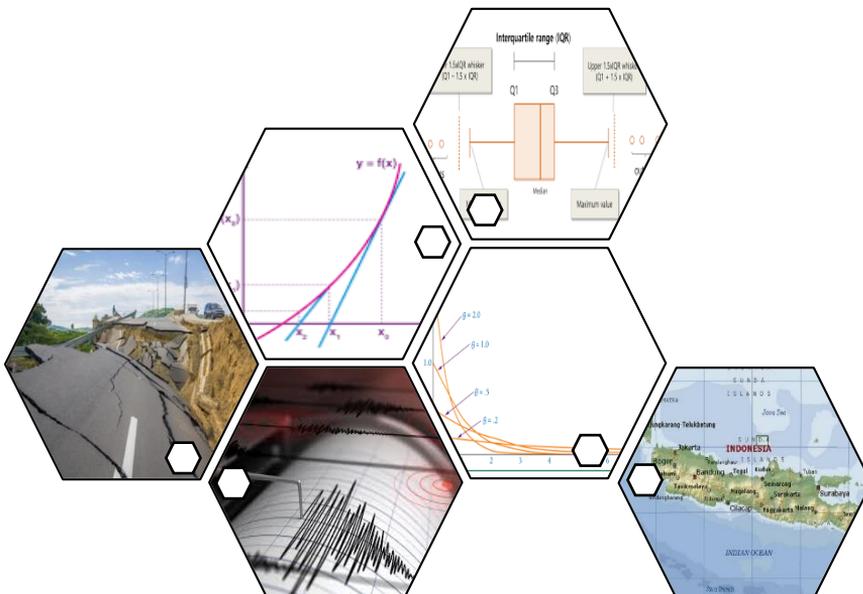


**PEMODELAN DATA WAKTU ANTAR KEJADIAN GEMPA BUMI YANG
MENGANDUNG *OUTLIER* MENGGUNAKAN DISTRIBUSI
EKSPONENSIAL TERKONTAMINASI**

**MODELING EARTHQUAKE INTER EVENT TIME DATA CONTAINING
OUTLIERS USING CONTAMINATED EXPONENTIAL DISTRIBUTION**



NURHIDAYA L

H062222007



**PROGRAM STUDI MAGISTER STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS HASANUDDIN
MAKASSAR
2024**

**PEMODELAN DATA WAKTU ANTAR KEJADIAN GEMPA BUMI YANG
MENGANDUNG *OUTLIER* MENGGUNAKAN DISTRIBUSI
EKSPONENSIAL TERKONTAMINASI**

NURHIDAYA L

H062222007



**PROGRAM STUDI MAGISTER STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS HASANUDDIN
MAKASSAR
2024**

**PEMODELAN DATA WAKTU ANTAR KEJADIAN GEMPA BUMI YANG
MENGANDUNG *OUTLIER* MENGGUNAKAN DISTRIBUSI
EKSPONENSIAL TERKONTAMINASI**

Tesis
sebagai salah satu syarat untuk mencapai gelar magister

Program Studi Magister Statistika

Disusun dan diajukan oleh

NURHIDAYA L
H062222007

Kepada

**PROGRAM STUDI MAGISTER STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS HASANUDDIN
MAKASSAR
2024**

TESIS

PEMODELAN DATA WAKTU ANTAR KEJADIAN GEMPA BUMI YANG MENGANDUNG *OUTLIER* MENGGUNAKAN DISTRIBUSI EKSPONENSIAL TERKONTAMINASI

NURHIDAYA L
H062222007

telah dipertahankan di depan Panitia Ujian Magister pada 7 Agustus 2024
dan dinyatakan telah memenuhi syarat kelulusan

pada

Program Studi Magister Statistika
Departemen Statistika
Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Hasanuddin
Makassar

Mengesahkan:

Pembimbing Utama



Dr. Erna Tri Herdiani, S.Si., M.Si.
NIP. 19750429 200003 2 001

Pembimbing Pendamping



Prof. Dr. Nurtiti Sunusi, S.Si., M.Si.
NIP. 19720117 199703 2 002

Ketua Program Studi
Magister Statistika



Dr. Erna Tri Herdiani, S.Si., M.Si.
NIP. 19750429 200003 2 001

Dekan Fakultas Matematika dan Magister
Ilmu Pengetahuan Alam
Universitas Hasanuddin



Dr. Eng. Amiruddin, M.Si.
NIP. 19720515 199702 1 002

PERNYATAAN KEASLIAN TESIS DAN PELIMPAHAN HAK CIPTA

Dengan ini saya menyatakan bahwa, tesis berjudul "Pemodelan Data Waktu Antar Kejadian Gempa Bumi Yang Mengandung *Outlier* Menggunakan Distribusi Eksponensial Terkontaminasi" adalah benar karya saya dengan arahan dari tim pembimbing (Dr. Erna Tri Herdiani, S.Si., M.Si. dan Prof. Dr. Nurtiti Sunusi, S.Si., M.Si.). Karya ilmiah ini belum diajukan dan tidak sedang diajukan dalam bentuk apapun kepada perguruan tinggi manapun. Sumber informasi yang berasal atau dikutip dari karya yang diterbitkan maupun tidak diterbitkan dari penulis lain telah disebutkan dalam teks dan dicantumkan dalam Daftar Pustaka tesis ini. Sebagian dari tesis ini akan dipublikasikan di *Iranian Journal of Geophysics* sebagai artikel dengan judul "*The contaminated exponential distribution to model earthquake inter event time data containing outliers*".

Dengan ini saya melimpahkan hak cipta dari karya tulis saya berupa tesis ini kepada Universitas Hasanuddin.

Makassar, 7 Agustus 2024



Nurhidaya L
NURHIDAYA L
NIM. H062222007

UCAPAN TERIMA KASIH

Segala puji hanya milik Allah *Subhanallahu Wa Ta'ala* atas limpahan rahmat dan hidayah-Nya kepada penulis. Shalawat dan salam tercurahkan kepada Rasulullah *Shallallahu 'Alaihi Wa sallam*, keluarganya, *tabi'in, tabi'ut tabi'in*, serta orang-orang sholeh yang haq hingga kadar Allah berlaku atas diri mereka. *Alhamdulillahirobbil'aalamiin*, berkat rahmat dan kemudahan dari Allah *Subhanallahu Wa Ta'ala*, penulis dapat menyelesaikan tesis berjudul "Pemodelan Data Waktu Antar Kejadian Gempa Bumi yang Mengandung *Outlier* Menggunakan Distribusi Eksponensial Terkontaminasi" sebagai salah satu syarat memperoleh gelar magister pada Program Studi Magister Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin.

Penulis menyadari bahwa penyelesaian tesis ini tidak lepas dari bantuan dan dorongan dari berbagai pihak. Oleh karena itu, penulis menyampaikan terima kasih yang sebesar-besarnya kepada Ibu Dr. Erna Tri Herdiani, S.Si., M.Si., selaku pembimbing utama dan Ibu Prof. Dr. Nurtiti Sunusi, S.Si., M.Si., selaku pembimbing pertama yang dengan sabar telah menyisihkan waktu dan pikirannya di tengah berbagai kesibukan dan prioritasnya, selalu memberikan panduan, dorongan, motivasi, dan kemudahan kepada penulis dalam menyelesaikan tesis ini. Ucapan terima kasih juga penulis sampaikan kepada Ibu Prof. Dr. Dr. Georgina Maria Tinungki, M.Si., Ibu Dr. Anna Islamiyati, S.Si., M.Si., dan Bapak Dr. Nirwan, M.Si., selaku tim penguji yang telah memberikan kritikan membangun, arahan dalam penyempurnaan penyusunan tesis ini, serta waktu yang telah diluangkan untuk penulis.

