1)(0,1,0)^{12}$	5	SARIMA$(1,1,2)(0,1,0)^{12}$
2	SARIMA$(0,1,2)(0,1,0)^{12}$	6	SARIMA$(2,1,0)(0,1,0)^{12}$
3	SARIMA$(1,1,0)(0,1,0)^{12}$	7	SARIMA$(2,1,1)(0,1,0)^{12}$
4	SARIMA$(1,1,1)(0,1,0)^{12}$	8	SARIMA$(2,1,2)(0,1,0)^{12}$

Table 4. Diagnostic Test Results of the SARIMA Model

Diagnostic	Statistical Test		Statistical Table	Assumptions
Normality	$D_{hit} = 0,15168$		$D_{(0,05;68)} = 0.1637$	Normality assumed
Independency	Lag 12	$Q = 4,6302$	$\chi^2_{(0,05;6)} = 12.5916$	No autocorrelation
	Lag 24	$Q = 20,505$	$\chi^2_{(0,05;18)} = 28.8693$	No autocorrelation
	Lag 36	$Q = 29,593$	$\chi^2_{(0,05;30)} = 43.7730$	No autocorrelation

cause the residual is stable again, which is below the significance limit, immediately after the intervention time. From the diagram, it can also be determined that the order $r = 1$ because one period after the intervention there appears to be a decline process that begins to form a linear pattern.

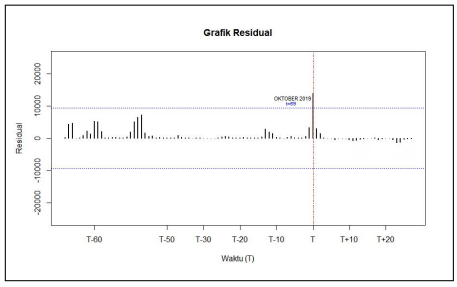


Figure 9. Residual Response Diagram

Based on the data pattern of the number of hotspots, the function used is pulse because the condition when the intervention occurs directly at time $t = 69$ and its influence is reduced. Therefore, the intervention model formed is:

$$Z_t = \frac{\omega_0}{1-\delta_1 B} P_t^{(69)} + \frac{\theta_2(B)}{\phi_2(B)(1-B)(1-B^{12})}\alpha_t$$

where $P_t^{(69)}$ will have a value of 0 when $t \neq 69$ and a value of 1 when $t = 69$.

Based on the parameter significance test and model diagnostics, it was obtained that the SARIMA$(2,1,2)(0,1,0)_{12}$ pulse function intervention model with order $b = 0$, $s = 0$, and $r = 1$ has significant parameters and meets the model diagnostics. After obtaining the noise model and the intervention order is known, the next step is to conduct a significance test as explained in Table 5.

Table 5 shows that all intervention parameters are significant because the $|t_{\text{hit}}|$ value for each parameter is greater than $t_{(0.025,90)} = 1.9822$. After conducting a diagnostic examination of the intervention model through the Kolmogorov–Smirnov test and the Ljung–Box test, the results of which are summarized in Table 6. It is concluded that the intervention model meets the requirements for residual normality and all lags are independent. The fulfillment of this assumption indicates that the intervention

Table 5. Significance Testing of Intervention Models

Parameter	t_{hit}	$t_{(0,025;90)}$	Decision
$\hat{\phi}_1 = 0,195666$	5,1434	1,9822	Significant
$\hat{\phi}_2 = -0,947343$	$-23,4329$		Significant
$\hat{\theta}_1 = -0,361393$	$-8,3405$		Significant
$\hat{\theta}_2 = 0,999984$	13,4724		Significant
$\widehat{\omega}_0 = -0,028275$	$-2,3110$		Significant
$\hat{\delta}_1 = -0,865029$	$-9,4557$		Significant

model is suitable for use in predicting the emergence of hotspots in Indonesia.

3.4. *Comparison of Accuracy*

To evaluate the forecasting performance of the SARIMA model and the intervention model, three complementary error metrics are used: MAPE (Mean Absolute Percentage Error), RMSE (Root Mean Square Error), and MAE (Mean Absolute Error). MAPE measures the average relative percentage deviation, RMSE penalizes larger errors more heavily due to the squared term and is expressed in the same unit as the data, while MAE provides a straightforward average of absolute deviations. Together, these metrics provide a more robust and comprehensive assessment of forecast accuracy than any single metric alone. All metrics are calculated over the testing period from January 2022 to December 2022. Lower values for all metrics indicate higher forecasting accuracy. The results are presented in Table 7. Based on the results, the intervention model demonstrates superior performance across all three metrics: a MAPE of 8.06% (compared to 36.93% for SARIMA), a substantially lower RMSE (8.45), and a lower MAE (6.67). This consistently superior performance across multiple metrics confirms that the intervention model provides more accurate forecasts, particularly in capturing extreme variations in the number of hotspots in Indonesia during the observed period. The relatively high MAPE of the SARIMA model is partly attributable to its sensitivity to small actual values (e.g., January with only 23 hotspots), where even moderate absolute errors produce large percentage deviations, further validating the utility of reporting RMSE and MAE alongside MAPE.

