ISSN 2090-3359 (Print)
ISSN 2090-3367 (Online)

ΑΔΣ

# Advances in Decision Sciences

Volume 30
Issue 2
June 2026

亞洲大學
ASIA UNIVERSITY

SCIENTIFIC & BUSINESS WORLD

Published by Asia University, Taiwan and Scientific and Business World

# Forecasting the Appearance Frequency of Rafflesia arnoldii in Bengkulu, Indonesia, Using Discrete-valued Time Series Modeling

**Jose Rizal**
Department of Mathematics, The University of Bengkulu, Bengkulu, Indonesia
***Corresponding author* Email:** jrizal04@unib.ac.id

**Nur Afandi**
Department of Mathematics, The University of Bengkulu, Bengkulu, Indonesia
**Email:** nafandi@unib.ac.id

**Gusman Juliadi**
Natural Resources Conservation Center Bengkulu-Lampung, Bengkulu, Indonesia
**Email:** g_juliadi@yahoo.co.id

**Indah Wahyuliani**
Department of Statistics, The University of Bengkulu, Bengkulu, Indonesia
**Email:** indahwahyuliani15@gmail.com

**Cinta Rizki Oktarina**
Department of Mathematics, Batam Institute of Technology, Batam, Indonesia
**Email:** cinta@iteba.ac.id

Received: November 15, 2024; First Revision: July 11, 2025;

Last Revision: January 25, 2026; Accepted: February 28, 2026;

**Published: March 4, 2026**

## Abstract

**Purpose:** This study aims to develop a probabilistic forecasting model to estimate the frequency of Rafflesia arnoldii appearances, a rare flower species endemic to Bengkulu Province, Indonesia. Motivated by the need for effective conservation and ecotourism planning, the study addresses a gap in the literature by applying discrete-valued time series models to ecological count data.

**Design/methodology/approach:** This study analyzes monthly data on the frequency of Rafflesia arnoldii appearances from January 2016 to December 2023. We applied and compared multiple discrete-valued time series models, including Poisson INAR, Negative Binomial INAR, Zero-Inflated INAR, INARMA, NB-INARMA, and INGARCH. The models were fitted using Maximum Likelihood Estimation and the EM algorithm. Model performance was evaluated using AIC, BIC, RMSE, and MAAPE, with post-estimation diagnostics conducted via Ljung-Box, Jarque-Bera, and Lagrange Multiplier tests.

**Findings:** The analysis found that the optimal models were INAR(1)-INGARCH(1,1) for South Bengkulu, and INAR(1)-INGARCH(1,1) for Kepahiang. These models effectively captured the discrete, overdispersed, and zero-inflated characteristics of the data, producing reliable short-term forecasts. The variation in model performance across regencies highlights underlying ecological differences and reinforces the need for region-specific forecasting strategies.

**Research limitations/implications:** Limitations include a focus on four regencies and the exclusion of exogenous environmental variables such as rainfall and temperature. Implications suggest that future studies should expand the data scope, incorporate environmental covariates, and explore spatio-temporal extensions to improve predictive accuracy. In addition, future extensions may explore integer-valued time series frameworks designed to accommodate non-stationary count data, including signed-thinning or generalized INAR formulations, to further enhance model flexibility and predictive performance.

**Practical implications:** The forecasting framework can inform the scheduling of conservation monitoring and guide ecotourism planning by anticipating peak flowering periods. This study’s primary relevance to Decision Sciences lies in providing a quantitative, data-driven framework for conservation resource allocation and ecotourism planning under uncertainty, transforming ecological data into actionable management strategies.

**Originality/value:** This study pioneers the application of discrete-valued time series models in forecasting the occurrences of a rare and sporadically flowering plant species. It contributes to the Decision Sciences literature by demonstrating how probabilistic modeling can be applied to biodiversity management under uncertainty.

**Keywords:** Rafflesia arnoldii, Discrete-valued, INARMA-INGARCH models, Bootstrap Analysis.

**JEL classifications:** C22, C35, C53, Q23, Q56.

## 1. Introduction

Rafflesia arnoldii is an obligate parasitic plant known for its massive blooms, the largest flowers in the world. This plant lacks leaves, grows on vines, and is incapable of photosynthesis. Endemic to Sumatra, particularly in southern regions such as Bengkulu, Jambi, and South Sumatra, Rafflesia arnoldii is classified as a rare and protected species (Collins et al., 1991). According to the International Union for Conservation of Nature (IUCN), its conservation status is endangered. Priatna et al. (1989) highlight that conservation priority should be given to this species due to its limited population and rarity in the wild. Rafflesia’s scarcity stems from its unique biology, including an annual life cycle, dependency on specific liana species for parasitism, and challenging reproduction, among other factors (Tolod et al., 2021). Figure 1 shows the five most commonly found species of Rafflesia in Bengkulu Province (Susatya, 2011).

**Figure 1.** Five types of Rafflesia flowers that are commonly found in Bengkulu Province, on Sumatra, Indonesia.

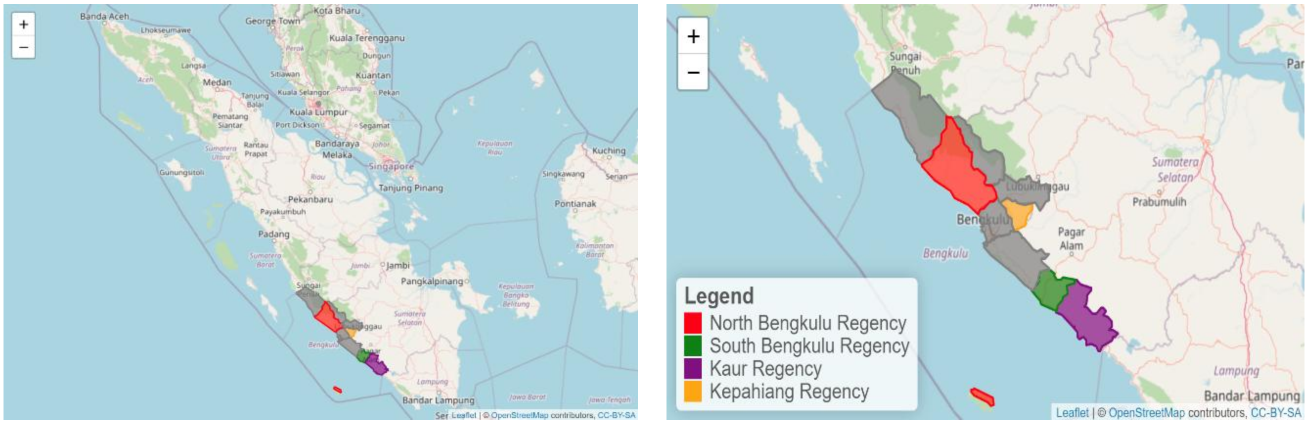
Rafflesia tuan-mudae Rafflesia patma Rafflesia hasseltii

Rafflesia arnoldii Rafflesia pricei

*Note*: The figure displays Rafflesia tuan-mudae, Rafflesia patma, Rafflesia hasseltii, Rafflesia arnoldii, and Rafflesia pricei, representing the most frequently reported species in Sumatra. Species identification and regional occurrence are based on Susatya (2011).

In Bengkulu Province, Indonesia, several new, healthy subpopulations of Rafflesia arnoldii have been identified, allowing this rare flower to continue blooming, including in North Bengkulu, South Bengkulu, Kaur, and Kepahiang regencies, as can be seen in Figure 2. This indicates that, in the long term, the species may be able to maintain its population above the extinction threshold. However, to better anticipate potential extinction risks, it is important to analyze and predict future bloom occurrences. One approach is to develop a probability model based on empirical time series data on the frequency of Rafflesia arnoldii blooms.

**Figure 2.** Map of Rafflesia flower distribution on Sumatra Island and its distribution zones in Bengkulu Province.



Map of Rafflesia flower occurrences on Sumatra Island    Map of Rafflesia flower occurrence in Bengkulu Province

*Note*: The map illustrates the geographical distribution of Rafflesia species across Sumatra Island and highlights their distribution zones within Bengkulu Province. The spatial representation provides an ecological context for the study area analyzed in this research.

Classical time-series methods rooted in mathematical statistics, such as regression analysis, Kalman filtering, and the Box-Jenkins Autoregressive Moving Average (ARMA) framework, remain widely used in practice (Nie et al., 2012). The ARMA model, in particular, is designed for stationary processes under the assumption of an underlying (multivariate) Gaussian data-generating mechanism (Box et al., 2016; Chen et al., 1996). Although well-suited to continuous-valued data, this framework performs poorly for count processes, especially in low-frequency settings where discreteness and non-negativity undermine continuous approximations. Consequently, a class of discrete-valued time-series models has been developed to address these limitations (e.g., Harvey & Fernandes, 1989; Smith, 1979; Zeger & Qaqish, 1988). Notable examples include the Integer-valued Autoregressive (INAR) model (Al-Osh & Alzaid, 1987) and the Integer-valued Autoregressive Moving Average (INARMA) model (Mohammadipour & Boylan, 2012). However, both models are limited in capturing time-varying variance and conditional heteroscedasticity, motivating the development of the Integer-valued Generalized Autoregressive Conditional Heteroscedasticity (INGARCH) model. Analogous to the classical GARCH framework (Engle, 1982), the INGARCH model is specifically designed to model conditional heteroscedasticity in integer-valued time series.

The INGARCH model, introduced by Ferland et al. (2006), is designed to model discrete time series with small values and overdispersion issues that the INARMA model struggles to address (Reboredo et al., 2023). The Autoregressive Conditional Heteroscedasticity (ARCH) model refers to cases where the residuals of an INARMA model meet the white noise assumption yet exhibit random variance (Tsay, 2010). Because time series data often display heteroscedasticity, characterized by randomness in residual variance, the INGARCH model is better suited for such modeling. Like the GARCH model, INGARCH also addresses heteroscedasticity in residual variance (Kim & Jun, 2017). Prior research by Kim and Jun (2017) used the INGARCH model to perform a forecasting analysis for Apple Inc., utilizing count time series data. Their results showed that the best-fitting model was the Negative Binomial-INARCH(1), based on Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) selection criteria.

Although Rafflesia arnoldii has been studied extensively in terms of its biological, ecological, and conservation aspects (Susatya, 2011; Tolod et al., 2021), the forecasting of its blooming frequency using discrete-valued time series modeling has not yet been explored. Prior research on count data modeling (Kim & Jun, 2017; Reboredo et al., 2023; Weiß, 2021; Weiß et al., 2019) has demonstrated the strength of INAR, NB-INAR, and INARMA-INGARCH models in dealing with overdispersion, heteroscedasticity, and excess zeros, common features in ecological datasets. Building upon these advancements, this study pioneers the application of such models in a biodiversity context, particularly for rare and sporadically occurring flora.

