

Zero-Inflated Integer Autoregressive (ZINAR) Approach in Modeling Major Earthquakes in Sumatra Region

Pendekatan *Zero-Inflated Integer Autoregressive* (ZINAR) dalam Memodelkan Gempa Bumi Besar di Wilayah Sumatra

Indah Wahyuliani¹, Jose Rizal^{2*}, Sigit Nugroho³

^{1,2,3}*Department of Mathematics, Faculty of Mathematics and Natural Sciences, The University of Bengkulu, Indonesia*

E-mail: ¹indahwahyuliani15@gmail.com, ²jrizar04@unib.ac.id, ³snugroho@unib.ac.id

**Corresponding Author*

Abstract

Sumatra is one of the regions in Indonesia with the highest seismic activity due to the convergence of the Indo-Australian and Eurasian plates. Major earthquakes with magnitudes of more than 7 pose serious risks, making it essential to understand their frequency for effective disaster mitigation planning. This study aims to model the frequency of major earthquakes in Sumatra using the Zero-Inflated Integer Autoregressive (ZINAR) model, which accommodates discrete data with excess zeros and temporal dependence. The analysis indicates significant overdispersion and zero inflation, leading to the selection of the ZINAR model as the most suitable approach compared to alternative models such as INAR, NB-INAR, and NB-ZINAR. The ZINAR model demonstrates superior performance in capturing the occurrence patterns of large earthquakes and effectively identifies non-event periods, although its predictive accuracy for actual event timing remains limited. Overall, the ZINAR model proves effective for modeling complex earthquake data and provides valuable insights to support disaster mitigation efforts in the Sumatra region.

Keywords: Earthquake, Overdispersion, Zero-inflated, Integer Autoregressive model.

Abstrak

Sumatra merupakan salah satu wilayah di Indonesia dengan aktivitas seismik tertinggi akibat konvergensi Lempeng Indo-Australia dan Eurasia. Gempa besar dengan magnitudo lebih besar dari 7 menimbulkan risiko serius, sehingga penting untuk memahami frekuensinya sebagai dasar perencanaan mitigasi bencana. Penelitian ini bertujuan untuk memodelkan frekuensi gempa besar di Sumatra menggunakan model *Zero-Inflated Integer Autoregressive* (ZINAR), yang mampu menangani karakteristik data berupa nilai diskret, dominasi nol, dan ketergantungan temporal. Hasil analisis menunjukkan adanya overdispersi dan *zero inflation* yang signifikan, sehingga model ZINAR dipilih sebagai pendekatan paling sesuai dibandingkan model alternatif seperti INAR, NB-INAR, dan NB-ZINAR. Model ini menunjukkan kinerja terbaik dalam menggambarkan pola kejadian gempa besar dan mampu mengidentifikasi periode tanpa kejadian dengan baik, meskipun



prediksi terhadap waktu terjadinya gempa masih terbatas. Secara keseluruhan, model ZINAR efektif dalam memodelkan data gempa yang kompleks dan memberikan wawasan penting bagi upaya mitigasi bencana di wilayah Sumatra.

Kata kunci: Gempa bumi, Overdispersi, *Zero-inflated*, model Integer Autoregressive.

1. PENDAHULUAN

Indonesia, khususnya Pulau Sumatra, merupakan kawasan dengan tingkat aktivitas seismik yang sangat tinggi akibat interaksi tektonik antara lempeng Indo-Australia dan Eurasia. Mekanisme tumbukan dan subduksi antar lempeng tersebut menghasilkan rangkaian deformasi kerak bumi yang memicu kejadian gempa bumi besar sepanjang zona megathrust Sumatra. Meskipun kejadian gempa-gempa besar (magnitudo $M_w \geq 7$) relatif jarang terjadi, namun demikian setiap peristiwa memiliki potensi dampak bencana yang signifikan terhadap infrastruktur, keselamatan manusia, dan stabilitas wilayah [1].

Data frekuensi kejadian gempa bumi merupakan deret waktu diskrit (*count time series*) yang umumnya didominasi oleh nilai nol akibat jarang terjadinya kejadian gempa besar pada sebagian besar periode observasi, serta menunjukkan overdispersi, yakni varians yang melebihi nilai rata-rata, sebagaimana umum terjadi pada data kejadian langka atau ekstrem [2].

Dalam analisis deret waktu, model *Autoregressive Integrated Moving Average* (ARIMA) umum digunakan untuk memodelkan ketergantungan temporal melalui kombinasi komponen Autoregressive (AR) dan Moving Average (MA), yang merepresentasikan pengaruh nilai historis terhadap observasi saat ini [3]. Model ARIMA dirancang untuk data kontinu dan berdistribusi normal, sehingga kurang sesuai untuk deret waktu diskrit yang bersifat non-normal [4]. Sebagai alternatif, dikembangkan model *Integer Autoregressive* (INAR) yang secara khusus dirancang untuk data diskrit dengan mempertahankan sifat bilangan bulat melalui operator thinning [3]. Data hitung umumnya mengikuti distribusi Poisson atau Negative Binomial, tetapi sering menunjukkan kondisi overdispersi, yaitu varians yang melebihi rata-rata [5],[6].

Selain kondisi overdispersi, pemodelan frekuensi gempa bumi besar juga menghadapi permasalahan inflasi nol, yang ditandai oleh dominasi nilai nol akibat tidak terjadinya gempa besar pada sebagian besar periode observasi. Nilai nol tersebut dapat berasal dari proses kejadian itu sendiri maupun dari kondisi struktural yang meniadakan kemungkinan terjadinya peristiwa pada periode tertentu. Ketidakhadiran mekanisme eksplisit untuk menangani inflasi nol pada model diskrit konvensional, seperti INAR, berpotensi menimbulkan bias dan ketidaktepatan estimasi parameter [7]. Untuk mengatasi inflasi nol sekaligus mempertahankan ketergantungan temporal, dikembangkan model *Zero-Inflated Integer Autoregressive* (ZINAR), yang mengintegrasikan komponen *zero-inflated* dan struktur INAR dalam satu kerangka pemodelan [8]. Pendekatan ini memungkinkan pemisahan mekanisme pembentukan nol dan proses generasi nilai hitung positif secara sistematis. Sejumlah studi menunjukkan bahwa integrasi komponen autoregresif dalam model *zero-inflated* efektif dalam menangani data kejadian langka dengan variabilitas tinggi [9].

