Sains Malaysiana 52(2)(2023): 655-667
http://doi.org/10.17576/jsm-2023-5202-25

On the Earthquake Distribution Modeling in Sumatra by Cauchy Cluster Process: Comparing Log-Linear and Log-Additive Intensity Models

(Mengenai Pemodelan Taburan Gempa Bumi di Sumatera oleh Proses Kelompok Cauchy: Membandingkan Model Keamatan Log-Linear dan Log-Tambahan)

TABITA YUNI SUSANTO, ACHMAD CHOIRUDDIN* & JERRY DWI TRIJOYO PURNOMO

Department of Statistics, Faculty of Science and Data Analytics, Institut Teknologi Sepuluh Nopember (ITS), Surabaya, Indonesia

Received: 7 August 2022/Accepted: 16 November 2022

ABSTRACT

Inhomogeneous cluster point processes have been considered for modeling the distribution of earthquake epicenters with the spatial trend and clustering patterns. In particular, the spatial trend is assessed by the intensity model involving geological variables. However, for intensity with a log-linear form, it may be too restrictive and not appropriate for earthquake distribution. In this study, we consider the Cauchy cluster process with the log-additive intensity model to analyze the distribution of major earthquake occurrences in Sumatra, Indonesia. The estimation procedure follows the standard two-step estimation technique, where the first step adapts the method for the Generalized Additive Models (GAMs) using penalized iteratively reweighted least squares (PIRLS) algorithm, and the second step employs the second-order composite likelihood. For the earthquake analysis in Sumatra, the log-additive intensity shows more flexibility to determine the contribution of each geological factor, especially to capture the effect of the nearest distance to the fault which is far from log-linear. In addition, compared to the log-linear model, the Cauchy cluster process with a log-additive intensity model performs better with a smaller Akaike Information Criterion's (AIC) value and a sharper envelope K-function. The estimated number of mainshocks is around 114 with aftershocks spread by 14 km around the mainshocks. We detect three hotspots for the major earthquake in Sumatra: the northern part (Aceh and North Sumatra), the western part (Mentawai, Nias, and Simeulue), and Bengkulu.

Keywords: Disaster risk reduction; earthquake modeling; generalized additive models; spatial point process; subduction

ABSTRAK

Proses titik kelompok tidak homogen telah dipertimbangkan untuk memodelkan taburan pusat gempa bumi dengan arah aliran ruang dan corak kelompok. Khususnya, trend ruang dinilai oleh model keamatan yang melibatkan pemboleh ubah geologi. Walau bagaimanapun, untuk keamatan dengan bentuk log-linear, ia mungkin terlalu terkawal dan tidak sesuai untuk taburan gempa bumi. Dalam kajian ini, kami mempertimbangkan proses kelompok Cauchy dengan model keamatan log-tambahan untuk menganalisis taburan kejadian gempa bumi besar di Sumatera, Indonesia. Prosedur anggaran mengikut teknik anggaran dua langkah piawai dengan langkah pertama menyesuaikan kaedah untuk Model Tambahan Am (GAM) menggunakan algoritma kuasa dua terkecil ditimbang semula secara berulang (PIRLS) berhukum dan langkah kedua menggunakan kemungkinan komposit tertib kedua. Bagi analisis gempa bumi di Sumatera, keamatan log-tambahan menunjukkan lebih kefleksibelan untuk menentukan sumbangan setiap faktor geologi, terutamanya untuk menangkap kesan jarak terdekat dengan sesar yang jauh daripada log-linear. Di samping itu, berbanding model log-linear, proses kelompok Cauchy dengan model keamatan log-tambahan berprestasi lebih baik dengan nilai Kriteria Maklumat Akaike (AIC) yang lebih kecil dan fungsi K sampul yang lebih tajam. Anggaran bilangan gegaran utama adalah sekitar 114 dengan gegaran susulan tersebar sejauh 14 km di sekitar gegaran utama. Kami mengesan tiga titik panas untuk gempa bumi besar di Sumatera: bahagian utara (Aceh dan Sumatera Utara), bahagian barat (Mentawai, Nias dan Simeulue) dan Bengkulu.

Kata kunci: Model tambahan am; pemodelan gempa bumi; pengurangan risiko bencana; proses titik ruang; subduksi

Introduction

Indonesia is located at the confluence of tectonic plates, resulting in the territory of Indonesia having more than 295 active faults, making Indonesia an earthquake-prone area. Sumatra is one of the most active tectonic areas in the world due to the activity of the Indo-Australian and Eurasian plates in the subduction zone along the Sumatran Sea (Figure 1). Sumatra is crossed by faults in the earth's crust along the Bukit Barisan and faults in the earth's crust at the bottom of the Indian Ocean along the west coast of Sumatra which makes Sumatra the most vulnerable area to earthquakes (Sosilawati et al. 2017). There are three sources of earthquake threat in Sumatra, namely the subduction area, the Mentawai Fault System, and the Sumatra Fault System (Triyono 2015). For an area susceptible to earthquakes like Sumatra, it is major important to detect the earthquake hotspots due to seismic activities, so risk calculation and mitigation can be conducted more accurately. The hotspot detection can be quantified by studying the earthquake distribution concerning geological variables.