This modeling innovation not only fills a methodological gap in the literature but also contributes practical tools for early warning systems in conservation and ecotourism planning. Therefore, this study aims to develop a probabilistic forecasting framework to estimate the monthly frequency of Rafflesia arnoldii appearances using discrete-valued time series models. Specifically, we explore a set of INARMA-INGARCH-based models and identify those that best capture the statistical behavior of blooming counts in different regencies of Bengkulu Province. The models are evaluated using rigorous criteria, including AIC, BIC, Root Mean Squared Error (RMSE), and Mean Arctangent Absolute Percentage Error (MAAPE), and are validated via bootstrapping techniques.

Our approach offers forecasting tools that can be integrated into biodiversity management and conservation planning. It aligns with the principles of Decision Sciences by supporting optimal decisions in biodiversity management through probabilistic forecasting models. These models help policymakers and conservationists prioritize actions under ecological uncertainty, consistent with the mission of the International Association of Decision Sciences (IADS), which promotes data-driven decision-making across disciplines, including environmental sustainability and natural resource management.

The organization of this paper is as follows: Section 2 describes the properties of data used and the time series models applied, covering key concepts in discrete time series modeling, including the Poisson process, overdispersion, excess zeros condition, the INARMA-INGARCH model, and the criteria for selecting the best model, with each topic discussed in its own subsection. Section 3 then details the methodology applied to construct the time series model for count data. Section 4 presents the results of the data analysis and discusses how these findings relate to existing studies. Finally, Section 5 provides conclusions and suggests potential directions for future research.

## 2. Background and Theoretical Framework

### *2.1 Poisson Process*

The Poisson process, represented as $\{Y(t)\,; t \geq 0\}$, is a counting process composed of non-negative, integer-valued, and non-decreasing random variables (Olofsson & Andersson, 2012). It is defined by specific assumptions that quantify the number of events occurring within a specified time interval, according to Tijms (2003). A counting process $\{Y(t)\,; t \geq 0\}$ qualifies as a Poisson process if it satisfies several criteria: at time: at time $t_0$, the number of events is zero $(Y(t_0) = 0)$; the intensity of events, $\lambda(t)$, is constant; one event occurs in each time interval with stationary increments for $t_i \leq t_i + t$ where $i = 0, 1, 2, \dots, n$; the number of events in the interval $[t_i, t_i + t]$, given by $Y(t_i + t) - Y(y_i)$, follows the same distribution as $Y\big((t_i + t) - t_i\big) = Y(t)$; and the increments are independent, meaning the number of events in disjoint time intervals are mutually independent. These characteristics make the Poisson process an effective model for random events occurring over time (Kutoyants, 2023).

Count data consists of non-negative integer values and, by nature, follows a discrete probability distribution. A notable property of such data is that as the count increases, the mean tends to rise, often resulting in a corresponding increase in variance due to the mean variance relationship (Zhang et al., 2014). One widely used model for count data is the Poisson distribution, whose probability mass function is given by:

$$P(Y_t = y_t) = \frac{\left(e^{(-\lambda_t)}\, {\lambda_t}^{y_t}\right)}{y_t!},\ y_t = 0,1,2, \dots, \tag{1}$$

where $Y_t$ is a count-valued random variable at time $t$, $y_t$ is its realization, and $\lambda_t > 0$ denotes both the mean and the variance of the distribution. This equality of mean and variance is a defining feature of the Poisson model. Throughout this paper, we use $\lambda_t$ to denote a time-varying conditional mean (e.g., in Poisson and INGARCH models), whereas $\lambda$ denotes the constant mean of the Poisson innovations in INAR and INARMA models. However, this equality of mean and variance is often violated in real-world data due to overdispersion, a condition in which the conditional variance of the dependent variable exceeds the conditional mean (Cameron & Trivedi, 1990).

### *2.2 Overdispersion in Count Data*

Overdispersion in the Poisson distribution occurs when the variance of the variable exceeds its mean. This phenomenon can arise from several factors, including correlations between observations, violations of Poisson distribution assumptions, namely the condition that $\mathrm{Var}(Y_t) = \mathrm{E}(Y_t)$, the presence of excess zeros in the dataset, and outliers that can distort the results. Overdispersion can significantly affect the accuracy of the model, often leading to inflated deviance values that indicate a poor fit to the data. Mathematically, the overdispersion value can be calculated using $I_d = S^2_{Y_t}(\bar{y}_t)^{-1}$ ; $S^2_{Y_t}$ is the sample variance and $\bar{y}_t$ is the sample mean (Hardin & Hilbe, 2018).

We can use the Fisher test to detect overdispersion relative to Poisson distribution (Yang et al., 2009), where this process involves several key steps. First, the hypotheses are defined, with the null hypothesis $H_0\colon \mathrm{Var}(Y_t) = \mathrm{E}(Y_t)$ and the alternative hypothesis $H_1\colon \mathrm{Var}(Y_t) \neq \mathrm{E}(Y_t)$. Given significance level $\alpha$, sample size $n$, and number of estimated parameters $k$, the test statistic is defined as,

$$\zeta_{test} = \frac{2}{(n-k)} \sum_{t=1}^{n} \left[ y_t \log\left(\frac{y_t}{\lambda_t}\right) - (y_t - \lambda_t) \right], \tag{2}$$

where $\zeta_{test}$ denotes the overdispersion test statistic, $y_t$ is the observed count at time $t$, $\lambda_t$ is the expected value under the null, $n$ is the sample size, $k$ is the number of estimated parameters, and $df = n - k$ is the degrees of freedom. The null hypothesis is rejected if $\zeta_{test} > \chi^2(\alpha; df)$, which indicates the presence of overdispersion.

To address overdispersion, we may replace the Poisson distribution with more flexible alternative distributions, such as the Negative Binomial or Hermite distributions. These models are particularly useful in both equidispersion and overdispersion scenarios, as they do not require the variance to equal the mean.

### *2.3 Excess Zeros in Count Data*

In discrete time series data modeling, the excess zeros condition occurs when the observed count data contain an unusually high number of zeros compared to what is typically expected under standard count distributions, such as the Poisson or Negative Binomial distributions. This phenomenon is common across various fields, including ecology, economics, and healthcare, where certain events may happen infrequently or not at all in a significant portion of observations. For instance, in ecological studies, a particular species may be absent from many sampling locations, resulting in a substantial number of recorded zeros. The presence of excess zeros can skew the results of conventional statistical analyses. Traditional count models generally assume that the data follows a specific distribution in which the variance is equal to or proportional to the mean. However, when there are more zeros than expected, this assumption is violated, leading to biased parameter estimates and unreliable predictions. It should be noted that ignoring zero inflation can result in misleading conclusions and ineffective decisions.

In this study, we conducted a zero-inflated detection test to identify the presence of excess zeros in the data. The testing process began by establishing the hypotheses, namely the null hypothesis $(H_0)$ asserts that there are no excess zeros, while the alternative hypothesis $(H_1)$ suggests that there are indeed excess zeros. We set the significance level at $\alpha$ = 5% to determine statistical significance. The test statistic used to detect zero inflation is calculated as

$$H_{test} = \frac{(n_0 - n\hat{p}_0)^2}{n\hat{p}_0\,(1-\hat{p}_0) - n\bar{y}_t\hat{p}_0^2}, \tag{3}$$

where $H_{test}$ denotes the zero-inflation test statistic, $n_0$ is the actual count of zero observations in the dataset, $\hat{p}_0$ represents the estimated probability of observing a zero count in the dataset, $n$ is the sample size, and $\bar{y}_t$ is the sample mean of $Y_t$. The numerator captures the difference between the observed number of zeros ($n_0$) and the expected number of zeros ($n\hat{p}_0$), while the denominator accounts for the variability associated with the zero counts.

The criterion for rejecting the null hypothesis relies on the calculated test statistic $H_{test}$. If this value exceeds the critical value from the chi-squared ($\chi^2$) distribution table with 1 degree of freedom, we reject the null hypothesis. This outcome indicates significant evidence of excess zeros in the dataset, suggesting that a zero-inflated model may be more suitable for the analysis. This method aligns with the findings of Weiß et al. (2016), who emphasize the necessity of recognizing and appropriately modeling excess zeros in count data. By applying this testing approach, we ensure that our statistical analyses accurately reflect the underlying characteristics of the data, leading to more robust and reliable conclusions.

### *2.4 Integer-Valued Autoregressive Models (INAR)*

The ARIMA models are commonly used for analyzing time series data, data collected at regular intervals, such as daily, weekly, monthly, or annually. ARIMA is specifically designed for continuous time series data, where its autoregressive and moving average components effectively capture linear relationships. However, the model has significant limitations when applied to non-continuous or discrete data, such as counts or categorical values. ARIMA's reliance on differencing and linear components makes it ill-suited for discrete data, which often displays non-linear and complex patterns that ARIMA cannot adequately model. Unlike continuous data, which ARIMA can model as a blend of linear processes, discrete data do not exhibit these linear properties. As such, ARIMA is best suited for continuous time series data, and its effectiveness is limited when applied to datasets that do not meet these characteristics.

Designed for modeling count data, the INAR model extends the classical autoregressive (AR) framework and has been widely applied in disciplines such as economics, epidemiology, and social science. A key component of the INAR model is the binomial thinning operator $\alpha \circ Y$, introduced by Al-Osh and Alzaid (1987), which is defined as:

$$\alpha \circ Y = \sum_{i=1}^{Y} B_i,\ B_i \sim \text{Bernoulli}(\alpha),\ \alpha \in [0,1], Y \in \mathbb{Z}_+, \tag{4}$$

where $\alpha$ is the thinning parameter, $\{B_i\}_{i=1}^{Y}$ are i.i.d. Bernoulli random variables with success probability $P(B_i = 1) = \alpha$, and $Y$ is a non-negative integer-valued random variable. This operator ensures that the outcome remains integer-valued, thereby preserving the discrete nature of the process while allowing for probabilistic thinning of previous counts. The probability mass function of each Bernoulli trial with success probability $\alpha$ is given by:

$$P(B_i = b) = \alpha^b (1-\alpha)^{1-b}, b \in \{0,1\}. \tag{5}$$

Using this operator, the INAR(1) model is expressed as

$$Y_t = \alpha \circ Y_{t-1} + Z_t,\ Z_t \sim \text{Poisson}(\lambda), \tag{6}$$

where $Z_t$ is a sequence of i.i.d. non-negative integer-valued random variables, typically assumed to follow a Poisson distribution. This formulation enables the model to capture autoregressive behavior in count data while maintaining the discreteness required for integer-valued outcomes.

Parameter estimation is conducted using the Expectation-Maximization (EM) algorithm, which is an efficient iterative method for computing Maximum Likelihood (ML) estimates in the presence of missing or hidden data. The goal of ML estimation is to identify the model parameters that maximize the likelihood of the observed data. The EM algorithm consists of two key processes in each iteration: the Expectation step (E-step) and the Maximization step (M-step). During the E-step, the algorithm estimates the missing data based on the observed data and the current parameter estimates, employing conditional expectations; this is why it is referred to as the Expectation step. In the subsequent M-step, the algorithm maximizes the likelihood function by treating the estimated values from the E-step as if they were the actual missing data. The EM algorithm is structured to increase the likelihood with each iteration, ensuring convergence toward the optimal parameter estimates.