Penulis haturkan rasa terima kasih yang setulus-tulusnya serta penghargaan yang setinggi-tingginya untuk kedua orang tua penulis, Bapak Abd Latif dan Ibu Hasriati yang telah memberikan dukungan penuh. Kakak dan adik penulis, terima kasih telah memberikan dorongan dan dukungan untuk penulis serta untuk keluarga besar penulis, terima kasih atas doa dan dukungannya selama ini. Spesial untuk sahabat penulis yaitu Andi Sri Yulianti, terima kasih atas waktu yang telah diberikan kepada penulis sejak tahun 2018 hingga sekarang. Terima kasih juga atas bantuan, arahan, nasihat, dukungan, dorongan, dan motivasi yang telah diberikan selama proses perkuliahan dan penyusunan tesis ini. Terima kasih atas kebersamaan, serta segala suka dan dukanya. Teman seperjuangan di Program Studi Magister Statistika Angkatan 7, Nurul, Nuge, Rifka, Kak Haura, Kak Miftah, Kak Ani, Hikmah, Lili, Kak Ikka, Kak Nola, Mega, Isra, Kak Cici, Fira dan lainnya terima kasih atas bantuan, dukungan, dan kebersamaannya selama ini. Serta kepada pihak yang tidak dapat penulis tuliskan satu-persatu, terimakasih atas bantuannya.

Penulis menyadari bahwa masih banyak kekurangan dalam tesis ini, untuk itu dengan segala kerendahan hati, penulis memohon maaf.

Makassar, 7 Agustus 2024



Nurhidaya L

ABSTRAK

NURHIDAYA L. **Pemodelan Data Waktu Antar Kejadian Gempa Bumi yang Mengandung *Outlier* Menggunakan Distribusi Eksponensial Terkontaminasi** (dibimbing oleh Dr. Erna Tri Herdiani, S.Si., M.Si. dan Prof. Dr. Nurtiti Sunusi, S.Si., M.Si.)

Latar Belakang. Gempa bumi merupakan peristiwa yang sulit diprediksi. Data waktu antar kejadian gempa bumi umumnya bernilai positif. Salah satu model yang sering digunakan untuk menganalisis data bernilai positif adalah distribusi eksponensial. Namun, distribusi ini memiliki keterbatasan ketika memodelkan data waktu antar kejadian gempa bumi yang mengandung *outlier*. Dalam kondisi sebagian besar data mengikuti distribusi eksponensial, namun terdapat *outlier* akibat efek kontaminasi, penggunaan distribusi eksponensial terkontaminasi dapat menjadi solusi yang lebih tepat. **Tujuan.** Penelitian ini bertujuan untuk melakukan pemodelan eksponensial terkontaminasi pada data waktu antar kejadian gempa bumi periode 1971-2022 di Pulau Jawa dan sekitarnya. **Metode.** Adapun metode yang digunakan pada penelitian ini adalah maksimum *likelihood* dengan iterasi Newton Raphson. **Hasil.** Hasil penelitian menunjukkan bahwa pada data waktu antar kejadian gempa bumi periode 1971-2022 di Pulau Jawa dan sekitarnya, model eksponensial terkontaminasi memiliki nilai AIC yang lebih kecil dibandingkan dengan model eksponensial. **Kesimpulan.** Dari hasil penelitian, penggunaan distribusi eksponensial terkontaminasi lebih baik dalam memodelkan data waktu antar kejadian gempa bumi yang mengandung *outlier* dibandingkan distribusi eksponensial.

Kata kunci: distribusi eksponensial; distribusi eksponensial terkontaminasi; *maximum likelihood*; Newton Raphson; *outlier*

ABSTRACT

NURHIDAYA L. **Modeling Earthquake Inter Event Time Data Containing Outliers Using Contaminated Exponential Distribution** (supervised by Dr. Erna Tri Herdiani, S.Si., M.Si. and Prof. Dr. Nurtiti Sunusi, S.Si., M.Si.).

Background. Earthquakes are events that are difficult to predict. The inter event time data for earthquakes generally have positive values. One commonly used model for analyzing positive valued data is the exponential distribution. However, this distribution has limitations when modeling inter event time data for earthquakes that contain outliers. In cases where most of the data follow an exponential distribution, but there are outliers due to contamination effects, using a contaminated exponential distribution can be a more appropriate solution. **Objective.** This study aims to model contaminated exponential distribution on inter event time data for earthquakes from 1971-2022 in Java and its surrounding areas. **Method.** The research results indicate that, for the inter event time data for earthquakes from 1971-2022 in Java and its surrounding areas, the contaminated exponential model has a smaller AIC value compared to the exponential model. **Results.** The results show that the contaminated exponential model has a lower AIC value compared to the exponential model. **Conclusion.** From the results of the study, the use of a contaminated exponential distribution is better for modeling inter event time data for earthquakes that contain outliers compared to the exponential distribution.

Keywords: exponential distribution; contaminated exponential distribution; maximum likelihood; Newton Raphson; outlier

DAFTAR ISI

halaman

HALAMAN JUDUL	i
PERNYATAAN PENGAJUAN.....	ii
HALAMAN PENGESAHAN.....	iii
PERNYATAAN KEASLIAN TESIS.....	iv
UCAPAN TERIMA KASIH.....	v
ABSTRAK	vi
ABSTRACT	vii
DAFTAR ISI	viii
DAFTAR TABEL	xi
DAFTAR GAMBAR	xii
DAFTAR LAMPIRAN	xiii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	2
1.3 Batasan Masalah	3
1.4 Tujuan	3
1.5 Manfaat.....	3
1.6 Kajian Teori.....	3
1.6.1 Distribusi Ekspensial	3
1.6.2 Distribusi Ekspensial Terkontaminasi.....	4
1.6.3 Deteksi <i>Outlier</i>	5
1.6.4 Uji Kecocokan Distribusi Ekspensial	6
1.6.5 Metode <i>Maximum Likelihood</i>	6

1.6.6 Metode Newton Raphson.....	7
1.6.7 <i>Akaike Information Criterion</i>	8
1.6.8 Model Regresi Linier, Kuadratik, dan Kubik	8
1.6.9 Gempa Bumi	10
1.6.10 Kerangka Konseptual.....	12
BAB II METODE PENELITIAN.....	13
2.1 Sumber Data.....	13
2.2 Langkah Analisis.....	13
BAB III HASIL DAN PEMBAHASAN	16
3.1 Deskripsi Data Waktu Antar Kejadian Gempa Bumi Periode 1971-2022 di Pulau Jawa dan Sekitarnya	16
3.2 Pendeteksian <i>Outlier</i>	16
3.3 Uji Kecocokan Distribusi	18
3.4 Estimasi Parameter Model Eksponensial Terkontaminasi dan Model Eksponensial	19
3.4.1 Estimasi Parameter Model Eksponensial Terkontaminasi	19
3.4.2 Estimasi Parameter Model Eksponensial.....	37
3.5 Pemodelan Waktu Antar Kejadian Gempa Bumi Periode 1971-2022 di Pulau Jawa dan Sekitarnya	39
3.5.1 Pemodelan Waktu Antar Kejadian Gempa Bumi Periode 1971-2022 di Pulau Jawa dan Sekitarnya dengan Distribusi Eksponensial.....	39
3.5.2 Pemodelan Waktu Antar Kejadian Gempa Bumi Periode 1971-2022 di Pulau Jawa dan Sekitarnya dengan Distribusi Eksponensial Terkontaminasi.....	40
3.5.3 Pemilihan Model Terbaik.....	43
3.6 Perkiraan Peluang Waktu Antar Kejadian Gempa Bumi Kedepan.....	43
BAB IV PENUTUP	45
4.1 Kesimpulan	45
4.2 Saran	45
DAFTAR PUSTAKA.....	46