The government has developed the Probabilistic Seismic Hazard (PSHA) and the Earthquake Early Warning System (EEWS). The EEWS includes three systems i.e., a monitoring system which detects earthquakes upstream, an automatic processing system which processes data quickly, and an information dissemination system (Ibrahim 2019). The PSHA method is a probabilistic earthquake hazard analysis method by calculating and combining the uncertainties of the magnitude, location, and time of the earthquake (Nasional 2017). Nevertheless, the PSHA method assumes that earthquakes originate from the Poisson process, meaning that earthquake events which produce ground shaking at one point for a certain level are independent events from other earthquake events. Moreover, the PSHA method only considers the hazard of the mainshocks, albeit aftershock occurrences also often cause major issues. In addition, PSHA only considers the uncertainty factor of the magnitude, location, and time of the earthquake, whereas geological factors such as active faults and subduction zones also may influence the occurrence of earthquakes.

Spatial point process becomes a standard tool for earthquake modeling when the focus is to model the spatial distribution of locations of earthquake occurrences (Geng, Shi & Hu 2021; Iftimi, Cronie & Montes 2019; Türkyilmaz, van Lieshout & Stein 2013). When the past studies rely mainly on stationary point processes (Ogata 2006, 1988), recent studies employ non-stationary point processes by involving geological factors in the intensity model to enhance interpretability and model improvement (Anwar, Yaseen & Mahmood 2023; Choiruddin et al. 2021; D'Angelo et al. 2022; Siino et al. 2017). In particular, the log-linear intensity model is employed to improve the prediction of earthquake distribution (Choiruddin et al. 2021; D'Angelo et al. 2022; Husain & Choiruddin 2021).

Within the point process framework, modeling of the distribution of earthquake occurrences in Sumatra has been carried out using a different point process model (Choiruddin et al. 2023; Pratiwi et al. 2021; Pratiwi, Rini & Mangku 2018), especially research by Choiruddin, Susanto and Metrikasari (2021) and Trisnisa et al. (2019) which considers geological factors. Sumatra has a unique pattern where spatial trend due to spatial covariate makes a stronger effect than the clustering due to seismic activity (Choiruddin, Susanto & Metrikasari 2021). However, capturing the spatial trend effect through the log-linear intensity model may not be appropriate especially when the effect of geological variables is not log-linear (Figure 3). For example, Figure 3(b) depicts the log-linear model (red dashed line) for capturing the effect of fault is a very crude estimation for the intensity model (grey line). Therefore, a more flexible model should be considered to cover the nonlinearity relationship and to improve the log-linear model.

This study focuses on comparing the performance of log-linear and log-additive intensities to model the distribution of earthquake locations in Sumatra involving two geological factors: faults and subduction. The Cauchy cluster process is employed as it performs best to model earthquake distribution among other cluster processes (Choiruddin et al. 2021; Choiruddin, Susanto & Metrikasari 2021). The log-additive intensity model is considered to relax the log-linear assumption model. Jalilian (2017) and Youngman and Economou (2017) demonstrate the good performance of the log-additive intensity for point process models applied to Ecology and Environmental Science. We intend to extend the model in Seismology by mapping the earthquake risk prediction in Sumatra. Earthquake risk prediction maps can detect hotspots and can be used as guidelines to be more alert in dealing with earthquake disasters and to reduce the impact of earthquakes.

Study Area and Variable Description

The study area is located in the Sumatra region (Figure 1), which includes the location of the occurrence of earthquakes, subduction zones, and active faults.

Fault and subduction coordinates are taken from Peta Sumber dan Bahaya Gempa Indonesia (Nasional 2017). According to USGS (https://earthquake.usgs.gov/earthquakes/search), from 2004 to 2018 there have been 10868 earthquakes in Sumatra, with 1195 of them having a magnitude M≥ 5 i.e., the Bengkulu earthquake (2004, M=7.3), the Aceh-Andaman earthquake (2004, M=9), and the North Sumatra earthquake (2005, M=8.6) (BMKG 2019). Figure 1 shows that the earthquake locations are distributed inhomogeneously. Earthquake occurrences tend to be clustered in certain areas (Aceh, Simeulue, Nias, Mentawai, West Sumatra, and Lampung). Earthquake densities are high in areas close to subduction zones and faults, widespread with a radius of ±300 km (Figure 2). The geological covariates in this study are shown in Table 1.