The parameter estimation procedure follows the Expectation-Maximization (EM) algorithm as described by Sammaknejad et al. (2019). First, initial values for the model parameters are specified. For the INAR model, these initial estimates are obtained using moment-based estimation via the Conditional Least Squares (CLS) method, where the initial parameter value is given by $\theta^{(0)} = \frac{Cov\ (Y_t, Y_{t-1})}{Var\ (Y_{t-1})}$, with both the covariance and variance computed from the sample data. Second, the Expectation (E) step is performed by computing the conditional expectation of the complete-data log-likelihood function, treating the latent variables as random. Specifically, if $Z_t$ denotes the unobserved (missing) data and $Y_t$ represents the observed data, the quantity $\mathrm{E}[\log \mathrm{L}(\theta; \mathrm{Y_t}, \mathrm{Z_t}) | \mathrm{Y_t}, \theta^{(t)}]$ is evaluated. Third, the Maximization (M) step updates the parameter estimates by maximizing the expected log-likelihood obtained in the E step with respect to $\theta$, yielding $\theta^{(t+1)} = \arg\max_{\theta} \mathrm{E}[\log \mathrm{L}(\theta; Y_t, Z_t) | Y_t, \theta^{(t)}]$. Finally, convergence is assessed by monitoring either changes in the log-likelihood function or successive updates of the parameter estimates. If convergence criteria are not satisfied, the E and M steps are iteratively repeated until convergence is achieved.

### *2.5 Integer-Valued Autoregressive Moving Average Models (INARMA)*

Aly and Nadjib (1994) extended the INAR and INMA frameworks by combining them into a unified model, analogous to the classical ARMA process, resulting in the integer-valued autoregressive moving average (INARMA) model. The general structure of the INARMA(1,1) process is defined as follows:

$$Y_t = \alpha \circ Y_{t-1} + Z_t + \beta \circ Z_{t-1}, \tag{7}$$

where $\alpha, \beta \in [0,1]$ are model parameters, $\alpha \circ Y_{t-1}$ and $\beta \circ Z_{t-1}$ denote binomial thinning operations, and $\{Z_t\}$ is an i.i.d. sequence of non-negative integer-valued innovations with mean $\mu$ and variance $\sigma^2$, independent of the past process $\{Y_s: s < t\}$. Based on Equation 7, caution is needed because the model involves multiple thinning operations.

To ensure valid stochastic properties, several independence assumptions are generally made (Weiß et al., 2019). Specifically, the Bernoulli sequences $\{B_{t,i}\}$ and $\{C_{t,i}\}$ associated with $\alpha \circ Y_{t-1}$ and $\beta \circ Z_{t-1}$ are assumed to be mutually independent and also independent of the innovation process $\{Z_t\}$. To verify the stationarity and invertibility of the INARMA(1,1) process, we examine its general form. Two conditions must be satisfied: $\alpha < 1$ and $\beta < 1$. This can be demonstrated by examining the first moment, autocovariance, and autocorrelation of the INARMA(1,1) process (Mohammadipour & Boylan, 2012).

Maximum Likelihood Estimation (MLE) estimates unknown parameters by maximizing the likelihood function of the observed data (Gujarati & Porter, 2009). Since INAR models are Markov processes, MLE can be applied conveniently (Dungey et al., 2019). Under the assumption of Poisson innovations, the log-likelihood of the INARMA model is given in Equation 8 (Weiß et al., 2019).

$$\log L(\alpha,\lambda,\beta) = \log(a) + \sum_{t=2}^{n} \log\left(\sum_{i=0}^{b}\sum_{j=0}^{c} \binom{y_{t-1}}{i} \alpha^{i}(1-\alpha)^{y_{t-1}-i} \binom{Z_{t-1}}{j} \beta^{j}(1-\beta)^{Z_{t-1}-j}\vartheta\right), \tag{8}$$

where $a = \frac{e^{-\left(\frac{\lambda}{1-\alpha}\right)}\left(\frac{\lambda}{1-\alpha}\right)^{y_1}}{y_1!}$, $b = \min(y_{t-1}, y_t)$, $c = \min(Z_{t-1}, y_t - i)$, and $\vartheta = \frac{e^{-\lambda}\lambda^{y_t-i-j}}{(y_t-i-j)!}$. Here, $y_t$ denotes the observed count at time $t$, $n$ is the sample size, $\{Z_t\}$ are the i.i.d. innovation terms following Poisson $(\lambda)$, and $\alpha, \beta \in [0,1)$ are the binomial thinning parameters. The parameters are estimated by maximizing (8) subject to $\alpha, \beta \in [0,1)$ and $\lambda > 0$; standard errors can be obtained from the observed information matrix.

Diagnostic testing is essential for evaluating the adequacy of the INARMA model and ensuring valid statistical inference. Residual analysis is used to verify key assumptions, including white noise, normality, and homoscedasticity. These tests help detect issues such as autocorrelation or non-constant variance that may undermine forecast reliability. Among them, the white-noise test assesses whether residuals are uncorrelated and randomly distributed over time, an expected property in well-specified time series models. The white-noise test examines the null hypothesis $H_0\colon \rho_1 = \cdots = \rho_k = 0$ and $H_1\colon \exists\, j \in \{1,2,\ldots,k\}$ such that $\rho_j \neq 0$, where $\rho_k$ denotes the autocorrelation of lag $k$ (Wei, 2006). The test statistic is computed as:

$$Q_{test} = n(n+2)\sum_{k=1}^{K} \frac{\hat{\rho}_k^2}{n-k}, \tag{9}$$

where $Q_{test}$ denotes the Ljung Box test statistic, $n$ is the sample size, $\hat{\rho}_k$ is the sample autocorrelation at lag $k$, and $K$ is the number of lags used. The null hypothesis $H_0$ is rejected if $Q_{test} > \chi^2(\alpha; K-p-q)$, where $\chi^2(\alpha; K-p-q)$ is the upper $\alpha$-quantile of the chi-square distribution with $K-p-q$ degrees of freedom, $\alpha$ is the significance level, and $p$ and $q$ are the autoregressive and moving average orders of the fitted model. Rejection of $H_0$ indicates that the residuals are not white noise.

The normality test is used to assess whether the residuals of the INARMA model follow a normal distribution. Following Gujarati and Porter (2009), the hypotheses for the test are stated as: $H_0$: The residuals are normally distributed and $H_1$: The residuals are not normally distributed. To perform the test, the necessary values of the significance level $\alpha$, sample size $n$, and degrees of freedom $df$ are specified. The test statistic, known as the Jarque Bera statistic, is computed as follows:

$$JB_{test} = \frac{n}{6}\left[\left(\frac{1}{n}\sum_{t=1}^{n}{\xi_t}^{3}\left(\left(\frac{1}{n}\sum_{t=1}^{n}{\xi_t}^{2}\right)^{-\frac{3}{2}}\right)\right)^{2} + 0.25\left(\left(\frac{\frac{1}{n}\sum_{t=1}^{n}{\xi_t}^{4}}{\left(\frac{1}{n}\sum_{t=1}^{n}{\xi_t}^{2}\right)^{2}}\right) - 3\right)^{2}\right], \tag{10}$$

where $JB_{test}$ denotes the Jarque-Bera statistic used to test the null hypothesis of normality, $\xi_t = z_t - \bar{z}$, with $z_t$ representing the residual at time $t$, and $\bar{z}$ denoting the sample mean of the residuals. The null hypothesis $H_0$ is rejected if: $JB_{test} > \chi^2(\alpha; 2)$ or the p-value $(JB_{test}) < \alpha$, indicating that the residuals are not normally distributed.

Heteroskedasticity in the residual variance of an INARMA model indicates that the variance of residuals varies inconsistently over time. To detect this phenomenon, the Lagrange Multiplier (LM) test is commonly employed. According to Tsay (2010), the test operates under the following hypotheses: $H_0$ proposes that the residual variance of the model is homoscedastic (constant over time), while $H_1$ suggests heteroskedasticity. To conduct the test, the significance level $\alpha$, sample size $n$, and degrees of freedom must be specified. The LM test statistic is calculated using the following formula:

$$LM_{test} = nR^2, \tag{11}$$

where $LM_{test}$ denotes the Lagrange Multiplier statistic, $n$ is the sample size, and $R^2$ is the coefficient of determination from the auxiliary regression of squared residuals $u_t^2$ on their lagged values up to order $k$. Under the null hypothesis, $LM_{test} \sim \chi^2(k)$, where $k$ is the number of lag terms. The null is rejected if $LM_{test} > \chi^2(\alpha; k)$, indicating the presence of heteroskedasticity.

### *2.6 Integer-Valued GARCH Models (INGARCH)*

The Integer-valued Generalized Autoregressive Conditional Heteroscedasticity (INGARCH$(p, q)$) model developed by Ferland et al. (2006) has two parameters, where $p$ is the autoregressive parameter and $q$ is the moving average parameter. Let $Y_t$ be a non-negative integer-valued time series that follows a Poisson distribution. The INGARCH(p,q) model, as introduced by Ferland et al. (2006), is defined as:

$$\lambda_t = \gamma_0 + \sum_{i=1}^{p} \gamma_i Y_{t-i} + \sum_{j=1}^{q} \delta_j \lambda_{t-j}, \tag{12}$$

where $\lambda_t = E[Y_t | \mathcal{F}_{t-1}]$ denotes the conditional mean of $Y_t$ given the past information set $\mathcal{F}_{t-1}$, is the conditional mean of $Y_t$ given past information, $\gamma_0 > 0$ is the intercept, $\gamma_i \geq 0$ for $i = 1, \ldots, p$ are the autoregressive coefficients (INARCH parameters), $\delta_j \geq 0$ for $j = 1, \ldots, q$ are the moving average coefficients (INGARCH parameters), and the stationarity condition requires $\sum_{i=1}^{p} \gamma_i + \sum_{j=1}^{q} \delta_j < 1$ (Weiss, 2018).

### *2.7 Model Selection Criteria*

Model selection is a crucial aspect of statistical modeling, especially when working with time series data. In this study, we utilize several criteria to evaluate the performance of the INARMA-INGARCH models, with a focus on the Akaike Information Criterion (AIC) (Akaike, 1974) and the Bayesian Information Criterion (BIC) (Schwarz, 1978) as key indicators of model fit and complexity.

The AIC is expressed as follows:

$$\mathrm{AIC} = -2\ln\left(\hat{\mathcal{L}}\right) + 2k, \tag{13}$$

where $k$ represents the number of estimated parameters in the model, and $\hat{\mathcal{L}}$ is the maximum likelihood of the model. The AIC balances goodness of fit with model complexity, penalizing models with more parameters. A lower AIC value indicates a better-fitting model, making it a valuable tool for model selection. Similarly, the BIC is expressed as follows:

$$\mathrm{BIC} = -2\ln(\hat{\mathcal{L}}) + k\ln(n), \tag{14}$$

where $n$ is the sample size. Compared to AIC, BIC applies a stronger penalty for complexity, particularly in large samples, thus favoring simpler models. Like AIC, lower BIC values indicate a better fit.