LAMPIRAN49

DAFTAR TABEL

nomor	halaman
1. Deskripsi data waktu antar kejadian gempa bumi periode 1971-2022 di pulau Jawa dan sekitarnya.....	16
2. Hasil uji kecocokan distribusi eksponensial.....	19
3. Peluang waktu antar kejadian gempa bumi periode 1971-2022 di pulau Jawa dan sekitarnya dengan menggunakan model eksponensial.....	40
4. Hasil estimasi parameter model waktu antar kejadian gempa bumi periode 1971-2022 di pulau Jawa dan sekitarnya	40
5. Peluang waktu antar kejadian gempa bumi periode 1971-2022 di pulau Jawa dan sekitarnya dengan menggunakan model eksponensial terkontaminasi....	41
6. Perbandingan nilai peluang antara model eksponensial dan model eksponensial terkontaminasi	42
7. Perbandingan nilai peluang untuk waktu antar kejadian yang diurutkan	42
8. Nilai AIC model eksponensial dan model eksponensial terkontaminasi	43
9. Peluang waktu antar kejadian gempa bumi periode 1971-2022 di pulau Jawa dan sekitarnya yang diurutkan dengan menggunakan model eksponensial terkontaminasi	43
10. Koefisien determinasi model regresi linier, kuadratik, dan kubik	44

DAFTAR GAMBAR

nomor	halaman
1. Kurva distribusi eksponensial.....	3
2. Kerangka konseptual.....	12
3. Peta pulau Jawa dan sekitarnya	13
4. <i>Boxplot</i> data waktu antar kejadian gempa bumi periode 1971-2022 di pulau Jawa dan sekitarnya.....	17

DAFTAR LAMPIRAN

nomor	halaman
1. Data kemunculan gempa bumi tektonik di pulau Jawa dan sekitarnya	49
2. Waktu antar kejadian kemunculan gempa bumi tektonik di pulau Jawa dan sekitarnya.....	50
3. <i>Output</i> uji kecocokan distribusi eksponensial dengan uji Kolmogorov Smirnov menggunakan <i>software</i> SPSS.	51
4. Peluang waktu antar kejadian gempa bumi periode 1971-2022 di pulau Jawa dan sekitarnya yang diurutkan dengan menggunakan model eksponensial dan model eksponensial terkontaminasi.	52
5. Daftar riwayat hidup	52

BAB I

PENDAHULUAN

1.1 Latar Belakang

Gempa bumi adalah kejadian yang sulit diprediksi. Gempa bumi adalah gerakan atau getaran pada permukaan bumi yang disebabkan oleh tenaga endogen. Tenaga endogen berasal dari dalam bumi yang dipicu oleh perubahan pada kulit bumi. Gempa bumi terdiri dari tiga jenis yaitu gempa bumi vulkanik, runtutan dan tektonik (Fadilah & Ronoatmojo, 2022). Data waktu antar kejadian gempa bumi merupakan data bernilai positif. Distribusi eksponensial adalah salah satu model yang paling diakui untuk data bernilai positif (Okhli & Nooghabi, 2021). Oleh karena itu, data waktu antar kejadian gempa bumi dapat dimodelkan dengan menggunakan distribusi eksponensial.

Penelitian tentang distribusi data waktu antar kejadian gempa bumi adalah penelitian oleh Chen dkk. pada tahun 2013. Penelitian tersebut membahas tentang distribusi data waktu antar kejadian gempa bumi di Taiwan. Hasil yang diperoleh pada penelitian tersebut adalah penggunaan distribusi Gamma disarankan untuk memodelkan data waktu antar kejadian gempa bumi di Taiwan. Penelitian lainnya yaitu penelitian oleh Sunusi dkk. pada tahun 2013. Pada penelitian tersebut data waktu antar kejadian gempa bumi di wilayah Nusa Tenggara berdistribusi eksponensial. Penelitian lain yang juga membahas tentang distribusi data waktu antar kejadian gempa bumi adalah penelitian oleh Pasari & Dikshit pada tahun 2015. Penelitian tersebut membahas tentang distribusi data waktu antar kejadian gempa bumi di Kachchh, India. Hasil yang diperoleh pada penelitian tersebut adalah data waktu antar kejadian gempa bumi di Kachchh, India mengikuti distribusi eksponensial.

Pada data waktu antar kejadian gempa bumi, seringkali terdapat *outlier*. *Outlier* adalah suatu pengamatan yang tidak mengikuti sebagian besar pola data (Herdiani dkk., 2021). Meskipun distribusi eksponensial memberikan banyak keuntungan dalam analisis data, distribusi eksponensial memiliki keterbatasan dalam memodelkan data yang mengandung *outlier*. Kehadiran *outlier* dalam data riil seringkali disebabkan oleh kontaminasi yang dapat berdampak pada hasil estimasi parameter model (Punzo & Maruotti, 2016).

Kontaminasi pada dasarnya mengacu pada titik-titik data yang secara teoritis tidak mengikuti distribusi dari titik-titik data lainnya (Okhli & Nooghabi, 2021). Kontaminasi berarti nilai-nilai yang menyimpang atau tidak sejalan dengan pola umum yang diikuti oleh sebagian besar data. Untuk mengatasi situasi sebagian besar data mengikuti distribusi eksponensial, namun ada *outlier* yang dihasilkan oleh efek kontaminasi, Okhli dan Nooghabi pada tahun 2021 mengembangkan model baru yang dikenal sebagai model eksponensial terkontaminasi.

Distribusi eksponensial terkontaminasi adalah model statistik yang menggabungkan distribusi eksponensial dengan elemen kontaminasi. Dalam distribusi eksponensial terkontaminasi, terdapat komponen tambahan pada distribusi

eksponensial yang mengikuti distribusi yang berbeda. Hal ini menciptakan situasi sebagian besar data mengikuti distribusi eksponensial, tetapi ada beberapa data yang dipengaruhi oleh komponen kontaminasi.

Penelitian tentang distribusi eksponensial terkontaminasi dilakukan oleh Okhli & Nooghabi pada tahun 2021. Penelitian tersebut membahas tentang distribusi eksponensial terkontaminasi pada data klaim asuransi. Hasil yang diperoleh pada penelitian tersebut adalah distribusi eksponensial terkontaminasi lebih baik daripada distribusi eksponensial. Penelitian lainnya yaitu penelitian oleh Mahdavi dkk. pada tahun 2022 yang membahas tentang distribusi *Contaminated Weighted Exponential* (CWE) pada data klaim asuransi. Distribusi CWE digunakan untuk memodelkan data yang mengalami *outlier* ringan. Pada penelitian tersebut distribusi eksponensial terkontaminasi digunakan sebagai pembandingan dari distribusi CWE. Hasil yang diperoleh adalah distribusi CWE lebih baik daripada distribusi eksponensial terkontaminasi. Hal ini menunjukkan bahwa untuk memodelkan data yang mengalami *outlier* ringan lebih baik menggunakan distribusi CWE.