Berdasarkan kondisi seismotektonik Pulau Sumatra, pemodelan frekuensi kejadian gempa besar menjadi aspek krusial dalam kajian mitigasi kebencanaan. Secara statistik, data frekuensi kejadian gempa bumi besar memiliki karakteristik khas berupa deret waktu diskrit yang didominasi oleh nilai nol, menunjukkan overdispersi yang kuat, serta mengandung ketergantungan temporal antarperiode observasi. Dalam konteks tersebut, model ZINAR dapat dipandang sebagai pendekatan yang paling sesuai secara metodologis karena mengintegrasikan mekanisme *zero-inflated* untuk mengakomodasi kelebihan nilai nol dan struktur *Integer Autoregressive* untuk merepresentasikan ketergantungan temporal pada data hitung. Oleh karena itu, penelitian ini bertujuan untuk menerapkan dan mengevaluasi kinerja model ZINAR dalam memodelkan frekuensi kejadian gempa bumi besar di Pulau Sumatra. Selain menilai keunggulan model ZINAR dalam menangani inflasi nol dan

overdispersi, penelitian ini juga diharapkan dapat memberikan pemahaman yang lebih komprehensif terhadap dinamika kejadian gempa yang bersifat langka.

2. KAJIAN LITERATUR

2.1 Diagnosa Model Awal Data Deret Waktu Diskrit

2.1.1 Overdispersi

Kondisi overdispersi merujuk pada kondisi ketika variansi data melebihi nilai rata-rata sampel berdasarkan model distribusi Poisson yang mengasumsikan kesetaraan antara nilai rata-rata dan variansi. Fenomena ini umum dijumpai pada data hitung, khususnya pada kejadian langka dan ekstrem, dan berdampak signifikan terhadap kinerja model statistik berbasis distribusi Poisson [7]. Ketidakesesuaian asumsi tersebut dapat menyebabkan estimasi parameter yang bias dan tidak efisien, kesalahan standar yang tereduksi secara tidak realistis, serta kesimpulan inferensial yang menyesatkan akibat nilai p -value yang terlalu kecil [10]. Identifikasi overdispersi dapat dilakukan melalui beberapa pendekatan, antara lain dengan membandingkan nilai *deviance* atau *Pearson chi-square* dari distribusi Poisson terhadap derajat kebebasan, serta melalui rasio variansi terhadap rata-rata, di mana nilai rasio yang lebih besar dari satu mengindikasikan adanya overdispersi [10]. Apabila kondisi overdispersi terdeteksi, penggunaan model alternatif seperti distribusi Negative Binomial menjadi pendekatan yang umum, karena model ini memungkinkan variansi lebih besar daripada rata-rata [11]. Selain itu, model *Generalized Poisson* dan *Quasi-Poisson* juga dapat digunakan untuk mengakomodasi variabilitas berlebih melalui penambahan parameter dispersi atau relaksasi asumsi distribusi Poisson [6].

Dalam kondisi tertentu, transformasi data dapat digunakan untuk mereduksi overdispersi, meskipun pendekatan ini berpotensi memengaruhi interpretasi hasil analisis. Oleh karena itu, deteksi dan penanganan overdispersi merupakan langkah krusial untuk memperoleh estimasi yang akurat dan inferensi statistik yang andal. Pengujian overdispersi juga dapat dilakukan menggunakan Fisher test dengan memanfaatkan parameter dispersi yang diperoleh dari rasio antara variansi dan rata-rata data [12].

$$\hat{I}_d := \frac{S^2}{\bar{X}} \text{ dengan range } [0, \infty). \quad (2.1)$$

Keterangan dari variabel-variabel pada Persamaan (2.1) sebagai berikut: $S^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2$, dan $\bar{X} = \frac{1}{n} \sum_{i=1}^n (y_i)$. Terima H_0 jika \hat{I}_d bernilai 1 dan tolak H_0 jika \hat{I}_d bernilai tidak sama dengan satu. Sebuah observasi dikatakan equidispersi ketika nilai $\hat{I}_d = 1$, kemudian dikatakan overdispersi apabila nilai $\hat{I}_d > 1$, sedangkan untuk nilai $\hat{I}_d < 1$ dikatakan underdispersi. Setelah dilakukan pengujian deteksi overdispersi, selanjutnya dilakukan deteksi *Zero-Inflated*, untuk mengetahui apakah terdapat pengaruh dari angka nol yang berlebih atau lebih tinggi dari yang diharapkan

2.2.2 Zero-Inflated

Kondisi *zero-inflated* dalam pemodelan statistik merujuk pada persentase keberadaan nilai nol dalam jumlah yang lebih besar daripada yang diharapkan berdasarkan distribusi Poisson atau Binomial. Fenomena ini umum dijumpai pada data kejadian langka, di mana sebagian besar observasi tidak mengalami peristiwa, sementara sebagian kecil mengalami kejadian tertentu [7]. Ketidampampuan model distribusi standar (Poisson) dalam mengakomodasi kelebihan nilai nol dapat menyebabkan distorsi estimasi dan kegagalan model dalam menangkap pola data [13].

Deteksi kondisi *zero-inflated* umumnya dilakukan melalui perbandingan model distribusi standar dengan model zero-inflated, seperti *Zero-Inflated Poisson (ZIP)* atau *Zero-Inflated Negative Binomial (ZINB)*, serta melalui evaluasi kecocokan model dan eksplorasi visual data [14]. Untuk penanganannya, model ZIP mengombinasikan komponen binomial dalam memodelkan probabilitas

kemunculan nol dan komponen Poisson untuk kejadian non-nol [15], sedangkan model ZINB lebih sesuai untuk data yang sekaligus menunjukkan overdispersi [13][15]. Alternatif lain meliputi model Hurdle dan model-model campuran (*mixture models*), yang memisahkan mekanisme pembentukan nol dan proses kejadian positif [16]. Pemahaman dan penanganan kondisi *zero-inflated* secara tepat merupakan aspek penting dalam menghasilkan analisis statistik yang akurat. Deteksi *zero-inflated* dapat dilakukan menggunakan zero test berikut ini [17].

$$T = \frac{(\hat{p}_0 \exp(\bar{Y}) - 1)^2}{\exp(\bar{Y})(1 - \exp(\bar{Y})) - \bar{Y}} \quad (2.2)$$

dengan $\hat{p}_0 = \frac{1}{n} \sum_{t=1}^n \mathbb{1}_{\{Y_t=0\}}$. Hipotesis nol ditolak jika nilai statistik uji T pada Persamaan (2.2) lebih besar dari pada nilai tabel sebaran χ^2 dengan derajat bebas 1.