While AIC and BIC are both useful for comparing models, if they yield different conclusions, the likelihood ratio test ($\mathrm{LR}_{test}$) can provide additional clarity. The $\mathrm{LR}_{test}$ is commonly used to compare the goodness of fit between two statistical models, particularly when one model is a special case of the other. In the context of the $\mathrm{LR}_{test}$, the hypotheses are typically formulated as follows: Null hypothesis ($H_0$): The simpler model is adequate for the data vs Alternative hypothesis ($H_1$): The more complex model provides a significantly better fit to the data than the simpler model. To test these hypotheses, the $\mathrm{LR}_{test}$ is calculated using the maximum likelihoods of the two models:

$$\mathrm{LR}_{test} = -2\left(\ln(\hat{\mathcal{L}}_0) - \ln(\hat{\mathcal{L}}_1)\right), \tag{15}$$

where $\mathrm{LR}_{test}$ is the likelihood ratio test statistic, $\hat{\mathcal{L}}_0$ is the likelihood of the simpler model, and $\hat{\mathcal{L}}_1$ is the likelihood of the more complex model. The $\mathrm{LR}_{test}$ in Equation 15 follows a chi-squared distribution with degrees of freedom equal to the difference in the number of parameters between the two models (denoted by $v$). The null hypothesis is rejected if $\mathrm{LR}_{test} > \chi^2(\alpha; v)$, or p-value $< \alpha$, which means one of the methods performs better than the other in favor of the alternative hypothesis. While AIC and BIC are useful for identifying well-fitting models and avoiding overfitting, they do not directly measure predictive accuracy. To address this, RMSE quantifies the average magnitude of prediction errors, while MAAPE offers a scale-sensitive alternative that remains robust in the presence of zero values. The formulas for RMSE and MAAPE are as follows:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{t=1}^{n}\left(Y_t - \hat{Y}_t\right)^2}, \tag{16}$$

$$\mathrm{MAAPE} = \frac{1}{n}\sum_{t=1}^{n}\arctan\left(\frac{|Y_t - \hat{Y}_t|}{|Y_t|}\right), \tag{17}$$

where $Y_t$ denotes the observed value at time $t$, $\hat{Y}_t$ is the model's forecast, and $n$ is the total number of observations.

## 3. Data and Methodology

This study employs a discrete-valued time series modeling framework, INARMA-INGARCH, to predict the monthly frequency of Rafflesia arnoldii appearances in Bengkulu Province, Indonesia. Given the ecological characteristics of the data, including the integer-valued outcomes, the low frequency, overdispersion, and the potential zero inflation, this modeling approach allows for a more accurate representation of the temporal and stochastic dynamics observed in rare-event ecological processes.

### *3.1 Data Description*

Let the random variable $Y_t$ represent the monthly frequency of Rafflesia arnoldii blooms, with each data point capturing the number of flowers observed in one-month intervals. This historical data, obtained from the Natural Resources Conservation Center Bengkulu-Lampung in Bengkulu, Indonesia, includes monthly bloom counts for different regencies and serves as a crucial resource for analyzing the distribution and frequency of Rafflesia arnoldii appearances. This study analyzes monthly blooming patterns of Rafflesia arnoldii across four Bengkulu regencies, namely North Bengkulu, South Bengkulu, Kaur, and Kepahiang, using 96 observations per regency from January 2016 to December 2023. The data reveal spatial and temporal dynamics, including bloom frequency, seasonality, and potential periodicity.

Given the unique characteristics of the time series data, it is essential to first introduce an appropriate time series model that aligns with these properties in the sections that follow. Additionally, common challenges in discrete time series modeling, such as overdispersion (where the variance exceeds the mean) and an excess of zero counts, must be addressed to ensure accurate model selection and interpretation. These conditions, along with their implications for model choice, are discussed to provide a solid foundation for the analysis.

### *3.2 Rationale for Discrete-Valued Time Series Models*

Conventional time series models such as ARMA or ARIMA (Box et al., 2016) are designed for continuous and Gaussian-distributed data. These assumptions do not hold for count data like the monthly blooming of Rafflesia arnoldii, which are inherently non-negative integers and often characterized by low frequency, many zeros, and fluctuating variance. To overcome these limitations, we adopt a class of models tailored for discrete time series. The INAR (Integer-Valued Autoregressive) model (Al-Osh & Alzaid, 1987) uses a binomial thinning operator to ensure that all model outputs remain integer-valued. However, the standard INAR may be insufficient when count data exhibit overdispersion, variance exceeding the mean, a common feature in ecological datasets. To handle this, the Negative Binomial INAR (NB-INAR) model allows for a more flexible variance structure (Weiß, 2021). Furthermore, in the presence of excessive zeros, as often found in rare species data, the Zero-Inflated INAR (ZINAR) model introduces a zero-inflation component to model the structural and sampling zeros separately (Weiß et al., 2016).

Subsequently, to capture both autoregressive and moving average components, we apply the INARMA model (Mohammadipour & Boylan, 2012), which extends INAR by incorporating lagged innovations.

However, these models assume constant conditional variance, which may not reflect the heteroscedastic nature of real-world ecological systems. For this reason, we integrate the INGARCH (Integer-valued Generalized Autoregressive Conditional Heteroscedasticity) process (Ferland et al., 2006) to model conditional variance as a dynamic process influenced by past outcomes and expectations. Thus, the INARMA-INGARCH framework, supported by NB and ZI extensions, allows us to model both the mean and the volatility structure of the data within a unified, flexible probabilistic framework.

### *3.3 Pre-modeling Diagnostics*

We obtained monthly counts of Rafflesia arnoldii bloom occurrences from four regencies in Bengkulu, namely South Bengkulu, Kepahiang, Kaur, and North Bengkulu, covering the period from January 2016 to December 2023. Exploratory data analysis was conducted to examine temporal patterns, potential seasonality, and variance behavior. We used time-series plots to visualize the data and employed the Augmented Dickey Fuller (ADF) test to check for stationarity. When a series exhibited unit roots or non-stationarity, first-order differencing was applied. This step is essential because many time-series models, including INARMA, assume weak stationarity in the count process.

**Table 1.** Recapitulation of the overdispersion and excess zeros test results with a significance level ($\alpha$) of 5%

| | ADF Test | | Overdispersion Test | | | Excess Zero Test | | | Linearitas Test | |
|---|---|---|---|---|---|---|---|---|---|---|
| Regency | $ADF_{test}$; p-value | Conclusion | DI | $\zeta_{test}$; p-value | Decision $H_0$ | Prop (%) | $H_{test}$; p-value | Decision $H_0$ | $\chi^2{}_{test}$; p-value | Decision $H_0$ |
| North Bengkulu | -3.904; 0.017 | I(0) | 1.390 | 2.081; 0.018 | Rejected | 36 | 12.318; 0.000 | Rejected | 0.066; 0.967 | Accepted |
| South Bengkulu | -3.674; 0.031 | I(0) | 3.670 | 2.976; 0.001 | Rejected | 74 | 84.651; 0.000 | Rejected | 0.835; 0.658 | Accepted |
| Kaur | -2.276; 0.462 | I(1) | 1.770 | 3.321; 0.000 | Rejected | 37 | 17.413; 0.000 | Rejected | 0.939; 0.625 | Accepted |
| Kepahiang | -3.994; 0.013 | I(0) | 1.660 | 2.348; 0.080 | Accepted | 80 | 6.690; 0.009 | Rejected | 1.622; 0.444 | Accepted |

*Note*: The ADF test checks stationarity ($H_0$: unit root, non-stationary). “I(0)” indicates stationarity at the level, “I(1)” after first differencing. The dispersion index (DI) is the ratio of variance to mean, $\zeta_{test}$ tests overdispersion ($H_0$: $Var(Y_t) = E(Y_t)$). $H_{test}$ is a Chi-square test for excess zeros ($H_0$: no excess zeros). “Rejected” means $H_0$ rejected at α = 0.05; Prop (%) represents the proportion of zero observations. Linearitas test evaluates whether the conditional mean function is linear ($H_0$: the relationship between $Y_t$ and $t$ is linear).

Table 1 summarizes the results of the ADF test, the overdispersion test, and the excess zeros test at a 5% significance level. The ADF results indicate that three series are stationary at level (I(0)), while one requires first differencing (I(1)). The overdispersion test results, based on the dispersion index (DI) of Cameron and Trivedi (1990), reveal significant overdispersion across all regencies. Similarly, the Chi-square test for excess zeros shows that three regencies exhibit significant zero inflation.

Based on Table 1, the Kaur series indicates I(1) behavior according to the ADF test. However, linear differencing $Y_t - Y_{t-1}$ is not suitable for count data as it may produce negative values and violate the non-negativity assumption of standard INAR/Poisson models. The literature on integer-valued time series also acknowledges that non-stationary count data require different treatment from real-valued ARIMA models (Kim & Park, 2008). Therefore, since the Kaur series does not satisfy the non-negativity and stationarity assumptions required for INAR-based count time series models, this series is not included in subsequent modeling. These diagnostics further confirm overdispersion (DI > 1) in all remaining regencies, justifying the use of negative binomial-based models. In addition, the evidence of significant zero inflation (Hardin & Hilbe, 2018) motivates the use of zero-inflated specifications. Together, these results support the adoption of NB-INAR and ZINAR as benchmark models before applying the full INARMA-INGARCH framework. Furthermore, the linearity test confirms that all remaining series exhibit a linear mean structure, indicating that linear integer-valued autoregressive specifications are appropriate for subsequent modeling.

### *3.4 Modeling and Estimation Strategy*

This study adopts a discrete-valued time series modeling framework that integrates an INARMA structure with a dynamic INGARCH component to capture the key characteristics of Rafflesia arnoldii blooming data. The model is designed to accommodate common features of ecological count series, including autocorrelation, overdispersion, and conditional heteroscedasticity.

The core specification is based on the INARMA(1,1) process (Equations 6 and 7), which incorporates autoregressive and moving average dynamics through binomial thinning operations. This structure preserves the integer-valued nature of the data while capturing short-term temporal dependencies through lagged observations and past innovations. The conditional mean of the innovation process is modeled using an INGARCH(1,1) component (Equation 12), which allows the expected count to evolve dynamically based on past observations and previous conditional means. This feature provides flexibility in adapting to irregular blooming patterns commonly observed in ecological time series.

To address distributional issues, the model is extended where necessary. In its standard form, the innovation term $Z_t$ follows a Poisson distribution. However, when overdispersion is present, a Negative Binomial distribution is used for $Z_t$, while retaining the INARMA INGARCH structure. For datasets with a high proportion of structural zeros, a Zero-Inflated INARMA formulation (Equation 4) is applied, allowing a dual process to generate excess zeros alongside the count-generating mechanism. Model specifications are selected individually for each regency, based on the statistical characteristics of the corresponding dataset. This flexible approach enables localized adaptation and supports more accurate forecasting of rare ecological events.