Data yang digunakan pada penelitian ini adalah data waktu antar kejadian gempa bumi tektonik periode 1971-2022 di pulau Jawa dan sekitarnya. Hasil studi eksplorasi pada data waktu antar kejadian gempa bumi tektonik periode 1971-2022 di pulau Jawa dan sekitarnya menunjukkan bahwa data berdistribusi eksponensial. Hal ini dapat diketahui berdasarkan bentuk kurva dari data tersebut yang condong ke kanan dan terus menurun seiring meningkatnya data waktu antar kejadian gempa bumi tektonik periode 1971-2022 di pulau Jawa dan sekitarnya. Selain itu, data tersebut juga mengandung *outlier*. Beberapa model yang dapat digunakan untuk memodelkan data yang berdistribusi eksponensial tetapi terdapat *outlier* pada data tersebut adalah distribusi *generalized exponential* (Li & Hao, 2016), distribusi eksponensial terkontaminasi (Okhli & Nooghabi, 2021), dan distribusi CWE (Mahdavi dkk., 2022). Distribusi CWE digunakan untuk memodelkan data yang mengalami *outlier* ringan. Pada data waktu antar kejadian gempa bumi tektonik periode 1971-2022 di pulau Jawa dan sekitarnya terdapat data *outlier* yang ekstrim sehingga data tersebut lebih cocok dimodelkan dengan menggunakan distribusi eksponensial terkontaminasi. Oleh karena itu, penulis tertarik melakukan penelitian dengan judul "Pemodelan Data Waktu Antar Kejadian Gempa Bumi yang Mengandung *Outlier* Menggunakan Distribusi Eksponensial Terkontaminasi".

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Bagaimana bentuk estimasi parameter model eksponensial terkontaminasi pada data waktu antar kejadian gempa bumi tektonik periode 1971-2022 di pulau Jawa dan sekitarnya menggunakan metode *maximum likelihood*?
2. Bagaimana pemodelan data waktu antar kejadian gempa bumi tektonik periode 1971-2022 di pulau Jawa dan sekitarnya yang mengandung *outlier* menggunakan distribusi eksponensial terkontaminasi?

1.3 Batasan Masalah

Berdasarkan permasalahan yang telah diuraikan, batasan masalah dalam penelitian ini mencakup data kemunculan gempa bumi tektonik yang terjadi pada periode 1971-2022 di pulau Jawa dan sekitarnya dengan magnitudo ≥ 5 dan kedalaman 0-100 km.

1.4 Tujuan

Berdasarkan rumusan masalah yang telah diuraikan, maka tujuan penelitian ini adalah sebagai berikut:

1. Memperoleh bentuk estimasi parameter model eksponensial terkontaminasi pada data waktu antar kejadian gempa bumi tektonik periode 1971-2022 di pulau Jawa dan sekitarnya menggunakan metode *maximum likelihood*.
2. Memperoleh model eksponensial terkontaminasi pada data waktu antar kejadian gempa bumi tektonik periode 1971-2022 di pulau Jawa dan sekitarnya.

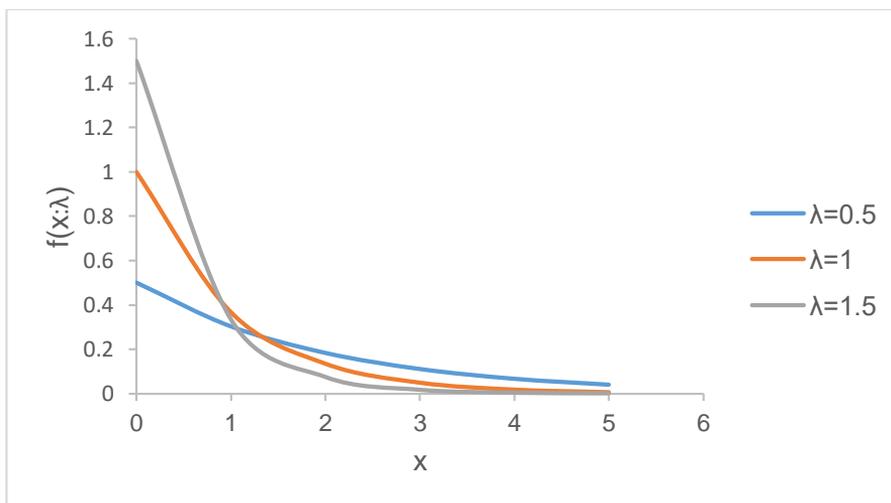
1.5 Manfaat

Menambah wawasan, informasi, dan pengetahuan tentang estimasi parameter dan pemodelan eksponensial terkontaminasi dengan menggunakan metode *maximum likelihood* dan sebagai bahan referensi bagi yang membutuhkan.

1.6 Kajian Teori

1.6.1 Distribusi Eksponensial

Distribusi eksponensial merupakan salah satu distribusi kontinu. Ciri utama dari distribusi ini adalah terdapat ekor di sebelah kanan kurvanya. Bentuk kurva distribusi eksponensial berubah-ubah tergantung pada nilai x dan λ yang digunakan. Adapun kurva dari distribusi eksponensial ditunjukkan pada Gambar 1 berikut:



Gambar 1. Kurva distribusi eksponensial

Fungsi kepadatan peluang variabel acak X yang berdistribusi eksponensial dengan parameter λ dinyatakan pada Persamaan (1) sebagai berikut (Okhli & Nooghabi, 2021):

$$f(x; \lambda) = \lambda e^{-\lambda x}, \quad x, \lambda > 0 \quad (1)$$

Rata-rata dan variansi distribusi eksponensial adalah $\mu = \frac{1}{\lambda}$ dan $\sigma^2 = \frac{1}{\lambda^2}$ (Ross, 2010). Adapun notasi dari variabel acak X yang berdistribusi eksponensial adalah $X \sim EXP(\lambda)$ (Rahmadiyah dkk., 2018).

Distribusi eksponensial sering diterapkan dalam berbagai konteks. Misalnya, dalam menghitung waktu antar kedatangan dan waktu pelayanan. Di bidang kesehatan, distribusi eksponensial digunakan untuk menganalisis data survival pasien sehingga dapat diketahui fungsi ketahanan hidup dalam situasi yang diamati (Fitria dkk., 2016). Selain itu, distribusi eksponensial sering digunakan untuk mengukur durasi waktu yang dibutuhkan dalam pelaksanaan aktivitas-aktivitas jasa tertentu.

1.6.2 Distribusi Eksponensial Terkontaminasi

Distribusi eksponensial terkontaminasi adalah model statistik yang menggabungkan distribusi eksponensial dengan unsur kontaminasi. Dalam distribusi eksponensial terkontaminasi, terdapat unsur tambahan dalam distribusi eksponensial yang mengikuti distribusi yang berbeda. Hal ini menciptakan situasi sebagian besar data mengikuti distribusi eksponensial, namun terdapat sejumlah data yang dipengaruhi oleh unsur kontaminasi.

Variabel acak yang berdistribusi eksponensial terkontaminasi dengan parameter λ , θ dan ϵ memiliki fungsi kepadatan peluang yang ditunjukkan pada Persamaan (2) sebagai berikut (Okhli dan Nooghabi, 2021):

$$f(x; \lambda, \theta, \epsilon) = (1 - \epsilon)\lambda e^{-\lambda x} + \epsilon\lambda\theta e^{-\lambda\theta x}, \quad x, \theta, \lambda > 0 \quad (2)$$

dengan:

x : nilai variabel acak

λ : parameter skala

θ : faktor kontaminasi

ϵ : proporsi campuran

Notasi dari variabel acak X yang berdistribusi eksponensial terkontaminasi adalah $X \sim CE(\lambda, \theta, \epsilon)$ dengan CE adalah singkatan dari *contaminated exponential*. Parameter $\epsilon \in [0, 1]$ dalam Persamaan (2) menunjukkan proporsi *outlier*.