2.2 Autoregressive Process (AR)

Model *Autoregressive* (AR) adalah model statistik yang digunakan untuk menganalisis data deret waktu, dimana nilai-nilai dalam deret waktu diprediksi berdasarkan nilai-nilai sebelumnya dalam deret tersebut. Secara khusus, dalam model AR, nilai observasi saat ini atau beberapa observasi terakhir dijelaskan sebagai kombinasi linear dari nilai-nilai observasi sebelumnya ditambah dengan sebuah *error* atau *noise*. Asumsi dari model AR adalah varian yang bersifat homoskedastik. Model AR(p) dapat ditulis sebagai berikut [11]:

$$Y_t = \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \dots + \alpha_p Y_{t-p} + Z_t. \quad (2.3)$$

Dalam konteks ini, Y_t menyatakan nilai deret waktu pada periode ke- t , sedangkan parameter $\alpha_1, \alpha_2, \dots, \alpha_p$ adalah koefisien autoregresif yang menggambarkan besarnya kontribusi dari masing-masing lag terhadap nilai saat ini. Secara umum model AR biasanya digunakan untuk memodelkan data yang bersifat kontinu. Hal ini dikarenakan pada model AR biasanya diasumsikan peubah acak Z_t (*error*) mengikuti distribusi normal dengan rata-rata nol dan varians konstan. Asumsi ini lebih mudah dipenuhi dalam data kontinu dibandingkan dengan data diskrit, terutama data hitungan [18]. Sehingga model AR tidak tepat lagi digunakan untuk memodelkan data hitungan. Oleh karena itu dikembangkan proses AR pada data *integer* menjadi model INAR [5].

2.3 Model Integer Autoregressive (INAR)

Model *Integer Autoregressive* (INAR) dikembangkan untuk memodelkan deret waktu diskrit berbentuk bilangan bulat, khususnya data hitung yang umumnya mengikuti distribusi Poisson atau Negative Binomial [19]. Dibandingkan model AR yang dirancang untuk data kontinu [18], model INAR secara eksplisit mempertahankan sifat diskrit data sehingga lebih sesuai untuk analisis deret waktu berbasis hitungan. Oleh karena itu, model INAR lebih sesuai untuk memodelkan data *count* dan telah banyak diterapkan dalam analisis kejadian langka dan frekuensi peristiwa [6][19].

Misalkan X adalah bilang bulat *non negative* yang bernilai acak dan $\alpha \in [0,1]$, binomial *thinning operator* (disimbolkan dengan “ \circ ”) didefinisikan sebagai berikut [20]:

$$\alpha \circ Y = \sum_{i=1}^Y X_i. \quad (2.4)$$

Notasi $\{X_i\}$ merupakan sebuah barisan yang *independent, identifiically distributed* (i.i.d) variabel acak Bernoulli yang *independent* terhadap Y , dengan $P(X_i = 1) = \alpha$. Sehingga binomial *thinning operator* pada Persamaan (2.4) dapat dijabarkan sebagai berikut:

$$\alpha \circ Y = X_1 + X_2 + \dots + X_Y = \sum_{i=1}^Y X_i. \quad (2.5)$$

Binomial thinning digunakan untuk memilih kejadian-kejadian dari suatu proses acak diskrit (misalnya, proses dengan kejadian yang terjadi pada titik waktu tertentu) dengan probabilitas tertentu. Persamaan (2.5) menjelaskan bahwa hasil dari *thinning* ($\alpha \circ Y$) adalah jumlah dari variabel acak Bernoulli X_1, X_2, \dots, X_y , yang menunjukkan berapa banyak Y kejadian yang dipilih. Setiap X_i bernilai 1 jika kejadian ke- i dipilih atau dengan probabilitas α , dan bernilai 0 jika tidak dipilih atau dengan probabilitas $1 - \alpha$. Dengan demikian hasil dari penjumlahan ini ($\sum_{i=1}^y X_i$) adalah jumlah kejadian yang terpilih dari Y kejadian yang diproses melalui operator *thinning* dengan probabilitas α . Analisis data deret waktu yang melibatkan proses *thinning* adalah model *Integer Autoregressive INAR(p)*, yang diformulasikan sebagai berikut:

$$Y_t = \alpha_1 \circ Y_{t-1} + \dots + \alpha_p \circ Y_{t-p} + Z_t. \quad (2.6)$$

Untuk kasus $p = 1$, yakni INAR (1), Persamaan (2.6) dapat ditulis sebagai berikut:

$$Y_t = \alpha \circ Y_{t-1} + Z_t. \quad (2.7)$$

Variabel Z_t adalah urutan variabel acak bilangan bulat non-negatif yang terdistribusi independen dan identik, yang disebut inovasi, dengan $E(Z_t) = \mu < \infty$, $\text{Var}(Z_t) = \sigma^2 < \infty$ dan saling bebas terhadap Y_{t-1} , untuk semua t . Inovasi variabel Z_t diasumsikan sebagai distribusi $ZIP(\rho, \lambda)$, dengan fungsi massa peluang sebagai berikut:

$$P(Z_t = k) = \rho I_{\{0\}}(k) + (1 - \rho) \frac{e^{-\lambda} \lambda^k}{k!}, \quad k = 0, 1, 2, 3, \dots \quad (2.8)$$

Untuk menganalisis deret waktu diskrit dengan kelebihan nol, dalam penelitian ini akan difokuskan pada model ZINAR (1), dengan mempertimbangkan bahwa inovasi mengikuti kelas model *Zero-Inflated (ZI)*, yang disebut model ZINAR (1).

2.4 Model Zero-Inflated Integer Autoregressive (ZINAR)

Proses ZINAR merupakan rangkaian waktu autoregresif orde- p dengan nilai bulat, yang memiliki inovasi ZIP dengan $Z_t \sim ZIP(\rho, \lambda)$. Karena proses ZINAR(p) merupakan perluasan alami dari proses INAR(p), beberapa hasil penting tentang kondisi stasioner untuk proses INAR(p) dapat dimanfaatkan. Secara khusus, Jin-Guan dan Yuan menunjukkan bahwa kondisi stasioner dari INAR(p) dan AR(p) relatif sama [21]. Misal $\{Z_t\}$ merupakan variabel acak bernilai bulat non negatif yang independen dan identik, dengan $E(Z_t) = \mu$, $\text{Var}(Z_t) = \sigma^2 < \infty$, dan di asumsikan bahwa $\alpha_i \in [0, 1]$, untuk $i = 1, \dots, p$.