### *3.5 Model Diagnostics and Selection*

The adequacy of each model was assessed through a combination of diagnostic testing, predictive performance evaluation, and statistical model selection. For model diagnostics, three standard statistical tests were applied to verify core assumptions. The Ljung-Box test (Equation 9) was used to detect residual autocorrelation, ensuring that model errors followed a white noise process. The Jarque Bera test (Equation 10) assessed residual normality, while the Lagrange Multiplier (LM) test (Equation 11) was employed to evaluate the presence of conditional heteroscedasticity. These diagnostics confirmed whether the residual structure of each fitted model aligned with theoretical expectations.

For model evaluation, both information-theoretic and predictive performance criteria were considered. The Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC), defined in Equations 13 and 14, were used to assess model fit while penalizing complexity. For nested models, the Likelihood Ratio (LR) test (Equation 15) was used to assess whether the more complex model significantly improved the fit over its simpler counterpart. Forecast accuracy was evaluated using RMSE and MAAPE (Equations 16 and 17). RMSE measures the average prediction error, while MAAPE provides a more stable and interpretable alternative to MAPE, especially for data with frequent zeros.

Final model selection for each regency considered low AIC and BIC values, strong predictive performance (RMSE and MAAPE), and satisfactory diagnostic results. When models performed similarly, ecological interpretability, parameter parsimony, and the LR test were also taken into account.

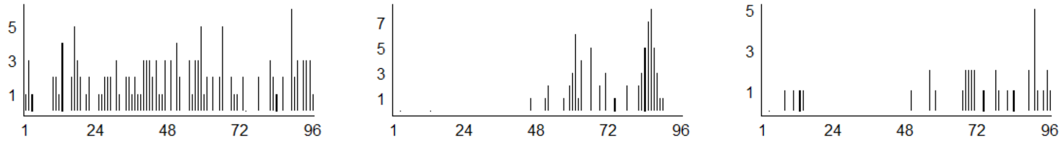
## 4. Results and Discussion

The study aims to understand the blooming patterns of Rafflesia arnoldii, a rare and unique endemic species known for its large size and distinct characteristics. By examining the blooming counts separately for each regency, the research seeks to uncover regional patterns and potential seasonal trends in blooming frequency. Since the data consists of monthly blooming counts, it requires discrete data modeling. A key step in the analysis is to check the equidispersion assumption, which compares the variance and mean of the counts to determine if they are approximately equal. This check helps identify if the data can be modeled with a simple Poisson distribution or if more complex models are needed to address overdispersion or underdispersion.

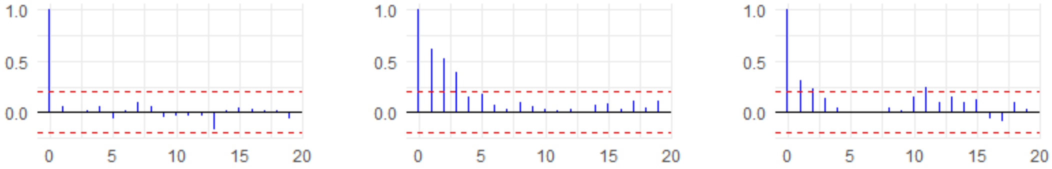
This research uses data on the number of Rafflesia arnoldii flowers blooming in four regencies of Bengkulu Province: North Bengkulu, South Bengkulu, and Kepahiang. The data includes monthly counts of blooms from January 2016 to December 2023, covering eight full years and resulting in 96 data points for each regency. The goal of the study is to analyze the blooming patterns of Rafflesia arnoldii, a rare and unique flower endemic to the region. Since the data is discrete, the study checks the assumption of equidispersion by comparing the expected value and variance of $Y_t$. Therefore, the following assumptions need to be verified.

Based on the results in Table 1, it is found that the regencies of North Bengkulu and South Bengkulu exhibit overdispersion, while Kepahiang is the only regency where the null hypothesis of equidispersion is not rejected at the 5% significance level. Overdispersion occurs when the variance of the data exceeds the mean, which is often caused by factors such as heterogeneity in the sample, the presence of outliers, or zero inflation. Since zero inflation is a common cause of overdispersion, it is important to detect it using the zero-inflation test. The test confirmed zero inflation in all four regencies, likely reflecting the rarity and sporadic nature of Rafflesia arnoldii blooms.

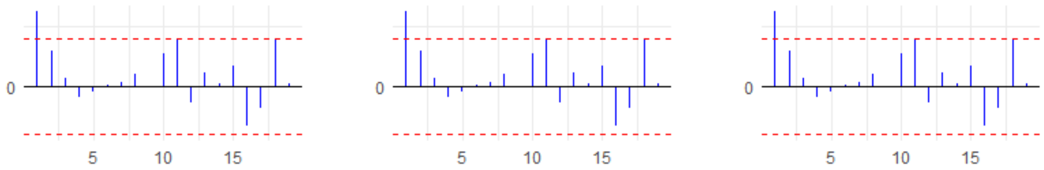
**Figure 3.** Visualization of data, ACF, and PACF plots of Rafflesia arnoldii for each regency.



North Bengkulu Data Plot South Bengkulu Data Plot Kepahiang Data Plot



North Bengkulu ACF Plot South Bengkulu ACF Plot Kepahiang ACF Plot



North Bengkulu PACF Plot South Bengkulu PACF Plot Kepahiang PACF Plot

*Note*: The figure presents the monthly time series of Rafflesia arnoldii appearance counts along with their Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots for each regency. These plots are used to examine serial dependence patterns and support the identification of appropriate discrete-valued time series models.

Understanding the blooming dynamics of Rafflesia arnoldii is essential for conservation efforts and the development of nature-based tourism in Bengkulu. Figure 3 presents bloom occurrences across four regencies from 2016 to 2023, highlighting substantial temporal variability and heterogeneous patterns. Although no consistent long-term trend is evident, episodic increases in certain years suggest the influence of seasonal or external environmental factors. ACF and PACF diagnostics reveal no temporal dependence in North Bengkulu, leading to its exclusion from further analysis. In contrast, South Bengkulu and Kepahiang exhibit overdispersion and excess zeros, thereby justifying the application of the Zero-Inflated Integer-Valued Autoregressive (ZINAR) model. Model performance is subsequently evaluated through comparison with the AR and INAR processes.

Table 2 summarizes the goodness of fit and parameter estimates for various models applied to South Bengkulu and Kepahiang regencies. For South Bengkulu, the ZINAR(1) model performs best, with the lowest AIC and BIC, highlighting the importance of zero-inflation, captured by the parameter $\rho$. Similarly, Kepahiang's data is best described by the ZINAR(1) model, suggesting that zero-inflation significantly improves fit, as indicated by low AIC and BIC values. Across all regencies, the models' additional parameters, including $\theta$ for overdispersion and $\rho$ for zero-inflation, demonstrate the need for flexible models to account for unique data characteristics like overdispersion and excess zeros in each regency's appearance data. In the next stage, the modeling approach was enhanced by applying ARMA models, which combine autoregressive and moving average components to better capture temporal dependencies.

**Table 2.** Goodness of fit and parameter estimates for AR models and modifications

| Regency | Model | Goodness of Fit | | | Parameters | | | |
|---|---|---|---|---|---|---|---|---|
| | Applied | Log-Likelihood | AIC | BIC | $\alpha$ | $\lambda$ | $\theta$ | $\rho$ |
| South Bengkulu | AR(1) | -152.310 | 310.600 | 318.100 | 0.615$^{***}$ | | | |
| | INAR(1) | -104.247 | 214.400 | 221.900 | 0.433$^{***}$ | 0.445$^{***}$ | | |
| | NBINAR(1) | -95.473 | 196.900 | 204.400 | 0.239$^{**}$ | 0.593$^{***}$ | 0.172$^{***}$ | |
| | ZINAR(1) | -92.562 | 191.100 | 198.600 | 0.250$^{**}$ | 2.453$^{***}$ | | 0.762$^{***}$ |
| Kepahiang | AR(1) | -76.690 | 159.300 | 166.800 | 0.309$^{***}$ | | | |
| | INAR(1) | -57.310 | 120.600 | 128.100 | 0.239$^{**}$ | 0.214$^{***}$ | | |
| | NBINAR(1) | -126.515 | 152.200 | 160.000 | 0.234 | 0.215$^{***}$ | 0.884 | |
| | ZINAR(1) | -56.487 | 118.900 | 126.400 | 0.222 | 0.482 | | 0.546$^{**}$ |

*Note*: $\alpha$ is the autoregressive parameter; $\lambda$ is the innovation parameter; $\theta$ is the overdispersion parameter; $\rho$ is the zero-inflation parameter. AIC and BIC are goodness-of-fit measures, with lower values indicating better fit. Empty cells indicate parameters not applicable for that model. Parameters marked with asterisks are statistically significant: $^{***}$ at the 1% level, $^{**}$ at the 5% level, and $^{*}$ at the 10% level.

To accommodate the count nature and variability of the data, three distributions, namely Poisson, Negative Binomial, and Hermite, were integrated. Each addresses specific characteristics such as overdispersion or excess zeros, enabling more robust and accurate model fitting tailored to the dataset's unique properties. Table 3 presents the goodness of fit results and parameter estimates for ARMA and INARMA models applied to South Bengkulu and Kepahiang, where "n.a." indicates that the model estimation failed to converge. Among the models, HER-INARMA(1,1) consistently shows the best fit across all regencies, with the lowest AIC and BIC values, indicating its effectiveness in capturing the data's characteristics. For South Bengkulu, the ARMA(1,1) model has a higher AIC and BIC compared to HER-INARMA(1,1), which includes key parameters such as $\mu = 0.386$, $\alpha = 0.592$, $\beta = 0.538$, and $\theta = 0.923$. For Kepahiang Regency, both the HER-INARMA(1,1) and POI-INARMA(1,1) models effectively captured the data's characteristics, with the Hermite and Negative Binomial distributions offering robust solutions for distinct data structures.