Fungsi kepadatan peluang pada Persamaan (2) merupakan campuran dari dua fungsi kepadatan peluang dengan proporsi campuran ϵ . Jika x dihasilkan dari distribusi eksponensial dengan parameter λ , fungsi kepadatan peluangnya dinyatakan pada Persamaan (1). Adapun jika x muncul dari fungsi kepadatan peluang yang dinyatakan pada Persamaan (3) berikut, maka x merupakan *outlier*.

$$f(x; \lambda\theta) = \lambda\theta e^{-\lambda\theta x}, \quad x, \lambda, \theta > 0 \quad (3)$$

Jika $\theta < 1$ terjadi *outlier* atas sedangkan jika $\theta > 1$ terjadi *outlier* bawah (Okhli dan Nooghabi, 2021).

1.6.3 Deteksi *Outlier*

Hawkins (1980) mengatakan bahwa *outlier* adalah pengamatan yang sangat menyimpang dari pengamatan lain (Domanski, 2020). Adapun definisi *outlier* menurut Herdiani dkk. (2021) adalah sebagai suatu pengamatan yang tidak mengikuti sebagian besar pola data. Ghazali (2009) juga mendefinisikan *outlier* sebagai data dengan karakteristik unik yang terlihat berbeda jauh dari pengamatan lain dan muncul dalam bentuk nilai ekstrim (Shodiqin dkk., 2018). Penyimpangan *outlier* berupa data yang terlalu kecil maupun terlalu besar.

Pendeteksian *outlier* dapat dilakukan dengan melakukan visualisasi data, salah satunya dengan menggunakan *boxplot*. *Boxplot* adalah tampilan grafis dari data. Median, penyebaran atau variabilitas, penyimpangan dari simetris, dan pengamatan yang terletak jauh dari sebagian besar data (*outlier*) dapat dilihat melalui *boxplot*. *Boxplot* menampilkan tiga kuartil, nilai minimum dan maksimum dari data pada kotak persegi panjang. Kotak meliputi jarak antar kuartil dengan garis kiri (atau bawah) merupakan kuartil pertama Q_1 dan garis kanan (atau atas) merupakan kuartil ketiga Q_3 . Median $Q_2 = \bar{x}$ pada *boxplot* berupa garis yang ditarik pada kotak yang terletak di antara Q_1 dan Q_3 . Adapun *whisker* adalah garis di kedua ujung yang meluas ke nilai-nilai ekstrim (Montgomery, 2009). Suatu pengamatan dikatakan *outlier* jika pengamatan tersebut terletak di luar *whisker*.

Pendeteksian *outlier* dapat dilakukan pula dengan metode Tukey. Metode ini menggunakan kuartil dalam mendeteksi *outlier*. Interval yang digunakan pada metode Tukey dinyatakan pada Persamaan (4) berikut (Saleem dkk., 2021):

$$[Q_1 - 1.5 IQR, Q_3 + 1.5 IQR] \quad (4)$$

dengan *IQR* (*Interquartile Range*) adalah jarak antar kuartil yang diperoleh dari selisih antara kuartil ketiga dan kuartil pertama. Data yang berada diluar interval pada Persamaan (4) merupakan *outlier* (Kannan dkk., 2015).

Letak kuartil ke- i disimbolkan dengan Q_i dapat diketahui dengan menggunakan Persamaan (5) berikut (Sudjana, 1996):

$$Q_i = \text{data ke } \frac{i(n+1)}{4}, \quad i = 1, 2, 3 \quad (5)$$

dengan:

n : jumlah data

1.6.4 Uji Kecocokan Distribusi Eksponensial

Pengujian kecocokan distribusi eksponensial dapat dilakukan dengan menggunakan uji Kolmogorov-Smirnov. Uji Kolmogorov-Smirnov adalah uji statistik yang digunakan untuk menentukan apakah suatu sampel berasal dari populasi dengan distribusi tertentu (Oktaviana & Irhamah, 2021). Definisi lainnya adalah uji Kolmogorov-Smirnov merupakan uji kecocokan yang digunakan untuk mengukur seberapa baik distribusi sampel acak sama dengan distribusi teoretis (Dimitrova dkk., 2020). Dalam konteks pengujian kecocokan distribusi eksponensial, uji Kolmogorov-Smirnov adalah salah satu metode yang bisa digunakan untuk mengidentifikasi tingkat kesesuaian antara distribusi sampel data dan distribusi eksponensial yang dihipotesiskan. Hipotesis dalam uji Kolmogorov-Smirnov yang digunakan dalam pengujian kecocokan distribusi eksponensial adalah sebagai berikut:

H_0 : Data mengikuti distribusi eksponensial

H_1 : Data tidak mengikuti distribusi eksponensial

Adapun statistik uji yang digunakan dalam uji Kolmogorov-Smirnov dinyatakan pada Persamaan (6) berikut (Oktaviana & Irhamah, 2021):

$$D = \max_{1 \leq i \leq n} \left(F(X_i) - \frac{i-1}{N}, \frac{i}{N} - F(X_i) \right) \quad (6)$$

dengan:

n : jumlah data

$F(X_i)$: fungsi distribusi kumulatif dari distribusi eksponensial

Kriteria pengujiannya adalah H_0 ditolak jika nilai $D > D_\alpha$ dengan D_α adalah nilai kritis yang diperoleh dari tabel Kolmogorov-Smirnov atau $p - value < \alpha$ (Felysia dkk., 2021).

1.6.5 Metode *Maximum Likelihood*

Metode *maximum likelihood* adalah pendekatan statistik yang umum digunakan dalam penaksiran parameter distribusi data. Metode *maximum likelihood* adalah suatu pendekatan yang bertujuan untuk mencari penaksir parameter dengan nilai yang paling mungkin (maksimum) dengan cara memaksimalkan fungsi *likelihood* (Purba, 2020). Dalam konteks ini, misalkan X_1, X_2, \dots, X_n adalah variabel acak yang independen dan berdistribusi identik dengan fungsi kepadatan peluang $f(x; \theta)$ dengan θ adalah parameter yang tidak diketahui dan merupakan parameter yang akan diestimasi. Fungsi *likelihood* dari variabel acak ini dapat dirumuskan dalam Persamaan (7) berikut (Amalia dkk., 2022):

$$L(\theta; x_1, x_2, \dots, x_n) = \prod_{i=1}^n f(x_i; \theta) \quad (7)$$

Untuk melakukan estimasi parameter θ , diperkenalkan fungsi \ln *likelihood*, yang didefinisikan dalam Persamaan (8) berikut:

$$l = \ln L(\theta; x_1, x_2, \dots, x_n) = \sum_{i=1}^n \ln f(x_i; \theta) \quad (8)$$

Langkah selanjutnya dalam metode *maximum likelihood* adalah mencari nilai θ yang memaksimalkan fungsi \ln *likelihood* pada Persamaan (8). Hal ini dilakukan dengan cara menurunkan fungsi \ln *likelihood* terhadap θ kemudian disamakan dengan nol, yang ditunjukkan pada Persamaan (9) berikut:

$$\frac{dl}{d\theta} = 0 \quad (9)$$

Persamaan (9) menunjukkan bahwa untuk mengestimasi parameter yang paling sesuai dengan data yang dimiliki, dicari nilai θ yang memenuhi Persamaan (9).

Estimasi parameter model eksponensial terkontaminasi menggunakan metode *maksimum likelihood* dibahas dalam Bab III. Dalam Bab III dijelaskan proses estimasi parameter, termasuk turunan parsial pertama dari fungsi \ln *likelihood* terhadap setiap parameter. Hasil dari turunan parsial pertama fungsi \ln *likelihood* ini menghasilkan persamaan yang implisit, sehingga tidak memungkinkan estimasi parameter secara langsung. Oleh karena itu, metode Newton-Raphson digunakan sebagai solusi untuk mengestimasi parameter.