Estimasi parameter yang digunakan dalam model ZINAR(1) adalah menggunakan algoritma *Expectation-Maximization (E-M)*. Algoritma E-M adalah prosedur iteratif yang efisien untuk menghitung estimasi *Maximum Likelihood (ML)* dalam keberadaan data yang hilang atau tersembunyi. Tujuannya untuk memperkirakan parameter model yang membuat data yang diamati menjadi paling mungkin [8]. Setiap iterasi dari algoritma EM terdiri dari dua proses yaitu, langkah ekspektasi (E) dan langkah Memaksimumkan (M). Pada Langkah E, data yang hilang diperkirakan berdasarkan data yang diamati dan estimasi saat ini dari parameter model menggunakan ekspektasi bersyarat.

Pada langkah M, fungsi *likelihood* dimaksimalkan dengan asumsi bahwa data yang hilang diketahui. Estimasi data yang hilang dari langkah E digunakan sebagai pengganti data yang sebenarnya hilang. Konvergensi akan tercapai seiring meningkatnya nilai *likelihood* pada setiap iterasi. Langkah-langkah estimasi parameter dengan metode E-M sebagai berikut [4]:

1. E-step: Misalkan $\hat{\theta}^{(k)}$ adalah estimasi Langkah ke- k saat ini dari θ .

Dengan menggunakan properti ekspektasi kondisional, dapat dihitung fungsi $Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(k)})$ yang diberikan oleh: $Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(k)}) = E[\ell_c(\boldsymbol{\theta}|\mathbf{y}_c)|\mathbf{y}, \hat{\boldsymbol{\theta}}^{(k)}]$

2. M-step: maksimumkan $Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(k)})$ terhadap $\boldsymbol{\theta}$, sehingga diperoleh $\hat{\boldsymbol{\theta}}^{(k+1)}$.

Fungsi Q ditentukan berdasarkan ekspektasi-ekspektasi tertentu yang harus diketahui atau dihitung terlebih dahulu. Ekspektasi-ekspektasi ini memberikan informasi yang diperlukan untuk membentuk fungsi Q , yang digunakan dalam algoritma E-M untuk memperbarui estimasi parameter model, sebagai berikut: $\hat{s}_t^{(k)} = E[S_t|\mathbf{y}, \hat{\boldsymbol{\theta}}^{(k)}]$, $\hat{w}_t^{(k)} = E[W_t|\mathbf{y}, \hat{\boldsymbol{\theta}}^{(k)}]$ dan $Q_t^*(\boldsymbol{\lambda}|\boldsymbol{\theta}^{(k)}) = E[(1 - W_t) \log(h(y_t - S_t|\boldsymbol{\lambda})|\mathbf{y}, \hat{\boldsymbol{\theta}}^{(k)})]$. Dengan demikian, diperoleh formula estimasi parameter sebagai berikut: $\hat{\alpha}^{(k+1)} = \frac{\sum_{t=2}^n \hat{s}_t}{\sum_{t=2}^n y_{t-1}}$, $\hat{\rho}^{(k+1)} = \frac{\sum_{t=2}^n \hat{w}_t}{n-1}$, dan $\hat{\boldsymbol{\lambda}}^{(k+1)} = \arg \max_{\boldsymbol{\lambda}} \{\sum_{t=2}^n Q_t^*(\boldsymbol{\lambda}|\boldsymbol{\theta}^{(k)})\}$.

3 METODE PENELITIAN

3.1 Sumber Data dan Variabel Penelitian

Data empiris yang digunakan dalam penelitian ini berupa frekuensi kejadian gempa bumi besar dengan kekuatan magnitudo momen (M_w) ≥ 7 selama periode 1960-2024. Data tersebut diperoleh dari basis data resmi *United States Geological Survey* (USGS) [22], yang merupakan lembaga kredibel dan banyak dijadikan rujukan dalam penelitian kebumihan. Pemilihan rentang waktu tersebut dimaksudkan untuk memperoleh gambaran yang komprehensif mengenai pola kejadian gempa bumi besar dalam jangka panjang.

Variabel penelitian yang digunakan bersumber dari data empiris tersebut, yaitu frekuensi kejadian gempa bumi dengan magnitudo $M_w \geq 7$. Variabel ini dipilih karena gempa bumi dengan magnitudo besar memiliki dampak signifikan terhadap lingkungan dan aktivitas manusia, sehingga relevan untuk dianalisis dalam konteks kajian kebencanaan dan dinamika seismotektonik.

3.2 Tahapan Analisis Penelitian

Untuk mencapai tujuan penelitian, analisis data dilakukan melalui serangkaian tahapan yang terstruktur dan saling berkaitan. Tahapan ini mencerminkan alur kerja penelitian mulai dari pengumpulan data hingga penentuan model terbaik berdasarkan kinerja statistik. Tahap awal diawali dengan pengumpulan dan penyiapan data empiris yang menjadi dasar analisis. Data yang telah dikumpulkan selanjutnya diuji karakteristik statistiknya untuk mengidentifikasi adanya overdispersi. Apabila data tidak menunjukkan overdispersi, maka digunakan model Poisson-INAR (Poi-INAR). Sebaliknya, jika terdeteksi overdispersi, analisis dilanjutkan dengan pengujian terhadap keberadaan inflasi nol (*zero-inflation*). Pada tahap pengujian *zero-inflation*, ditentukan apakah data memiliki kelebihan nilai nol. Jika inflasi nol tidak terdeteksi, maka digunakan model Negative Binomial-INAR (NB-INAR). Namun, apabila data menunjukkan karakteristik inflasi nol, maka diterapkan model *zero-inflated* yang sesuai. Dalam hal ini, model *Zero-Inflated* INAR (ZINAR) digunakan untuk data yang mengalami inflasi nol tanpa overdispersi, sedangkan model Negative Binomial *Zero-Inflated* INAR (NB-ZINAR) diterapkan pada data yang sekaligus menunjukkan overdispersi dan inflasi nol.