**Table 3.** Goodness of fit and parameter estimates for ARMA models and modifications

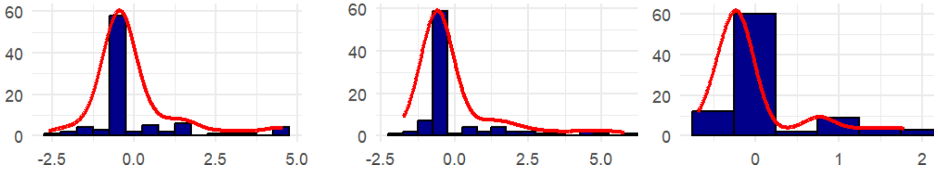
| Regency | Model | Goodness of Fit | | | Parameters | | | |
|---|---|---|---|---|---|---|---|---|
| | Applied | Log-llk | AIC | BIC | $\mu$ | $\alpha$ | $\beta$ | $\theta$ |
| South Bengkulu | ARMA(1,1) | -157.695 | 323.391 | 344.015 | 0.675$^{*}$ | 0.775$^{***}$ | -0.257$^{***}$ | n.a |
| | POI-INARMA(1,1) | n.a | n.a | n.a | n.a | n.a | n.a | n.a |
| | NB-INARMA(1,1) | n.a | n.a | n.a | n.a | n.a | n.a | n.a |
| | HER-INARMA(1,1) | -91.998 | 193.995 | 228.993 | 0.386$^{***}$ | 0.592$^{***}$ | 0.538$^{***}$ | 0.923$^{***}$ |
| Kepahiang | ARMA(1,1) | -75.763 | 159.524 | 180.149 | 0.277$^{***}$ | 0.648$^{***}$ | -0.375 | n.a |
| | POI-INARMA(1,1) | -56.533 | 121.067 | 149.065 | 0.090$^{***}$ | 0.775$^{***}$ | 0.748$^{***}$ | n.a |
| | NB-INARMA(1,1) | -56.054 | 122.109 | 157.107 | 0.148$^{***}$ | 0.551$^{**}$ | 0.495$^{***}$ | 3.333 |
| | HER-INARMA(1,1) | -54.713 | 119.425 | 154.424 | 0.152$^{***}$ | 0.528$^{***}$ | 0.478$^{***}$ | 0.577$^{***}$ |

*Note*: $\mu$ is the mean parameter, $\alpha$ the autoregressive parameter, $\beta$ the moving-average parameter, and $\theta$ a model-specific parameter such as overdispersion or shape. AIC and BIC measure model fit, with lower values indicating better fit. "n.a." indicates model estimation did not converge. Parameters marked with asterisks are statistically significant: $^{***}$ at the 1% level, $^{**}$ at the 5% level, and $^{*}$ at the 10% level.

However, due to conflicting AIC and BIC values between the two models, a likelihood ratio test was conducted. Since the LR statistic was smaller than the chi-squared critical value of 3.84 at the 5% significance level, the null hypothesis stating that POI-INARMA(1,1) fits as well as HER-INARMA(1,1) was accepted. Accordingly, POI-INARMA(1,1) was selected as the best model for Kepahiang Regency, highlighting the importance of aligning model structure with the statistical nature of rare and sporadic ecological events such as Rafflesia arnoldii blooming.

In light of these findings, and to further demonstrate the advantages of the proposed models, a comparison was also made with traditional time series approaches. The AR(1) and ARMA(1,1) models, which assume continuous, normally distributed time series, were applied to the same datasets across regencies. As shown in Tables 2 and 3, these models consistently yielded higher AIC and BIC values compared to the discrete-valued alternatives, highlighting their limited ability to accommodate key features such as non-negativity, overdispersion, and excess zeros. By contrast, the NB-INAR, ZINAR, and NB-INARMA-INGARCH models employed in this study provided superior statistical fit and greater ecological interpretability, particularly in capturing the irregular, count-based dynamics of Rafflesia blooming events (Ferland et al., 2006; Kim & Jun, 2017; Weiß, 2021). Furthermore, these models enhance decision-making in conservation contexts by enabling reliable forecasting of rare events, which is vital for resource planning, visitor management, and biodiversity protection. This comparative analysis confirms the methodological advantages of discrete-valued time series models over traditional ARMA-based approaches for modeling rare ecological phenomena.

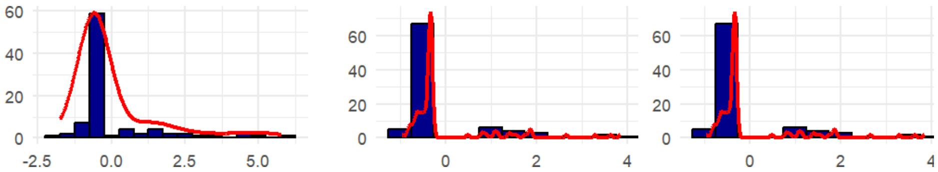
**Figure 4.** Histogram of residuals for each model applied across the studied regencies.



INAR(1)
South Bengkulu

NB-INAR(1)
South Bengkulu

INAR(1)
Kepahiang



NB-INAR(1)
Kepahiang

POI-INARMA(1,1)
Kepahiang

NB-INARMA(1,1)
Kepahiang

*Note*: The histograms display the distribution of residuals for each fitted discrete-valued time series model across the studied regencies. The plots are used to assess model adequacy and examine whether the residuals exhibit approximately random behavior.

After selecting the best model for each approach, the next essential step is model diagnostic testing, which involves a series of hypothesis tests to verify that the model satisfies the necessary assumptions and performs effectively. Figure 4 displays histograms of residuals for each model applied across South Bengkulu and Kepahiang, providing an initial check of key assumptions in the Model Diagnostic Testing stage. These histograms help assess the normality, independence, and homoscedasticity of the residuals. If the residuals are symmetrically distributed in a bell shape, it suggests that the normality assumption is likely met. On the other hand, if skewness or heavy tails are evident, the model may not be fully capturing the data's structure. Similarly, if clear patterns are visible in the histograms, it could indicate autocorrelation or non-constant variance, highlighting areas where the model needs improvement.

The next step in the diagnostic process involves applying formal tests to rigorously verify these assumptions and ensure model performance. The Ljung-Box Test will check for autocorrelation in the residuals, confirming their independence. The Jarque-Bera Test will assess whether the residuals follow a normal distribution, formally testing the assumption of normality. Finally, the Lagrange-Multiplier (LM) Test will examine whether the residuals exhibit heteroscedasticity, ensuring that their variance remains constant over time. These tests will provide a more detailed validation of the results observed in the histograms and help ensure that the models are robust, accurate, and well-specified for the data.

The first test, the Ljung-Box test, will check for significant autocorrelation in the residuals, ensuring that the model has captured all underlying patterns in the data. Next, the Jarque-Bera test will assess the residual distribution for normality by examining skewness and kurtosis to identify any deviations that could indicate model misspecification. Lastly, the Lagrange-Multiplier test is used to detect heteroscedasticity, ensuring that the variance of the residuals remains constant over time. If any of these tests reveal issues, the model may need refinement or adjustment to ensure reliable and valid results.

**Table 4.** Model diagnostic testing: verifying assumptions and ensuring model performance

| Regency | Model Applied | Ljung-Box Test | | Jarque Bera Test | | Lagrange Multiplier Test | |
|---|---|---|---|---|---|---|---|
| | | $Q_{test}$ | p-value | $JB_{test}$ | p-value | $LM_{test}$ | p-value |
| South Bengkulu | INAR(1) | 1.042 | 0.310 | 100.130 | 0.000 | 26.313 | 0.001 |
| | NB-INAR(1) | 13.271 | 0.000 | 164.800 | 0.000 | 28.875 | 0.004 |
| | POI-INARMA(1,1) | n.a | n.a | n.a | n.a | n.a | n.a |
| | NB-INARMA(1,1) | n.a | n.a | n.a | n.a | n.a | n.a |
| Kepahiang | INAR(1) | 0.131 | 0.720 | 68.685 | 0.000 | 13.315 | 0.350 |
| | NB-INAR(1) | n.a | n.a | n.a | n.a | n.a | n.a |
| | POI-INARMA(1,1) | 0.149 | 0.690 | 146.910 | 0.000 | 9.599 | 0.650 |
| | NB-INARMA(1,1) | 0.149 | 0.690 | 146.850 | 0.000 | 9.596 | 0.650 |

*Note*: The Ljung Box statistic ($Q_{test}$) assesses whether residuals are free from autocorrelation ($H_0$: residuals are uncorrelated). The Jarque Bera statistic ($JB_{test}$) tests for normality of residuals ($H_0$: residuals are normally distributed). The Lagrange Multiplier statistic ($LM_{test}$) evaluates homoscedasticity ($H_0$: constant residual variance). A p-value below 0.05 indicates rejection of the null hypothesis at the 5% significance level. "n.a." denotes that the model estimation failed to converge.

Table 4 shows the results of diagnostic tests for models applied to South Bengkulu, and Kepahiang regencies, where "n.a." indicates that the model estimation failed to converge. In South Bengkulu, the INAR(1) and NB-INAR(1) models exhibit heteroscedasticity in residual variance, as indicated by the Lagrange Multiplier test, but both fail the Ljung-Box test, suggesting autocorrelation in the residuals. The POI-INARMA(1,1) and NB-INARMA(1,1) models have missing values, implying incomplete evaluations. The POI-INARMA(1,1) and NB-INARMA(1,1) models exhibit heteroscedasticity, failing the LM test. In Kepahiang, the INAR(1) and POI-INARMA(1,1) models pass the Ljung-Box test but fail the Jarque-Bera test, indicating non-normality of residuals. Despite these issues, the Ljung-Box test is satisfied for most models, suggesting no autocorrelation. Given the diagnostic test results, all models show at least one issue, such as autocorrelation, heteroscedasticity, or non-normality. To address these concerns, the models will be reanalyzed with INGARCH to correct for heteroscedasticity and autocorrelation.

**Table 5.** Goodness of fit and parameter estimates for the INARMA-INGARCH model

| Regency | Model Applied | Goodness of Fit | | | Parameters | | |
|---|---|---|---|---|---|---|---|
| | | Log-Likelihood | AIC | BIC | ω | $\alpha$ | $\beta$ |
| South Bengkulu | INAR(1)-INGARCH(1,1) | -75.514 | 157.028 | 164.528 | $0.007^{**}$ | $0.329^{***}$ | $0.643^{***}$ |
| | NB-INAR(1)-INGARCH(1,1) | -108.377 | 222.753 | 230.253 | 0.452 | $0.522^{***}$ | 0.108 |
| | POI-INARMA(1,1)-INGARCH(1,1) | n.a | n.a | n.a | n.a | n.a | n.a |
| | NB-INARMA(1,1)-INGARCH(1,1) | n.a | n.a | n.a | n.a | n.a | n.a |
| Kepahiang | INAR(1)-INGARCH(1,1) | -54.935 | 115.870 | 123.369 | $0.015^{*}$ | $0.121^{*}$ | $0.822^{***}$ |
| | NB-INAR(1)-INGARCH(1,1) | n.a | n.a | n.a | n.a | n.a | n.a |
| | POI-INARMA(1,1) INGARCH(1,1) | -88.288 | 182.577 | 190.076 | 0.059 | $0.206^{**}$ | $0.694^{***}$ |
| | NB-INARMA(1,1) INGARCH(1,1) | -89.837 | 187.674 | 197.917 | 0.058 | $0.205^{**}$ | $0.695^{**}$ |

*Note*: ω is the intercept parameter, $\alpha$ is the autoregressive parameter, and $\beta$ is the moving-average parameter of the conditional variance in the INGARCH component. AIC and BIC measure model fit, with lower values indicating a better fit. "n.a." indicates that the model estimation failed to converge or the parameter was not applicable for the corresponding model. Parameters marked with asterisks are statistically significant: $^{***}$ at the 1% level, $^{**}$ at the 5% level, and $^{*}$ at the 10% level.