1.6.6 Metode Newton Raphson

Metode Newton Raphson adalah metode iteratif yang digunakan untuk menemukan solusi optimal dari persamaan *nonlinear*. Metode ini memanfaatkan vektor turunan pertama dan matriks turunan kedua dari fungsi terkait (Purba, 2020). Misalkan $\hat{\theta}_k = (\hat{\lambda}, \hat{\theta}, \hat{\epsilon})^T$ sebagai estimasi ke- k dari vektor parameter $\theta = (\lambda, \theta, \epsilon)^T$. Selanjutnya, $S(\hat{\theta}_k)$ adalah vektor gradien yang dinyatakan pada Persamaan (10), dan $H(\hat{\theta}_k)$ adalah matriks Hessian yang dinyatakan pada Persamaan (11) sebagai berikut (Willis dkk., 2020):

$$S(\hat{\theta}_k) = \left[\frac{\partial l}{\partial \hat{\lambda}} \quad \frac{\partial l}{\partial \hat{\theta}} \quad \frac{\partial l}{\partial \hat{\epsilon}} \right]^T \quad (10)$$

$$H(\hat{\theta}_k) = \begin{bmatrix} \frac{\partial^2 l}{\partial \hat{\lambda}^2} & \frac{\partial^2 l}{\partial \hat{\lambda} \partial \hat{\theta}} & \frac{\partial^2 l}{\partial \hat{\lambda} \partial \hat{\epsilon}} \\ \frac{\partial^2 l}{\partial \hat{\theta} \partial \hat{\lambda}} & \frac{\partial^2 l}{\partial \hat{\theta}^2} & \frac{\partial^2 l}{\partial \hat{\theta} \partial \hat{\epsilon}} \\ \frac{\partial^2 l}{\partial \hat{\epsilon} \partial \hat{\lambda}} & \frac{\partial^2 l}{\partial \hat{\epsilon} \partial \hat{\theta}} & \frac{\partial^2 l}{\partial \hat{\epsilon}^2} \end{bmatrix} \quad (11)$$

Persamaan (12) berikut adalah persamaan yang digunakan dalam algoritma Newton Raphson (Willis dkk., 2020):

$$\hat{\theta}_{k+1} = \hat{\theta}_k - H(\hat{\theta}_k)^{-1} S(\hat{\theta}_k) \quad (12)$$

Adapun algoritma Newton Raphson diuraikan sebagai berikut:

1. Menentukan estimasi awal parameter-parameter yang akan diestimasi yaitu $\hat{\theta}_0 = [\hat{\lambda}_0 \ \hat{\theta}_0 \ \hat{\varepsilon}_0]^T$.
2. Menghitung vektor gradien dan matriks Hessian seperti yang dinyatakan pada Persamaan (10) dan (11) dengan mensubstitusi estimasi awal $\hat{\theta}_0$.
3. Menghitung $\hat{\theta}_{k+1}$ dengan menggunakan Persamaan (12).
4. Melakukan pengecekan untuk menentukan apakah $\hat{\theta}_{k+1}$ sudah mencapai kondisi optimal. Pengecekan dilakukan dengan menggunakan kondisi konvergensi $|\hat{\theta}_{k+1} - \hat{\theta}_k| \leq \varepsilon$, dengan ε adalah tingkat toleransi yang diinginkan dan memiliki nilai yang kecil.

1.6.7 Akaike Information Criterion

Akaike Information Criterion (AIC) adalah indikator yang pada dasarnya menghitung informasi yang hilang pada model tertentu. Pada AIC, semakin sedikit informasi yang hilang pada suatu model maka semakin baik model tersebut, yang menandakan bahwa model tersebut mampu menjelaskan data pengamatan dengan baik. AIC dinyatakan pada Persamaan (13) sebagai berikut (Rossi dkk., 2020):

$$AIC = -2l + 2k \quad (13)$$

dengan:

l : fungsi \ln *likelihood* dari model

k : jumlah parameter yang diestimasi pada model

Model dengan nilai AIC terkecil adalah model yang lebih baik (Zhang & Meng, 2023).

1.6.8 Model Regresi Linier, Kuadrat, dan Kubik

Analisis regresi adalah metode statistik yang digunakan untuk mengetahui kekuatan hubungan antara variabel respon dan variabel prediktor. Variabel respon adalah variabel yang dipengaruhi oleh variabel prediktor, sementara variabel prediktor adalah variabel yang menyebabkan perubahan pada variabel respon (Siregar dkk., 2023). Dalam regresi linier terdapat satu variabel respon dan satu variabel prediktor. Model regresi linier ditunjukkan pada Persamaan (14) (Almumtazah dkk., 2021).

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad (14)$$

dengan:

y_i : nilai variabel respon pada pengamatan ke- i untuk $i = 1, 2, \dots, n$

x_i : nilai variabel prediktor pada pengamatan ke- i untuk $i = 1, 2, \dots, n$

β_0 : intersep

β_1 : parameter regresi

ε_i : galat ke- i untuk $i = 1, 2, \dots, n$

Regresi *non-linier* adalah metode analisis regresi yang bertujuan mendapatkan model *non-linier* untuk mengidentifikasi hubungan antara variabel respon dengan variabel prediktor. Regresi kuadrat adalah bentuk analisis regresi *non-linier* dimana hubungan antara variabel respon dan satu atau lebih variabel prediktor dimodelkan

sebagai polinomial derajat dua. Model regresi kuadratik ditunjukkan pada Persamaan (15) (Siregar dkk., 2023).

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \varepsilon_i \quad (15)$$

dengan:

- y_i : nilai variabel respon pada pengamatan ke- i untuk $i = 1, 2, \dots, n$
- x_i : nilai variabel prediktor pada pengamatan ke- i untuk $i = 1, 2, \dots, n$
- β_0 : intersep
- β_1, β_2 : parameter regresi
- ε_i : galat ke- i untuk $i = 1, 2, \dots, n$

Model regresi kubik adalah bentuk analisis regresi di mana hubungan antara variabel dependen dan satu atau lebih variabel independen dimodelkan sebagai polinomial derajat tiga. Model regresi kubik ditunjukkan pada Persamaan (16) (Hermanto & Rizqika, 2019).

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \beta_3 x_i^3 + \varepsilon_i \quad (16)$$

dengan:

- y_i : nilai variabel respon pada pengamatan ke- i untuk $i = 1, 2, \dots, n$
- x_i : nilai variabel prediktor pada pengamatan ke- i untuk $i = 1, 2, \dots, n$
- β_0 : intersep
- $\beta_1, \beta_2, \beta_3$: parameter regresi
- ε_i : galat ke- i untuk $i = 1, 2, \dots, n$

Estimasi parameter model regresi linier, kuadratik dan kubik dapat dilakukan dengan menggunakan metode kuadrat terkecil. Tujuan metode kuadrat terkecil adalah untuk meminimalkan jumlah kuadrat galat. Adapun estimasi parameter model regresi linier, kuadratik dan kubik dapat dihitung dengan menggunakan Persamaan (17) (Augini dkk., 2019):

$$\hat{\beta} = (X'X)^{-1}(X'Y) \quad (17)$$

dengan:

$$\hat{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix}$$

$$X = \begin{bmatrix} 1 & X_{11} & X_{12} & \dots & X_{1p} \\ 1 & X_{21} & X_{22} & \dots & X_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n1} & X_{n2} & \dots & X_{np} \end{bmatrix}$$

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}$$

$\hat{\beta}_p$: parameter regresi variabel prediktor ke- p

X_{ik} : Nilai pengamatan variabel prediktor ke- k pada pengamatan ke- i

k : jumlah variabel prediktor dengan $k = 1, 2, \dots, p$

Adapun untuk menilai kebaikan suatu model, dapat digunakan koefisien determinasi. Koefisien determinasi adalah nilai yang menunjukkan seberapa besar variabel respon data dapat dijelaskan oleh variabel prediktor. Nilai koefisien determinasi dapat dihitung dengan menggunakan Persamaan (18) berikut:

$$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (18)$$

dengan:

y_i : nilai variabel respon pada pengamatan ke- i untuk $i = 1, 2, \dots, n$

\hat{y}_i : nilai prediksi pada pengamatan ke- i untuk $i = 1, 2, \dots, n$

\bar{y} : rata-rata dari nilai variabel respon

Jika nilai R^2 mendekati 1, maka hal tersebut menunjukkan bahwa data sesuai dengan model yang diperoleh (Nabila & Yotenka, 2021).