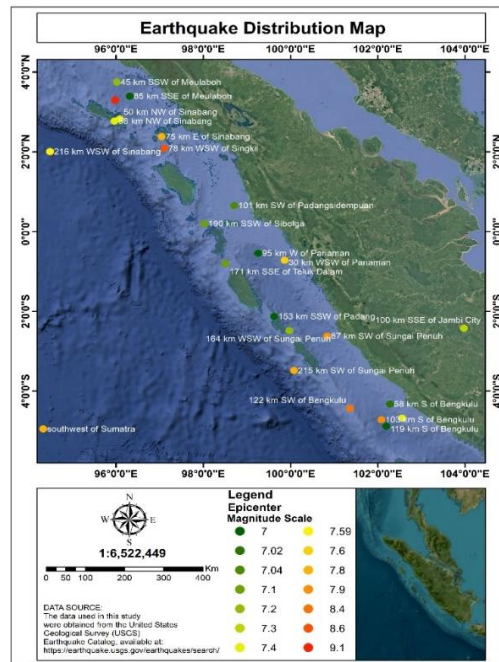
Seluruh model yang dihasilkan kemudian dievaluasi untuk menentukan model yang paling sesuai dengan karakteristik data empiris. Evaluasi kinerja model dilakukan menggunakan beberapa ukuran statistik, yaitu *Akaike Information Criterion* (AIC), *Bayesian Information Criterion* (BIC), *Mean Arc tangent Absolute Percentage Error* (MAAPE), dan *Root Mean Square Error* (RMSE). Model dengan nilai evaluasi paling optimal berdasarkan keempat kriteria tersebut selanjutnya ditetapkan sebagai model terbaik.

JURNAL MATEMATIKA, STATISTIKA DAN KOMPUTASI

Indah Wahyuliani, Jose Rizal, Sigit Nugroho

4. HASIL DAN PEMBAHASAN

Bab ini menyajikan hasil penerapan model ZINAR dalam memodelkan frekuensi kejadian gempa bumi besar ($M_w \geq 7$) di Pulau Sumatra. Analisis dilakukan berdasarkan formulasi model ZINAR sebagaimana dirumuskan pada Persamaan (2.5), (2.6), dan (2.7), dalam kajian literatur, yang mengintegrasikan komponen zero-inflated untuk mengakomodasi kelebihan nilai nol dan struktur model INAR untuk menangkap ketergantungan temporal pada data deret waktu diskrit. Estimasi parameter model dilakukan menggunakan pendekatan *maximum likelihood* dengan bantuan perangkat lunak bahasa program R.



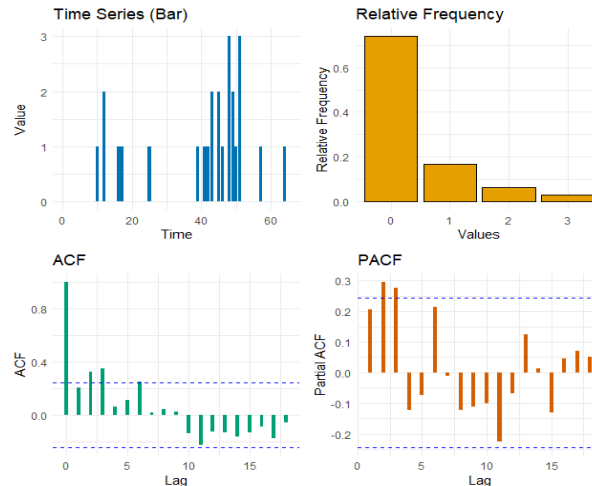
Gambar 4.1. Peta Sebaran Gempa dengan Kekuatan Magnitudo $M_w \geq 7$ di Pulau Sumatra.

Sebagai langkah awal dalam penyajian hasil, Gambar 4.1 menampilkan distribusi spasial episentrum gempa bumi bermagnitudo $M_w \geq 7$ yang terjadi di Pulau Sumatra selama periode pengamatan. Sebaran episentrum menunjukkan pola konsentrasi yang dominan di sepanjang pesisir barat Sumatra, sejalan dengan keberadaan zona subduksi megathrust Sumatra. Pola ini mengindikasikan bahwa interaksi tektonik antara Lempeng Indo-Australia dan Lempeng Eurasia merupakan faktor utama yang memengaruhi kejadian gempa bumi besar di wilayah tersebut.

Pada peta Gambar 4,1, titik-titik merah merepresentasikan lokasi episentrum gempa bumi dengan magnitudo $M_w \geq 7$, sedangkan label teks menunjukkan nama atau posisi relatif kejadian gempa terhadap kota terdekat. Penyajian informasi spasial secara langsung pada peta ini mempermudah interpretasi visual dan memungkinkan identifikasi cepat terhadap wilayah yang mengalami aktivitas seismik intensif tanpa memerlukan eksplorasi tambahan. Secara visual, tampak adanya pengelompokan (kluster) gempa yang cukup dominan di wilayah Sumatra Barat dan Bengkulu. Pola ini mengindikasikan bahwa kawasan tersebut merupakan zona dengan tingkat kerawanan gempa besar yang tinggi, sejalan dengan keberadaan zona subduksi aktif di sepanjang pesisir barat Sumatra. Kondisi ini menegaskan bahwa bagian barat Pulau Sumatra memiliki tingkat risiko seismik yang signifikan. Dengan demikian, hasil visualisasi ini menekankan pentingnya penguatan sistem pemantauan seismik, pengembangan sistem peringatan dini, serta peningkatan

JURNAL MATEMATIKA, STATISTIKA DAN KOMPUTASI
Indah Wahyuliani, Jose Rizal, Sigit Nugroho

kesiapsiagaan masyarakat, khususnya di wilayah pesisir barat Sumatra, sebagai bagian dari strategi mitigasi bencana gempa bumi dan tsunami.



Gambar 4.2. Visualisasi Frekuensi Gempa Berkekuatan Magnitudo $M_w \geq 7$ di Pulau Sumatra.

Gambar 4.2 menyajikan visualisasi deret waktu diskrit dan analisis statistik pendukung frekuensi kejadian gempa bumi bermagnitudo $M_w \geq 7$ di Pulau Sumatra. Secara visual, sebagian besar observasi bernilai 0 atau 1, dengan hanya sedikit periode yang mencatat nilai hingga 3 kejadian, yang menegaskan bahwa gempa bumi besar merupakan fenomena yang relatif jarang. Dominasi nilai nol, yang mencakup lebih dari 65% total observasi, mengindikasikan adanya inflasi nol (*zero-inflation*) dalam data. Selain itu, analisis fungsi autokorelasi menunjukkan nilai autokorelasi yang tinggi pada lag awal, khususnya lag-1, dengan pola penurunan bertahap pada lag berikutnya, yang mencerminkan adanya ketergantungan temporal. Pola ini diperkuat oleh fungsi autokorelasi parsial yang memperlihatkan signifikansi pada beberapa lag sebelum mendekati nol.