Table 5 presents the results of the goodness of fit and parameter estimates for the INGARCH models applied to the data from South Bengkulu and Kepahiang regencies. The table includes the log-likelihood, AIC, BIC, and parameter estimates for each model. These criteria help assess each model's performance, where lower AIC and BIC values indicate a better fit, and the parameters $\omega$, $\alpha$, and $\beta$ reflect the dynamics of the data in terms of the INGARCH model structure.

In South Bengkulu, the INAR(1)-INGARCH(1,1) model has a lower AIC (157.028) and BIC (164.528) than the NB-INAR(1)-INGARCH(1,1) model, indicating a better fit for the former. However, not avaliable (n.a.) for POI-INARMA(1,1) and NB-INARMA(1,1) suggest these models could not be evaluated. Finally, for Kepahiang, the INAR(1)-INGARCH(1,1) model has the lowest AIC and BIC, suggesting it provides the best fit among the models evaluated for this region.

The parameter estimates $\omega$, $\alpha$, and $\beta$ reflect the autoregressive and moving-average components of the model. For instance, in South Bengkulu, the INAR(1)-INGARCH(1,1) model shows $\omega = 0.007$, $\alpha = 0.329$, and $\beta = 0.643$, suggesting a moderate autoregressive effect and a stronger moving-average effect. Each region's best model selection will ultimately help in capturing the patterns and characteristics specific to that regency's data. We have applied three distinct approaches to model our discrete time series data: INAR, INARMA, and INARMA-INGARCH. Each model was chosen to capture different characteristics of the data, ranging from simpler autoregressive structures to models that account for moving averages and time-varying volatility. As we approach the final stage, we compare the goodness of fit of the best models from each approach for each regency. This comparison will utilize metrics such as AIC, BIC, and log-likelihood to assess which model offers the most accurate fit to the data. The model with the best performance for each regency will then be selected to forecast the test dataset, providing a foundation for evaluating the model's predictive accuracy. This comparative approach allows us to identify the model most suitable for capturing the underlying structure of the time series in each region, ultimately leading to more reliable forecasts.

**Table 6.** Recap of model selection for three discrete-valued time series approaches

| Regency | Model Applied | Goodness of Fit | | | ADF Test Error | | |
|---|---|---|---|---|---|---|---|
| | | Log-Likelihood | AIC | BIC | $ADF_{test}$ | p-value | Conclusion |
| South Bengkulu | ZINAR(1) | -92.562 | 191.100 | 198.600 | -3.124 | 0.113 | I(1) |
| | HER - INARMA(1,1) | -91.998 | 193.995 | 228.993 | -3.054 | 0.142 | I(1) |
| | INAR(1) - INGARCH(1,1) | -75.514 | **157.028** | **164.528** | **-3.443** | **0.053** | **I(0)** |
| Kepahiang | ZINAR(1) | -56.487 | 118.900 | 126.400 | -3.202 | 0.092 | I(1) |
| | POI - INARMA(1,1) | -56.533 | 121.067 | 149.065 | -4.148 | 0.010 | I(0) |
| | INAR(1) - INGARCH(1,1) | -54.935 | **115.870** | **123.369** | **-3.556** | **0.041** | **I(0)** |

*Note*: Log-Likelihood refers to the maximum of the log-likelihood function. AIC and BIC measure model fit, where lower values indicate a better model fit. The selected model for each regency is the one with the lowest information criterion among the alternatives. The bolded values indicate the lowest score for each criterion within a regency, representing the selected model. The ADF test checks stationarity ($H_0$: unit root, non-stationary). "I(0)" indicates stationarity at the level, "I(1)" after first differencing.

Table 6 provides a summary of the model selection results for three different discrete-valued time series approaches applied to data from South Bengkulu and Kepahiang regencies. For each model, the goodness of fit is evaluated using log-likelihood, Akaike Information Criterion (AIC), and Bayesian Information Criterion (BIC). Lower AIC and BIC values indicate a better model fit, helping us identify the most suitable model for each regency. In South Bengkulu, the INAR(1)- INGARCH(1,1) model demonstrates the best fit, with the lowest AIC (157.028) and BIC (164.528) among the tested models, outperforming the ZINAR(1) and HER-INARMA(1,1) models. In Kepahiang, the INAR(1)-INGARCH(1,1) model is the top choice, with an AIC of 115.870 and a BIC of 123.369, outperforming the ZINAR(1) and POI - INARMA(1,1) models. Overall, the INAR(1)-INGARCH(1,1) model is consistently selected as the best-fitting approach in South Bengkulu and Kepahiang, demonstrating the utility of combining INGARCH components with INAR models in addressing the data's characteristics across different regions. In addition, both selected specifications satisfy the error-stationarity assumption, as confirmed by the ADF test on model residuals, thereby ensuring the admissibility of the models for subsequent inference.

To ensure the reliability of the best model, we will apply the bootstrap technique. Bootstrapping is a resampling method that estimates the sampling distribution of a statistic by repeatedly sampling from the original data. In this case, we will perform 50 bootstrap repetitions for each selected model, using the parameters listed in Table 5. By using bootstrapping, we aim to achieve more robust estimates and improved variability in the model parameters, ultimately resulting in more accurate and reliable predictions for each regency. The bootstrap results, based on the parameters derived from the original data, are presented below.

**Table 7.** Bootstrap results of the selected model parameters at a 95% confidence level

| Regency | Selected Model | Estimation Parameters Model | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | MLE | 95% Confidence Interval Bootstrap | | | | |
| | | | | Lower | Mean | Upper | S.E. |
| South Bengkulu | INAR(1)-INGARCH(1,1) | $\omega$ | 0.007 | -0.036 | 0.029 | 0.094 | 0.033 |
| | | $\alpha$ | 0.329 | 0.025 | 0.361 | 0.697 | 0.169 |
| | | $\beta$ | 0.643 | 0.070 | 0.456 | 0.841 | 0.194 |
| Kepahiang | INAR(1)-INGARCH(1,1) | $\omega$ | 0.015 | -0.074 | 0.088 | 0.249 | 0.081 |
| | | $\alpha$ | 0.121 | 0.092 | 0.276 | 0.460 | 0.093 |
| | | $\beta$ | 0.822 | 0.386 | 0.655 | 0.924 | 0.135 |

*Note*: MLE denotes the maximum likelihood estimate of each parameter. The bootstrap results provide the 95% confidence interval (Lower, Mean, Upper) and the standard error (S.E.) based on resampling. Parameters are defined as follows: ω represents the intercept term, $\alpha$ is the autoregressive parameter, and $\beta$ is the moving-average parameter in the conditional variance equation. Estimates are reported for the selected model in each regency.

Table 7 presents the bootstrap results for the models applied to South Bengkulu and Kepahiang. South Bengkulu and Kepahiang used the same model. The confidence intervals (CIs) and standard errors indicate the precision of the parameter estimates. South Bengkulu shows generally narrower CIs and smaller standard errors, suggesting greater estimate stability. Kepahiang also exhibits relatively tight CIs, indicating consistent results. These findings imply reliable model estimates in South Bengkulu and Kepahiang. Overall, the bootstrap results underscore regional differences in estimate reliability.

In the final phase, forecasts of Rafflesia arnoldii blooms in South Bengkulu and Kepahiang were generated using the best-performing models identified earlier. The INAR(1)-INGARCH(1,1) model was applied to South Bengkulu and Kepahiang based on optimal AIC, BIC, and log-likelihood values. Following bootstrap validation, these models were used to produce six-month forecasts, offering valuable insights for conservation and nature-based tourism in Bengkulu Province, where Rafflesia arnoldii holds significant ecological and economic value.

**Table 8.** Forecasting results based on the selected time series model

| Period | South Bengkulu INAR(1)-INGARCH(1,1) | | Kepahiang INAR(1)-INGARCH(1,1) | |
|---|---|---|---|---|
| | Actual | Forecasting | Actual | Forecasting |
| July 2023 | 0 | 2 | 5 | 0 |
| August 2023 | 0 | 1 | 1 | 0 |
| September 2023 | 0 | 1 | 0 | 0 |
| October 2023 | 0 | 0 | 1 | 0 |
| November 2023 | 0 | 1 | 2 | 1 |
| December 2023 | 0 | 2 | 1 | 0 |
| RMSE | | 1.354 | | 2.198 |
| MAAPE (%) | | 1.571 | | 1.558 |

*Note*: Forecasting results are based on the selected model for each regency: INAR(1)-INGARCH(1,1) for South Bengkulu and INAR(1)-INGARCH(1,1) for Kepahiang. Actual refers to observed values, while Forecasting indicates the model's predicted values for each month. RMSE denotes the root mean square error, and MAAPE (%) represents the mean arctangent absolute percentage error, both used to assess forecast accuracy (lower values indicate better predictive performance).

Table 8 compares actual and forecasted bloom counts from July to December 2023 across the three regencies, revealing regional and temporal variations in accuracy. In South Bengkulu, the largest discrepancy occurred in July (forecast = 2; actual = 0), while October showed better alignment. The RMSE of 1.354 and MAAPE of 1.571 suggest moderate forecasting performance. Kepahiang exhibited the largest error in July (forecast = 0; actual = 5), with the highest RMSE (2.198), but the lowest MAAPE (1.558), suggesting better relative accuracy. Overall, Kepahiang showed superior relative accuracy, highlighting opportunities for further model refinement.

In summary, the modeling results across the three regencies provide both statistical validation and ecological insight. Overdispersion and significant zero inflation in South Bengkulu reflect the highly irregular and clustered blooming behavior of Rafflesia arnoldii, likely driven by habitat fragmentation, host vine variability, climatic factors, and human disturbance. The NB-INAR and ZINAR models effectively capture these episodic blooms and long dormant periods, which conventional models often overlook. In contrast, the superior performance of the Poisson-INARMA model in Kepahiang indicates a more stable and predictable blooming pattern, possibly supported by more homogeneous and less disturbed ecological conditions. This regional variation highlights the ecological heterogeneity of Rafflesia habitats and underscores the importance of location-specific model selection that integrates statistical diagnostics with ecological context. Practically, forecasts from the best-fitting models can inform conservation management, including monitoring schedules, resource allocation, and ecotourism planning, while reducing ecological disturbance. Overall, this study demonstrates that advanced discrete time-series models can translate sporadic ecological data into reliable forecasts, supporting data-driven conservation planning for rare species such as Rafflesia arnoldii.

## 5. Conclusions and Future Research

The motivation behind this study stems from the urgent need to protect Rafflesia arnoldii, one of the world’s rarest and most ecologically unique flowering plants. Despite its biological importance and endangered status, its blooming patterns remain largely unpredictable, limiting the effectiveness of conservation programs, ecotourism development, and scientific monitoring. In this context, constructing a robust probabilistic forecasting model is not only timely but also essential. Such a model empowers academics and practitioners, including ecologists, conservationists, policymakers, and local stakeholders, with a data-driven tool to anticipate rare blooming events and plan interventions accordingly. This study presents a comprehensive framework for forecasting the monthly blooming frequency of Rafflesia arnoldii using discrete-valued time series models, an approach seldom applied in ecological research. Through rigorous model selection and validation, we found that the INAR(1)-INGARCH(1,1) model provided the best fit for South Bengkulu and Kepahiang. These models significantly outperformed the traditional ARMA(1,1) model, which assumes continuity and normality, and is therefore unsuitable for modeling count data characterized by overdispersion and zero inflation.