1.6.9 Gempa Bumi

Gempa bumi adalah salah satu jenis bencana yang disebabkan oleh faktor alam yang kompleks dan melibatkan pergeseran lempeng tektonik di dalam kerak bumi, sehingga menghasilkan getaran dan perubahan geologis yang dapat menyebabkan kerusakan pada lingkungan dan kehidupan manusia. Gempa bumi merupakan berguncangnya bumi yang disebabkan oleh patahan aktif, tumbukan antar lempeng bumi, runtuhannya batuan dan aktifitas gunung api (Fadilah & Ronoatmojo, 2022). Gempa bumi merupakan fenomena alam yang terjadi secara acak yang tidak beraturan dalam ruang dan waktu (Pratiwi dkk., 2018). Gempa bumi sebagai kekuatan alam yang luar biasa, memiliki potensi untuk menimbulkan bencana yang besar dan merugikan. Gempa bumi mengakibatkan pergerakan tanah yang kuat dan getaran hebat, yang bisa menimbulkan kerusakan serius pada bangunan, infrastruktur, dan juga mengancam kehidupan manusia serta lingkungan di sekitarnya.

1. Jenis-jenis gempa bumi yaitu (Fadilah & Ronoatmojo, 2022):

a. Gempa bumi vulkanik

Gempa bumi vulkanik merupakan gempa berkekuatan kecil dan sifatnya lokal, yang terkait erat dengan aktivitas vulkanik pada gunung berapi. Proses terjadinya gempa ini diawali oleh naiknya fluida gunung api berupa gas, uap, dan magma dari dalam bumi yang bergerak menuju ke permukaan kawah. Saat fluida ini mendekati permukaan, tekanan yang terkumpul dapat menyebabkan retakan, dan retakan ini

merupakan sumber gempa bumi vulkanik. Getaran yang timbul dari gempa ini menyebar di sekitar rekahan tersebut dan merambat ke segala arah, seringkali dengan intensitas yang cukup rendah, tetapi tetap mampu menciptakan dampak yang signifikan di sekitar area vulkanik tersebut.

b. Gempa bumi runtuhan

Gempa bumi runtuhan merupakan gempa dengan kekuatan yang relatif kecil dan cenderung terjadi dalam skala lokal. Gempa ini muncul akibat runtuhannya batuan yang terjadi terutama di daerah kapur, dan penyebab utamanya seringkali dikaitkan dengan kondisi geologi di wilayah tersebut. Proses pada gempa bumi runtuhannya ini seringkali berawal dari adanya stalagtit yang mengalami keruntuhan di dalam gua-gua kapur yang terbentuk oleh proses pelarutan selama ribuan tahun. Ketika stalagtit atau struktur batuan lainnya jatuh, mereka dapat menciptakan getaran dan gelombang kejut di sekitarnya, yang merambat sebagai gempa bumi lokal dengan dampak yang dapat terasa di sekitar lokasi tersebut.

c. Gempa bumi tektonik

Gempa bumi tektonik adalah gempa bumi yang disebabkan oleh aktivitas di zona batas antar lempeng dan patahan, dimana pergeseran tektonik lempeng bumi secara periodik menghasilkan tekanan dan ketegangan besar. Getaran yang dihasilkan oleh gempa bumi tektonik ini merambat melalui kerak bumi, menjalar ke segala arah, dan dapat menciptakan gelombang seismik yang terasa di permukaan bumi.

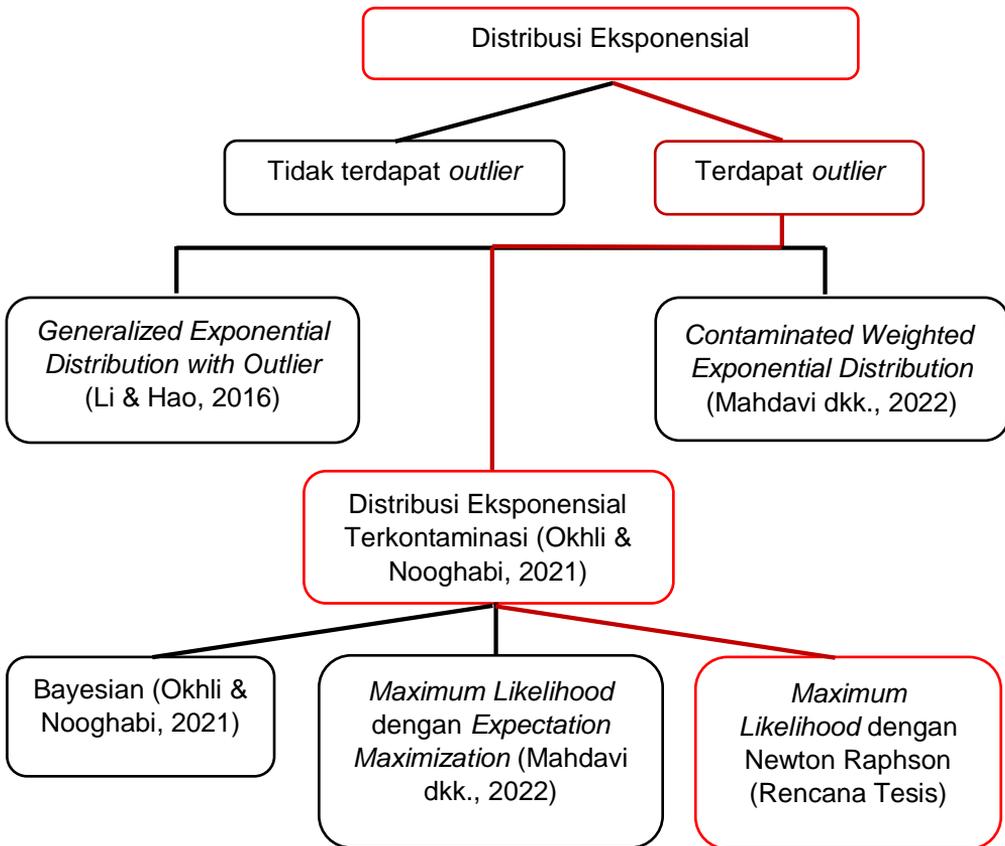
2. Parameter gempa bumi

Parameter gempa bumi adalah acuan mengenai besaran dan lokasi terjadinya gempa bumi. Besaran gempa bumi adalah ukuran yang menggambarkan tingkat kekuatan gempa dan diukur dengan memanfaatkan data yang tercatat oleh perangkat pencatat gempa, atau yang dikenal sebagai seismograf.