Then, a model is formed based on the identification results as a SARIMA intervention model SARIMA$(2,1,2)(0,1,0)^{12}$ with

Table 6. Diagnostic test of Intervention Model

Diagnostic	Statistical Test	Statistical Table	Assumptions
Normality	$D_{hit} = 0,10631$	$D_{(0,05;96)} = 0,1378$	Normality assumed
Independency	Lag 12 $Q = 9,4992$	No autocorrelation	No autocorrelation
	Lag 24 $Q = 15,651$	No autocorrelation	No autocorrelation
	Lag 36 $Q = 20,485$	No autocorrelation	No autocorrelation

Table 7. Comparison of Actual Data and Forecasts from SARIMA and Intervention Models

Period in 2022	Actual Data	SARIMA Model Forecast	Intervention Model Forecast
January	23	13	18
February	283	136	292
March	130	85	126
April	35	27	31
May	56	37	51
June	121	60	114
July	16	10	15
August	251	187	248
September	169	131	179
October	307	172	329
November	68	38	63
December	34	23	29
MAPE		36,93%	8,06%
RMSE		66,27	8,45
MAE		47,83	6,67

intervention parameters $b = 0$, $s = 0$, $r = 1$ as follows:

$$\begin{aligned}
Z_t = {} & Z_{t-1} + Z_{t-12} - Z_{t-13} + 0.195666Z_{t-1} - 0.195666Z_{t-2} \\
& - 0.195666Z_{t-13} + 0.195666Z_{t-14} - 0.947343Z_{t-2} + 0.947343Z_{t-3} \\
& + 0.947343Z_{t-14} - 0.947343Z_{t-15} - 0.865029Z_{t-1} + 0.865029Z_{t-2} \\
& + 0.865029Z_{t-13} - 0.865029Z_{t-14} + 0.169257Z_{t-2} - 0.169257Z_{t-3} \\
& - 0.169257Z_{t-14} + 0.169257Z_{t-15} - 0.819479Z_{t-3} + 0.819479Z_{t-4} \\
& + 0.819479Z_{t-15} - 0.819479Z_{t-16} - 0.028275P_t^{(69)} \\
& + 0.028275P_{t-1}^{(69)} + 0.028275P_{t-12}^{(69)} - 0.028275P_{t-13}^{(69)} \\
& + 0.005532P_{t-1}^{(69)} - 0.005532P_{t-2}^{(69)} + 0.005532P_{t-13}^{(69)} \\
& - 0.005532P_{t-14}^{(69)} - 0.026786P_{t-2}^{(69)} + 0.026786P_{t-3}^{(69)} \\
& + 0.026786P_{t-14}^{(69)} - 0.026786P_{t-15}^{(69)} + 0.361393a_{t-1} \\
& - 0.999984a_{t-2} - 0.312615a_{t-2} + 0.865015a_{t-3} + a_t
\end{aligned}$$

3.5. *Forecasting*

The results of forecasting the number of hotspots in Indonesia for 2023 to 2025 using the intervention model are explained in Table 8. Seasonal patterns remain clear, with significant spikes occurring in February, August, and October. Research [22] shows that forest fires in Kalimantan and Sumatra tend to increase in August and October due to extreme dry conditions. Meanwhile, the spike in February in this model reflects the delayed impact of drought in the previous period. Several studies also show that fire activity in early years can be triggered by dry conditions that continue from the El Niño phase in the previous year [23].

Table 8. Forecasting Result using Intervention Models

Month	Year 2023	Year 2024	Year 2025
January	14	12	12
February	292	242	201
March	137	137	120
April	28	28	28
May	40	35	34
June	106	88	76
July	16	15	13
August	233	240	230
September	140	127	125
October	284	231	205
November	66	60	52
December	29	29	28

4. CONCLUSION

The results of the study show that the intervention model with the pulse function has better forecasting performance compared to the SARIMA model in predicting the number of hotspots in Indonesia. Performance evaluation using three complementary metrics shows that the intervention model achieves a MAPE of 8.06% (vs. 36.93%), an RMSE of 8.45 (vs. 66.27), and an MAE of 6.67 (vs. 47.83), consistently outperforming the SARIMA model across all metrics. This indicates that the intervention model is more accurate, especially in capturing extreme changes that occur in a certain period. The forecasting results using the intervention model for the period 2023 to 2025 show a downward trend in the number of hotspots each year. However, seasonal patterns are still identified with a consistent increase in hotspots occurring in February, August, and October. The results of this

forecast can be used by the authorized government to address forest fire problems by monitoring the results of hotspot forecasts.

References

[1] R. Kurniawan, "Conservation of indonesian tropical forests as the lungs of the world," *Inkalindo Environmental Journal*, vol. 1, no. 1, pp. 62–66, 2020.

[2] W. Ma, Z. Feng, Z. Cheng, S. Chen, and F. Wang, "Identifying forest fire driving factors and related impacts in china using random forest algorithm," *Forests*, vol. 11, no. 5, p. 507, 2020.