Several important insights can be drawn from these findings. First, the application of discrete-valued models reveals temporal dynamics in blooming frequencies that would otherwise remain hidden.

Differences in model performance across regencies may reflect varying ecological conditions, human interference, or environmental factors. Second, the forecasting outputs generated by these models can directly inform decision-making in the field, for example, scheduling conservation work, regulating visitor access, and aligning ecotourism efforts with anticipated bloom cycles. The predictive capabilities offered here bridge the gap between statistical modeling and real-world biodiversity management.

This research offers a novel contribution to the literature by integrating INAR, ZINAR, and INARMA-INGARCH models into the context of rare plant bloom forecasting. Unlike previous ecological studies that rely primarily on static or descriptive methods, our work introduces a dynamic, probabilistic modeling approach tailored to count-based ecological data. This integration of statistical innovation and ecological forecasting provides a replicable framework for conservation and decision-support systems in other biodiversity contexts. Additionally, the forecasting results presented serve as practical tools for decision-makers in conservation agencies, local governments, and environmental planning. By providing predictive insights under uncertainty, this study contributes directly to Decision Sciences, particularly in the domains of operations research, statistical forecasting, and sustainability-focused decision modeling. Despite its contributions, this study has several limitations. The analysis is limited to four regencies in Bengkulu Province, and the models rely solely on historical count data of blooming events, without incorporating exogenous environmental variables such as rainfall, temperature, or host plant density. These factors may influence blooming patterns and should be integrated into future modeling efforts to enhance accuracy (Hardin & Hilbe, 2018; Zhang et al., 2014). Additionally, the forecasting results provide practical support for decision-making in conservation and environmental planning. By translating ecological uncertainty into probabilistic forecasts, this study contributes to Decision Sciences, particularly in statistical forecasting and sustainability-oriented decision modeling, and supports strategic planning in natural resource management.

Future research is recommended to expand the scope of data collection, both spatially and temporally, and to explore spatial-temporal extensions of INARMA-INGARCH models. Incorporating zero-inflated or hurdle models (Weiß et al., 2016) and covariate-based predictors could lead to more comprehensive ecological modeling. In addition, future extensions may consider integer-valued time series frameworks that explicitly accommodate non-stationary count data, such as those based on signed thinning operators or generalized INAR-type processes. These methodological developments will not only improve predictive accuracy but also provide deeper insights into the population dynamics of *Rafflesia arnoldii*, supporting more effective and localized conservation strategies.

## Acknowledgment

The authors sincerely thank the reviewers for their constructive comments, which greatly improved this manuscript. Financial support from the Faculty of Mathematics and Natural Sciences, The University of Bengkulu, under the *Penelitian Kerjasama dengan Dunia Usaha, Industri, Lembaga Riset, dan Pemerintah Daerah* scheme (Grant No. 2698/UN30.12/HK/2024) is gratefully acknowledged.

## References

Akaike, H. (1974). A New Look at The Statistical Model Identification. *IEEE Transactions on Automatic Control*, *19*(6), 716 723. https://doi.org/10.1109/TAC.1974.1100705

Al-Osh, M. A., & Alzaid, A. A. (1987). First-Order Integer-Valued Autoregressive (INAR(1)) process. *Journal of Time Series Analysis*, *8*(3), 261 275. https://doi.org/10.1111/j.1467-9892.1987.tb00438.x

Aly, E.-E. A. A., & Nadjib, B. (1994). On Some Integer-Valued Autoregressive Moving Average Models. *Journal of Multivariate Analysis*, *50*(1), 132 151. https://doi.org/10.1006/jmva.1994.1038

Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2016). *Time Series Analysis Forecasting and Control (5th ed.).* John Wiley & Sons, Inc. https://doi.org/10.1111/jtsa.12194

Cameron, A. C., & Trivedi, P. K. (1990). Regression-Based Tests for Overdispersion in The Poisson Model. *Journal of Econometrics*, *46*(3), 347 364. https://doi.org/10.1016/0304-4076(90)90014-K

Chen, C., Davis, R. A., & Brockwell, P. J. (1996). Order Determination for Multivariate Autoregressive Processes Using Resampling Methods. *Journal of Multivariate Analysis*, *57*(2), 175 190. https://doi.org/10.1006/jmva.1996.0028

Collins, N. M., Sayer, J. A., & Whitmore, T. C. (Eds.). (1991). *The conservation atlas of tropical forests*: Asia and the Pacifics. Palgrave Macmillan London. https://doi.org/10.1007/978-1-349-12030-7

Dungey, M., Martin, V. L., Tang, C., & Tremayne, A. (2019). A Threshold Mixed Count Time Series Model: Estimation and Application. *Studies in Nonlinear Dynamics and Econometrics*, *24*(2), 1 18. https://doi.org/10.1515/snde-2018-0029

Engle, R. F. (1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica*: *Journal of the Econometric Society*, 987-1007. https://doi.org/10.2307/1912773

Ferland, R., Latour, A., & Oraichi, D. (2006). Integer-Valued GARCH Process. *Journal of Time Series Analysis*, *27*(6), 923-942. https://doi.org/10.1111/j.1467-9892.2006.00496.x

Gujarati, D. N., & Porter, D. C. (2009). *Basic Econometrics*. McGraw-Hill/Irwin.

Hardin, J. W., & Hilbe, J. M. (2018). *Generalized Linear Models and Extensions*. Stata press.

Harvey, A. C., & Fernandes, C. (1989). Time Series Models for Count or Qualitative Observations. *Journal of Business & Economic Statistics*, 7(4), 407-417. https://doi.org/10.2307/1391639

Kim, H. Y., & Park, Y. (2008). A non-stationary integer-valued autoregressive model. *Statistical papers*, *49*(3), 485-502. https://doi.org/10.1007/s00362-006-0028-1

Kim, J. M., & Jun, S. (2017). Integer-Valued GARCH Processes for Apple Technology Analysis. *Industrial Management & Data Systems*, *117*(10), 2381-2399. http://dx.doi.org/10.1108/IMDS-01-2017-0023

Kutoyants, Y. A. (2023). *Introduction to the Statistics of Poisson Processes and Applications*. Springer Nature. https://doi.org/10.1007/978-3-031-37054-0

Mohammadipour, M., & Boylan, J. E. (2012). Forecast horizon aggregation in integer autoregressive moving average (INARMA) models. *Omega*, 40(6), 703-712. https://doi.org/10.1016/j.omega.2011.08.008

Nie, H., Liu, G., Liu, X., & Wang, Y. (2012). Hybrid of ARIMA and SVMs for Short-Term Load Forecasting. *Energy Procedia*, 16, 1455 1460. https://doi.org/10.1016/j.egypro.2012.01.229

Olofsson, P., & Andersson, M. (2012). *Probability, Statistics, and Stochastic Processes*. John Wiley & Sons. https://doi.org/10.1002/9781118231296

Priatna, D. R., Zuhud, E. A., & Alikodra, H. S. (1989). Kajian Ekologis Rafflesia Patma Blume di Cagar Alam Leuweung Sancang Jawa Barat. *Media Konservasi*, *2*(2). https://dx.doi.org/10.29243/medkon.2.2.%p

Reboredo, J. C., Queiruga, J. R. B., Ferreiro, J. O., & Santias, F. R. (2023). Forecasting Emergency Department Arrivals Using INGARCH Models. *Health Economics Review*, *13*(51). https://doi.org/10.1186/s13561-023-00456-5

Sammaknejad, N., Zhao, Y., & Huang, B. (2019). A review of the expectation maximization algorithm in data-driven process identification. *Journal of Process Control*, 73, 123 136. https://doi.org/10.1016/j.jprocont.2018.12.010

Schwarz, G. (1978). Estimating The Dimension of a Model. *The Annals of Statistics*, 461-464. https://doi.org/10.1214/aos/1176344136

Smith, J. Q. (1979). A Generalization of the Bayesian Steady Forecasting Model. *Journal of the Royal Statistical Society: Series B (Methodological)*, *41*(3), 375-387. https://doi.org/10.1111/j.2517-6161.1979.tb01092.x

Susatya. (2011). Raflesia Pesona Bunga Terbesar di Dunia. Direktorat Kawasan Konservasi dan Bina Hutan Lindung Dengan Pendanaan dari DIPA TA 2011.

Tijms, H. C. (2003). *A First Course in Stochastic Models*. John Wiley Sons Ltd. https://doi.org/10.1002/047001363X

Tolod, J. R., Galindon, J. M. M., Atienza, R. R., Duya, M. V., Fernando, E. S., & Ong, P. S. (2021). Flower and fruit development and life history of Rafflesia consueloae (Rafflesiaceae). *Philippine Journal of Science*, 150(S1), 321 334. https://doi.org/10.56899/150.S1.23

Tsay, R. S. (2010). *Analysis of Financial Time Series*. John Wiley & Sons. https://doi.org/10.1002/9780470644560

Wei, W. W. S. (2006). Time Series Analysis: Univariate and Multivariate Methods (2nd ed.). Pearson Education Inc.

Weiss, C. H. (2018). *An Introduction to Discrete-Valued Time Series*. John Wiley & Sons. https://doi.org/10.1002/9781119097013

Weiß, C. H. (2021). Stationary count time series models. *Wiley Interdisciplinary Reviews: Computational Statistics*, 13(1), e1502. https://doi.org/10.1002/wics.1502

Weiß, C. H., Feld, M. H.-J. M., Mamode Khan, N., & Sunecher, Y. (2019). INARMA Modeling of Count Time Series. *Stats*, *2*(2), 284 320. https://doi.org/10.3390/stats2020022

Weiß, C. H., Homburg, A., & Puig, P. (2016). Testing for Zero Inflation and Overdispersion in INAR(1) Models. *Springer*. https://doi.org/10.1007/s00362-016-0851-y

Yang, Z., Hardin, J. W., & Addy, C. L. (2009). A Score Test for Overdispersion in Poisson Regression Based on the Generalized Poisson-2 Model. *Journal of Statistical Planning and Inference*, *139*(4), 1514-1521. http://dx.doi.org/10.1016/j.jspi.2008.08.018

Zeger, S. L., & Qaqish, B. (1988). Markov Regression Models for Time Series: A Quasi-Likelihood Approach. *Biometrics*, 1019-1031. https://doi.org/10.2307/2531732

Zhang, H., Liu, Y., & Li, B. (2014). Notes on Discrete Compound Poisson Model with Applications to Risk Theory. *Insurance: Mathematics and Economics*, *59*, 325-336. https://doi.org/10.1016/j.insmatheco.2014.09.012.