Parameter gempa bumi meliputi (Shohaya dkk., 2013):

- a) Waktu kejadian gempa bumi
- b) Hiposenter
- c) Episenter
- d) Kedalaman sumber gempa bumi
- e) Kekuatan gempa bumi

1.6.10 Kerangka Konseptual



Gambar 2. Kerangka konseptual

BAB II

METODE PENELITIAN

2.1 Sumber Data

Data yang digunakan pada penelitian ini adalah data sekunder yaitu data yang diperoleh dari katalog *United States Geological Survey* (USGS) periode 1971-2022 di pulau Jawa dan sekitarnya. Data tersebut dapat diakses melalui *website* <https://earthquake.usgs.gov/earthquakes/search/>. Magnitude yang dipilih adalah ≥ 5 dengan kedalaman 0-100 km. Adapun batas titik koordinat yang dipilih adalah $5.5^{\circ} - 9.0^{\circ} LS$ dan $105.5^{\circ} - 114.5^{\circ} BT$ yang wilayahnya ditunjukkan pada Gambar 3 berikut.



Gambar 3. Peta pulau Jawa dan sekitarnya
(<https://earthquake.usgs.gov/earthquakes/search/>)

2.2 Langkah Analisis

Langkah-langkah analisis data dalam penelitian ini adalah sebagai berikut:

1. Deskripsi data waktu antar kejadian gempa bumi sebagai informasi awal tentang data tersebut.
2. Melakukan pendeteksian *outlier* terhadap data waktu antar kejadian gempa bumi menggunakan *boxplot* dan metode Tukey. Adapun langkah-langkah untuk mendeteksi *outlier* dengan metode Tukey adalah sebagai berikut:
 - a) Menghitung Q_1 dan Q_3 dengan menggunakan Persamaan (5) berikut:

$$Q_i = \text{data ke } \frac{i(n+1)}{4}, i = 1,2,3 \quad (5)$$

- b) Menghitung IQR yang merupakan selisih antara Q_3 dengan Q_1 .
- c) Mensubstitusi hasil yang diperoleh ke dalam Persamaan (4) berikut:

$$[Q_1 - 1.5 IQR, Q_3 + 1.5 IQR] \quad (4)$$

Data yang berada diluar interval pada Persamaan (4) merupakan *outlier*.

3. Melakukan pengujian distribusi eksponensial terhadap data waktu antar kejadian gempa bumi menggunakan uji Kolmogorov-Smirnov.

Hipotesis yang digunakan adalah:

H_0 : Data mengikuti distribusi eksponensial

H_1 : Data tidak mengikuti distribusi eksponensial

Adapun statistik ujinya dinyatakan pada Persamaan (6) sebagai berikut:

$$D = \max_{1 \leq i \leq n} \left(F(X_i) - \frac{i-1}{N}, \frac{i}{N} - F(X_i) \right) \quad (6)$$

H_0 ditolak jika nilai $D > D_\alpha$ dengan D_α adalah nilai kritis yang diperoleh dari tabel Kolmogorov-Smirnov atau p -value $< \alpha$.

4. Melakukan estimasi parameter model eksponensial terkontaminasi menggunakan metode *maximum likelihood* melalui iterasi Newton Raphson.
- Menentukan fungsi *likelihood* dari fungsi kepadatan peluang distribusi eksponensial terkontaminasi.

$$L(\lambda, \theta, \epsilon; x_1, x_2, \dots, x_n) = \prod_{i=1}^n f(x_i; \lambda, \theta, \epsilon)$$

- Menentukan fungsi *ln likelihood*.

$$l = \ln L(\lambda, \theta, \epsilon; x_1, x_2, \dots, x_n) = \ln \left[\prod_{i=1}^n f(x_i; \lambda, \theta, \epsilon) \right]$$

- Mencari turunan parsial pertama fungsi *ln likelihood* terhadap parameter λ , θ , dan ϵ kemudian disamakan dengan nol.
- Mencari turunan parsial kedua fungsi *ln likelihood* karena turunan parsial pertama fungsi *ln likelihood* terhadap parameter λ , θ , dan ϵ implisit.
- Melakukan estimasi parameter dengan menggunakan iterasi Newton Raphson yang algoritmanya adalah sebagai berikut:
 - Menentukan estimasi awal parameter-parameter yang akan diestimasi yaitu $\hat{\theta}_0 = [\hat{\lambda}_0 \quad \hat{\theta}_0 \quad \hat{\epsilon}_0]^T$.
 - Menghitung vektor gradien dan matriks Hessian seperti yang dinyatakan pada Persamaan (10) dan (11) dengan mensubstitusi estimasi awal $\hat{\theta}_0$.

$$S(\hat{\theta}_k) = \left[\frac{\partial l}{\partial \hat{\lambda}} \quad \frac{\partial l}{\partial \hat{\theta}} \quad \frac{\partial l}{\partial \hat{\epsilon}} \right]^T \quad (10)$$

$$H(\hat{\theta}_k) = \begin{bmatrix} \frac{\partial^2 l}{\partial \hat{\lambda}^2} & \frac{\partial^2 l}{\partial \hat{\lambda} \partial \hat{\theta}} & \frac{\partial^2 l}{\partial \hat{\lambda} \partial \hat{\epsilon}} \\ \frac{\partial^2 l}{\partial \hat{\theta} \partial \hat{\lambda}} & \frac{\partial^2 l}{\partial \hat{\theta}^2} & \frac{\partial^2 l}{\partial \hat{\theta} \partial \hat{\epsilon}} \\ \frac{\partial^2 l}{\partial \hat{\epsilon} \partial \hat{\lambda}} & \frac{\partial^2 l}{\partial \hat{\epsilon} \partial \hat{\theta}} & \frac{\partial^2 l}{\partial \hat{\epsilon}^2} \end{bmatrix} \quad (11)$$

- Menghitung $\hat{\theta}_{k+1}$ dengan menggunakan Persamaan (12) berikut:

$$\hat{\theta}_{k+1} = \hat{\theta}_k - H(\hat{\theta}_k)^{-1} S(\hat{\theta}_k) \quad (12)$$

- 4) Melakukan pengecekan untuk menentukan apakah $\hat{\theta}_{k+1}$ sudah mencapai kondisi optimal. Pengecekan dilakukan dengan menggunakan kondisi konvergensi $|\hat{\theta}_{k+1} - \hat{\theta}_k| \leq \varepsilon$, dengan ε adalah tingkat toleransi yang diinginkan dan memiliki nilai yang kecil.
5. Melakukan pemodelan waktu antar kejadian gempa bumi dengan menggunakan distribusi eksponensial dan eksponensial terkontaminasi.
6. Melakukan pemilihan model terbaik antara model eksponensial dan model eksponensial terkontaminasi dengan menggunakan nilai AIC yang dinyatakan pada Persamaan (13). Model dengan nilai AIC terkecil adalah model terbaik.

$$AIC = -2l + 2k \quad (13)$$

7. Melakukan penentuan model persamaan linier, kuadratik, dan kubik yang langkah-langkahnya sebagai berikut:
 - 1) Menentukan variabel respon dan variabel prediktor, dengan variabel respon adalah peluang waktu antar kejadian gempa bumi dan variabel prediktor adalah waktu antar kejadian gempa bumi.
 - 2) Menghitung data pada variabel x^2 dengan mengkuadratkan data variabel x dan menghitung data pada variabel x^3 dengan mengkubikkan data variabel x .
 - 3) Melakukan estimasi parameter model persamaan linier, kuadratik dan kubik dengan menggunakan Persamaan (17).

$$\hat{\beta} = (X'X)^{-1}(X'Y) \quad (17)$$

- 4) Melakukan pemodelan dengan menggunakan persamaan linier, kuadratik, dan kubik.
- 5) Melakukan pemilihan model terbaik antara model persamaan linier, kuadratik dan kubik yang diperoleh dengan menggunakan nilai koefisien determinasi yang dinyatakan pada Persamaan (18).

$$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (18)$$

- 6) Melakukan perkiraan peluang waktu antar kejadian gempa bumi kedepan berdasarkan model terbaik yang diperoleh.
8. Membuat kesimpulan dari hasil yang diperoleh.