[3] I. Albar, I. N. S. Jaya, B. H. Saharjo, and B. Kuncahyo, "Spatio-temporal typology of land and forest fire in sumatra," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 4, no. 1, pp. 83–90, 2016.

[4] Supari, F. Tangang, E. Salimun, E. Aldrian, A. Sopaheluwakan, and L. Juneng, "Enso modulation of seasonal rainfall and extremes in indonesia," *Climate Dynamics*, vol. 51, no. 7–8, pp. 2559–2580, 2018.

[5] R. D. Field *et al.*, "Indonesian fire activity and smoke pollution in 2015 show persistent nonlinear sensitivity to el niño-induced drought," *Proceedings of the National Academy of Sciences*, vol. 113, no. 33, pp. 9204–9209, 2016.

[6] S. Nurdiati, A. Sopaheluwakan, and P. Septiawan, "Spatial and temporal analysis of el niño impact on land and forest fire in kalimantan and sumatra," *Agromet*, vol. 35, no. 1, pp. 1–10, 2021.

[7] A. D. Nurhayati, B. H. Saharjo, L. Sundawati, Syartinilia, and M. A. Cochrane, "Forest and peatland fire dynamics in south sumatra province," *Forest and Society*, pp. 591–603, 2021.

[8] M. B. R. Prayoga and R. H. Koestoer, "Improving forest fire mitigation in indonesia: A lesson from canada," *Jurnal Wilayah dan Lingkungan*, vol. 9, no. 3, pp. 293–305, 2021.

[9] J. S. Sze, Jefferson, and J. S. H. Lee, "Evaluating the social and environmental factors behind the 2015 extreme fire event in sumatra, indonesia," *Environmental Research Letters*, vol. 14, no. 1, p. 015001, 2019.

[10] Copernicus, "Wildfires wreaked havoc in 2021, cams tracked their impact," 2021.

[11] D. M. Molina-Terrén *et al.*, "Analysis of forest fire fatalities in southern europe: Spain, portugal, greece and sardinia (italy)," *International Journal of Wildland Fire*, vol. 28, no. 2, pp. 85–98, 2019.

[12] J. Paudel, "Natural disasters and economic inequality: Insights from wildfires across the globe," UNU-WIDER, Tech. Rep., 2023.

[13] Endrawati, *Analisis Data Titik Panas (Hotspot) dan Areal Kebakaran Hutan dan Lahan Tahun 2016*. Jakarta: Direktorat Inventarisasi dan Pemantauan Sumber Daya Hutan, Ditjen Planologi Kehutanan dan Tata Lingkungan Kementerian Lingkungan Hidup dan Kehutanan, 2016.

[14] A. Adnan, G. Erda, Wamiliana, and E. Russel, "Modeling cointegrated nonstationary air pollution data: A forecasting study of no_2 and so_2 in indonesia (1950–2022)," *Science and Technology Indonesia*, vol. 11, no. 1, pp. 161–173, 2026.

[15] W. W. S. Wei, *Time Series Analysis: Univariate and Multivariate Methods*, 2nd ed. New York: Pearson Education, 2006.

[16] A. Ramadhani, S. Wahyuningsih, and M. Siringoringo, "Forecasting the number of foreign tourist visits to indonesia used intervention analysis with step function," *Jurnal Matematika, Statistika dan Komputasi*, vol. 19, no. 1, pp. 146–162, 2022.

[17] A. L. Schaffer, T. A. Dobbins, and S. A. Pearson, "Interrupted time series analysis using autoregressive integrated moving average (arima) models: A guide for evaluating large-scale health interventions," *BMC Medical Research Methodology*, vol. 21, no. 1, pp. 1–12, 2021.

[18] P. Chakorn and L. Jetsada, "The impact of insurgency in the deep south of thailand on the arrival of malaysian tourists to betong district, yala province using sarima with intervention model," *Kasetsart Journal of Social Sciences*, vol. 43, no. 1, pp. 81–87, 2022.

[19] A. R. Saputra, S. Wahyuningsih, and M. Siringoringo, "Peramalan jumlah titik panas provinsi kalimantan timur menggunakan analisis intervensi fungsi pulse," *Eksponensial*, vol. 12, no. 1, pp. 93–101, 2021.

[20] G. E. P. Box, G. M. Jenkins, and G. C. Reinsel, *Time Series Analysis: Forecasting and Control*, 4th ed. New Jersey: John Wiley and Sons, 2008.

[21] M. H. Lee, Suhartono, and B. Sanugi, "Multi input intervention model for evaluating the impact of the asian crisis and terrorist attacks on tourist arrivals," *Matematika*, vol. 26, no. 1, pp. 83–106, 2010.

[22] S. Nurdiati, A. Sopaheluwakan, and P. Septiawan, "Spatial and temporal analysis of el niño impact on land and forest fire in kalimantan and sumatra," *Agromet*, vol. 35, no. 1, pp. 1–10, 2021.

[23] S. Nurdiati, A. Sopaheluwakan, and P. Septiawan, "Joint distribution analysis of forest fires and precipitation in response to enso, iod, and mjo (study case: Sumatra, indonesia)," *Atmosphere*, vol. 13, no. 4, p. 537, 2022.