Secara deskriptif, data memiliki nilai minimum 0 dan maksimum 3 kejadian per periode, dengan nilai rata-rata sebesar 0,385 dan standar deviasi sebesar 0,743. Nilai standar deviasi yang lebih besar daripada rata-rata mengindikasikan adanya variabilitas yang tinggi serta kecenderungan overdispersi. Secara keseluruhan, karakteristik ini menunjukkan bahwa data frekuensi gempa besar di Sumatra bersifat langka, terinflasi nol, dan memiliki struktur autoregresif yang jelas, sehingga memerlukan pendekatan pemodelan deret waktu diskrit yang mampu mengakomodasi ketiga karakteristik tersebut secara simultan.

Kombinasi antara rata-rata yang rendah, nilai minimum yang dominan, dan standar deviasi yang cukup besar dibandingkan dengan mean mengindikasikan potensi overdispersi dan inflasi nol, dua karakteristik umum dalam data *count* yang menyulitkan model tradisional seperti Poisson dengan satu parameter. Oleh karena itu, pemilihan model deret waktu diskrit seperti ZINAR atau NB-INAR yang mampu menangani sifat tersebut menjadi penting dalam pemodelan data ini. Selanjutnya dilakukan diagnose model awal menggunakan deteksi overdispersi dan *zero-inflated* dalam data sehingga diperoleh hasil pada Tabel 4.1 sebagai berikut:

Tabel 4.1. Diagnosa Awal Data Deret Waktu Diskrit Data Gempa berkekuatan Magnitudo $M_w \geq 7$

| Sumber | <i>p-value</i> | Statistik Uji χ^2 | Parameter Dispersi | Keterangan | Model |
|----------------------|----------------|------------------------|--------------------|---------------|----------|
| Overdispersi | 0,035 | | 1,439 | Overdispersi | NB-ZINAR |
| <i>Zero-inflated</i> | 0,014 | 5,995 | | Zero Inflated | |

Berdasarkan hasil diagnosis awal karakteristik data deret waktu diskrit frekuensi kejadian gempa bumi bermagnitudo $M_w \geq 7$ yang disajikan pada Tabel 4.1, dilakukan pengujian untuk menentukan pendekatan pemodelan yang paling sesuai. Pengujian difokuskan pada dua karakteristik utama data count kejadian langka, yaitu overdispersi dan inflasi nol (*zero-inflation*). Hasil uji overdispersi menunjukkan nilai *p-value* sebesar 0,035 dan nilai statistik parameter dispersi sebesar 1,439. Karena *p-value* $< 0,05$, maka dapat disimpulkan bahwa data mengalami overdispersi secara signifikan, yakni varians data lebih besar dari nilai rata-ratanya. Overdispersi ini menandakan bahwa model Poisson tidak akan memberikan hasil estimasi yang optimal, sehingga perlu digunakan model alternatif yang dapat mengakomodasi varian yang lebih besar, seperti model Negative Binomial atau model deret waktu diskrit berbasis ZINAR.

Tabel 4.2. Estimasi Parameter Data Frekuensi Gempa Bumi Berkekuatan Magnitudo $M_w \geq 7$

| Metode | Ukuran Kebaikan Model | | | Parameter | | | |
|----------|-----------------------|---------|----------------|-----------|-----------|----------|--------|
| | AIC | BIC | Log likelihood | α | λ | θ | ρ |
| INAR | 106,309 | 110,309 | -51,154 | 0,193 | 0,318* | | |
| NB-INAR | 103,807 | 110,140 | -48,903 | 0,202 | 0,314* | 0,479 | |
| NB-ZINAR | 105,032 | 113,475 | -48,515 | 0,202 | 0,852* | 42,178* | 0,631* |
| ZINAR | 103,032 | 109,364 | -48,515 | 0,202 | 0,869* | | 0,639* |

Catatan: α adalah parameter autoregresif; λ adalah parameter inovasi; θ adalah parameter overdispersi; dan ρ adalah parameter inflasi-nol. AIC dan BIC merupakan ukuran *goodness-of-fit* (kecocokan model), di mana nilai yang lebih rendah menunjukkan kecocokan model yang lebih baik. Tanda bintang (*) menunjukkan bahwa parameter tersebut signifikan secara statistik pada tingkat signifikansi 5%.

Selanjutnya, hasil pengujian terhadap fenomena *zero-inflation* menunjukkan bahwa nilai *p-value* sebesar 0,014 dan nilai statistik uji sebesar 5,995. Dengan nilai *p-value* yang juga lebih kecil dari 0,05, maka dapat disimpulkan bahwa terdapat kelebihan nilai nol dalam data (*zero-inflated*) yang signifikan. Hal ini sesuai dengan karakteristik data yang telah teridentifikasi sebelumnya melalui visualisasi, di mana sebagian besar pengamatan bernilai nol. Berdasarkan dua hasil pengujian tersebut, dapat disimpulkan bahwa model deret waktu yang tepat untuk data kegempaan yang dimodelkan adalah model ZINAR.

Estimasi parameter frekuensi kejadian gempa bumi bermagnitudo $M_w \geq 7$ dilakukan menggunakan algoritma Expectation-Maximization (E-M) untuk memastikan kestabilan dan konsistensi estimasi pada data deret waktu diskrit. Tabel 4.2 menyajikan hasil estimasi parameter beserta ukuran kebaikan model dari empat pendekatan pemodelan yang dipertimbangkan, yaitu INAR, NB-INAR, NB-ZINAR, dan ZINAR, yang masing-masing dirancang untuk menangani karakteristik khusus data frekuensi kejadian gempa bumi besar. Evaluasi dilakukan berdasarkan nilai AIC, BIC, *log-likelihood*, dan signifikansi parameter utama dari masing-masing model.

Dari sisi pengujian kesesuaian model, model ZINAR menunjukkan performa terbaik dengan nilai AIC terendah (103,032) dan *log-likelihood* tertinggi (-48,515) dibandingkan model lainnya. Hal ini menunjukkan bahwa model ZINAR paling sesuai dalam merepresentasikan karakteristik data, khususnya dalam menangani fenomena overdispersi dan *zero-inflation*. Nilai parameter ρ sebesar 0,639 menunjukkan proporsi signifikan dari inflasi nol dalam data, sementara nilai λ sebesar 0,869 menandakan rata-rata kejadian yang lebih besar dibandingkan model lainnya.

Model NB-INAR dan INAR secara berturut-turut memiliki performa lebih rendah dibanding ZINAR, dengan nilai AIC masing-masing sebesar 103,807 dan 106,309. Model NB-INAR mengakomodasi overdispersi dengan menambahkan parameter $\theta=0,479$, namun tidak mempertimbangkan inflasi nol secara eksplisit, sehingga masih kalah dari ZINAR dalam konteks data ini. Sementara itu, model NB-ZINAR menghasilkan nilai AIC dan BIC yang jauh lebih tinggi (105,032 dan 113,475), serta *log-likelihood* yang lebih rendah (-48,515), meskipun secara teori model ini menggabungkan kemampuan menangani overdispersi dan *zero-inflation* sekaligus. Hasil ini mengindikasikan bahwa meskipun kompleksitas model lebih tinggi, NB-ZINAR tidak memberikan peningkatan performa yang signifikan dalam konteks dataset ini, bahkan cenderung *overfitting* atau tidak efisien.

Dari estimasi parameter model yang ditampilkan pada Tabel 4.2, nilai α sebagai parameter autoregresif relatif stabil di angka 0,19 sampai 0,20 pada semua model. Hasil ini menandakan adanya pengaruh moderat dari nilai masa lalu terhadap kejadian saat ini. Sementara nilai λ lebih tinggi pada model ZINAR dan NB-ZINAR, yang menunjukkan estimasi rata-rata kejadian yang lebih besar dalam model-model yang mempertimbangkan *zero-inflation*. Secara keseluruhan, hasil ini menegaskan bahwa model ZINAR merupakan pendekatan terbaik dalam memodelkan data deret waktu diskrit kejadian gempa bumi besar di Pulau Sumatra, karena mampu menangkap sifat data yang mengandung banyak nilai nol dan ketergantungan waktu, dengan hasil estimasi yang paling efisien dan akurat.

Berdasarkan hasil seleksi model, model ZINAR terpilih sebagai pendekatan yang paling sesuai untuk memodelkan frekuensi kejadian gempa bumi bermagnitudo $M_w \geq 7$. Selanjutnya, model ini digunakan untuk melakukan prediksi terhadap data kejadian gempa besar, dengan hasil prediksi disajikan pada Tabel 4.3. Dari tiga observasi yang dianalisis, model ZINAR mampu menghasilkan prediksi yang akurat pada dua periode awal, yang keduanya memiliki nilai aktual nol. Namun, pada satu periode dengan nilai aktual satu kejadian, model memprediksi nilai nol. Hasil ini menunjukkan bahwa model ZINAR memiliki kecenderungan prediksi yang konservatif, yaitu lebih sering mengestimasi tidak terjadinya gempa besar. Pola tersebut sejalan dengan karakteristik data empiris yang didominasi oleh nilai nol serta mencerminkan kemampuan model dalam menangkap struktur probabilistik kejadian langka. Meskipun demikian, hasil ini juga mengindikasikan adanya potensi underestimation pada periode dengan kejadian positif, yang menjadi tantangan umum dalam pemodelan dan prediksi fenomena ekstrem dan jarang terjadi.

Tabel 4.3. Hasil Prediksi Data Frekuensi Gempa Bumi Berkekuatan Magnitudo $M_w \geq 7$

| Prediksi Data Gempa Bumi | |
|--------------------------|---------|
| Testing | Predict |
| 0 | 0 |
| 0 | 0 |
| 1 | 0 |
| MAAPE | 1.327 |
| RMSE | 0.484 |

Kinerja prediksi model dievaluasi secara kuantitatif menggunakan *Mean Arctangent Absolute Percentage Error* (MAAPE) dan *Root Mean Square Error* (RMSE). Nilai MAAPE sebesar 1,327 menunjukkan bahwa secara relatif tingkat kesalahan prediksi masih cukup tinggi, yang terutama disebabkan oleh ketidaksesuaian prediksi pada periode dengan kejadian gempa bernilai non-nol.

Sebaliknya, nilai RMSE sebesar 0,484 mengindikasikan bahwa secara absolut deviasi antara nilai prediksi dan nilai aktual relatif kecil, yang sebagian besar dipengaruhi oleh dominasi nilai nol dalam data. Hasil evaluasi ini menunjukkan bahwa model memiliki kinerja yang baik dalam mengidentifikasi periode tanpa kejadian gempa besar, namun masih menunjukkan keterbatasan dalam memprediksi kejadian gempa yang benar-benar terjadi. Keterbatasan tersebut merupakan karakteristik umum dalam penerapan model deret waktu diskrit pada data dengan tingkat inflasi nol yang tinggi, di mana frekuensi kejadian non-nol yang sangat rendah menyebabkan model cenderung lebih menyesuaikan diri pada pola dominan berupa nilai nol. Hal ini menegaskan tantangan metodologis dalam pemodelan kejadian langka dan perlunya pengembangan pendekatan yang lebih adaptif terhadap kejadian ekstrem.

4 KESIMPULAN DAN SARAN

Pada bagian ini menyajikan rangkuman hasil utama penelitian serta implikasi metodologis dan praktis yang diperoleh dari penerapan model ZINAR dalam memodelkan frekuensi kejadian gempa bumi besar di Pulau Sumatra. Selain itu, juga dikemukakan beberapa rekomendasi pengembangan penelitian lanjutan yang relevan untuk meningkatkan kemampuan pemodelan dan interpretasi dinamika seismik di masa mendatang.

4.1 Kesimpulan

Penelitian ini telah menerapkan model ZINAR untuk memodelkan frekuensi kejadian gempa bumi besar ($M_w \geq 7$) di Pulau Sumatra periode 1960-2024. Hasil analisis menunjukkan bahwa data memiliki karakteristik overdispersi dan inflasi nol yang signifikan, sehingga model konvensional seperti Poisson-INAR kurang memadai. Dibandingkan model alternatif (INAR, NB-INAR, dan NB-ZINAR), model ZINAR memberikan kinerja terbaik berdasarkan nilai AIC terendah, log-likelihood tertinggi, serta estimasi parameter yang stabil dan signifikan. Secara substantif, model ZINAR mampu merepresentasikan ketergantungan temporal sekaligus fenomena *excess zeros* dalam deret waktu seismik, sehingga memberikan pemodelan yang lebih realistis terhadap kejadian gempa yang bersifat langka. Dari sisi metodologis, hasil ini menegaskan keunggulan pendekatan *zero-inflated count time series* dalam menangani data diskrit dengan distribusi tidak seimbang. Penerapan model ini berimplikasi pada peningkatan akurasi pemodelan risiko seismik dan mendukung perencanaan mitigasi bencana berbasis probabilistik.

4.2 Saran

Penelitian selanjutnya disarankan untuk mengembangkan model ZINAR dengan memasukkan variabel eksogen yang mencerminkan faktor geofisik, seperti kedalaman gempa, aktivitas sesar, dan dinamika subduksi. Selain itu, eksplorasi pendekatan multivariat, model Bayesian hierarkis, serta analisis spasio-temporal berpotensi meningkatkan kemampuan interpretasi dan prediksi model terhadap kompleksitas kejadian gempa di wilayah Indonesia.

UCAPAN TERIMA KASIH

Penulis menyampaikan terima kasih kepada Kementerian Pendidikan, Ilmu Pengetahuan, dan Teknologi Republik Indonesia atas dukungan pendanaan melalui Direktorat Riset dan Pengabdian kepada Masyarakat dalam skema Penelitian Tesis Magister, berdasarkan Kontrak No. 062/C3/DT.05.00/PL/2025. Disamping itu, hibah penelitian ini dikelola oleh Lembaga Penelitian dan Pengabdian kepada Masyarakat (LPPM) Universitas Bengkulu berdasarkan Kontrak No. 2886/UN30.15/PT/2025.

DAFTAR PUSTAKA

- [1] Al-Osh, M. A., & Alzaid, A. A., 1987. First-order integer-valued autoregressive (INAR (1)) process. *Journal of Time Series Analysis*, Vol. 8, No. 3, 261-275.
- [2] Borman, S., 2004. *The expectation maximization algorithm: A short tutorial*. Technical Report, University of California.
- [3] Box, G.E.P., & Jenkins, G. M., 1976. *Time Series Analysis: Forecasting and Control*. San Francisco: Holden-Day.
- [4] Cameron, A. C., & Trivedi, P. K., 1990. Regression-based tests for overdispersion in the Poisson model. *Journal of Econometrics*, Vol. 46, No. 3, 347-364.
- [5] Charlton, S., 2014. Zero-Inflated Count Time Series. In *Wiley StatsRef: Statistics Reference Online* (pp.1-5).
- [6] Fávero, L. P., Souza, R. D. F., Belfiore, P., Corrêa, H. L., & Haddad, M. F., 2021. Count data regression analysis: Concepts, overdispersion detection, zero-inflation identification, and applications with R. *Practical Assessment, Research & Evaluation*, Vol. 26, No. 13.
- [7] Garay, A. M., Medina, F. L., Cabral, C. R., & Lin, T. I., 2020. Bayesian analysis of the p-order integer-valued AR process with zero-inflated Poisson innovations. *Journal of Statistical Computation and Simulation*, Vol. 90, No. 11, 1943-1964.
- [8] Hadi, A. I., Farid, M., Mase, L. Z., Refrizon, Aditya, K. B., Fadli, D. I., & Sumanjaya, E., 2025. Seismic microzonation of kepahiang regency, Indonesia using the HVSR method: enhancing Urban resilience to mitigate earthquake vulnerability. *Indian Geotechnical Journal*, 1-15.
- [9] Hning, D. B., 1994. A note on a test for Poisson overdispersion. *Biometrika*, Vol. 81, No. 2, 418-419.
- [10] Hu, H., 2008. Poisson distribution and application. *A Course in Department of Physics and Astronomy*; Technical Report University of Tennessee at Knoxville: Knoxville, TN, USA
- [11] Jazi, M. A., Jones, G., & Lai, C. D., 2012. First-order integer valued AR processes with zero inflated Poisson innovations. *Journal of Time Series Analysis*, Vol. 33, No. 6, 954-963.
- [12] Jin-Guan, D., & Yuan, L., 1991. The integer-valued autoregressive (INAR (p)) model. *Journal of time series analysis*, Vol. 12, No. 2, 129-142.
- [13] Lambert, D., 1992. Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics*, Vol. 34, No. 1, 1-14.
- [14] Popović, P. M., Bakouch, H. S., & Ristić, M. M., 2025. A non-linear integer-valued autoregressive model with zero-inflated data series. *Journal of Applied Statistics*, Vol. 52, No. 6, 1195-1218.
- [15] Rizal, J., Nugroho, S., & Irwanto, A., 2016. Analisis Kejadian Gempa Bumi Tektonik di Wilayah Pulau Sumatera. *Jurnal Matematika*, Vol. 6, No. 1, 1-14.
- [16] Sathish, V., Mukhopadhyay, S., & Tiwari, R., 2022. Autoregressive and moving average models for zero-inflated count time series. *Statistica Neerlandica*, Vol. 76, No. 2, 190-218.

- [17] Kang, Y., Wang, S., Wang, D., & Zhu, F., 2023. Analysis of zero-and-one inflated bounded count time series with applications to climate and crime data. *Test*, Vol. 32, No. 1, 34-73.
- [18] Steutel, F. W., & van Harn, K., 1979. Discrete analogues of self-decomposability and stability. *The Annals of Probability*, Vol. 7, No. 5, 893-899.
- [19] Weiss, C. H., Homburg, A., & Puig, P., 2019. Testing for zero inflation and overdispersion in INAR (1) models. *Statistical Papers*, Vol. 60, No. 3, 823-848.
- [20] Yu, K. H., & Kim, S. G., 2025. The first-order seasonal integer-valued autoregression process with zero-inflated Poisson innovations; application to integer-valued seasonal data analysis with overdispersion. *Brazilian Journal of Probability and Statistics*, Vol. 39, No. 1, 39-58.
- [21] Zeileis, A., Kleiber, C., & Jackman, S., 2008. Regression models for count data in R. *Journal of Statistical Software*, Vol. 27, No. 8, 1-25.