This study explores the effect of geological variables through the log-additive intensity model as an alternative to the log-linear model. Figure 3 shows the estimated earthquake intensity as a function of each covariate with the gray area being a 95% confidence interval assuming an inhomogeneous Poisson point process. Nonparametric estimates do not assume a particular model form for the relationship between earthquake intensity and specific spatial covariates. Parametric estimates of the effects of geological variables on the log-linear intensity model are represented by red dotted lines. The nonparametric estimates and envelopes clearly show that the influence of geological variables is far from log-linear, particularly on the effect of faults. Therefore, it is of interest to extend the log-linear intensity model to the log-additive model.

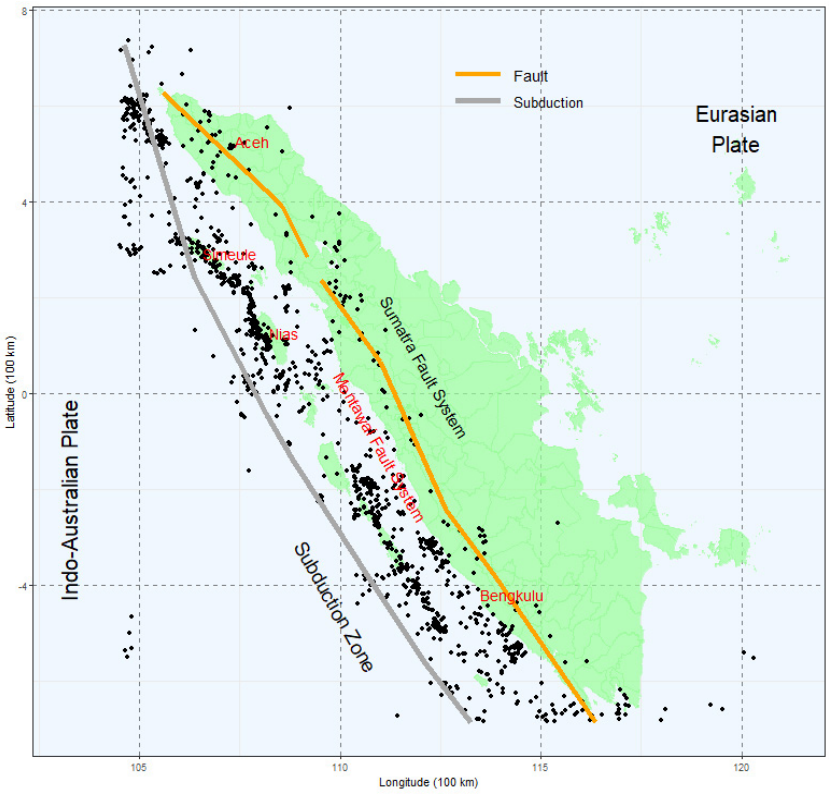
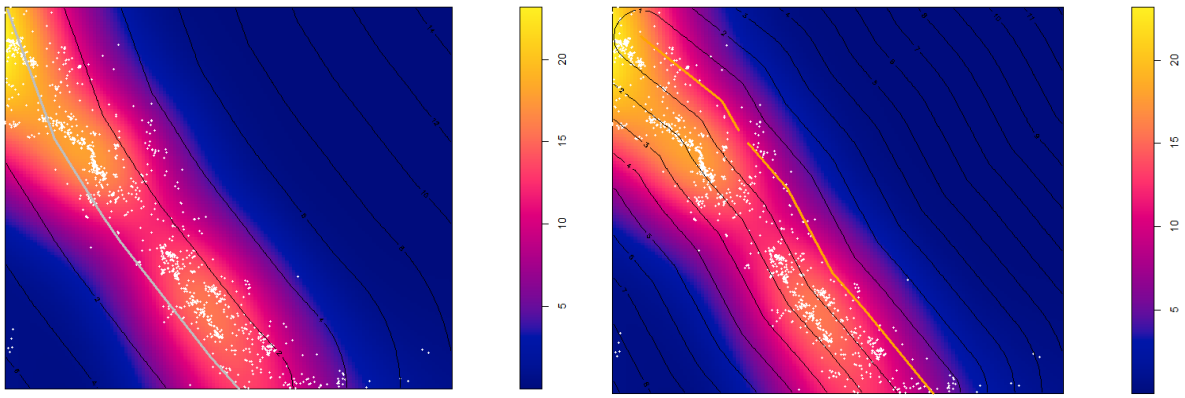
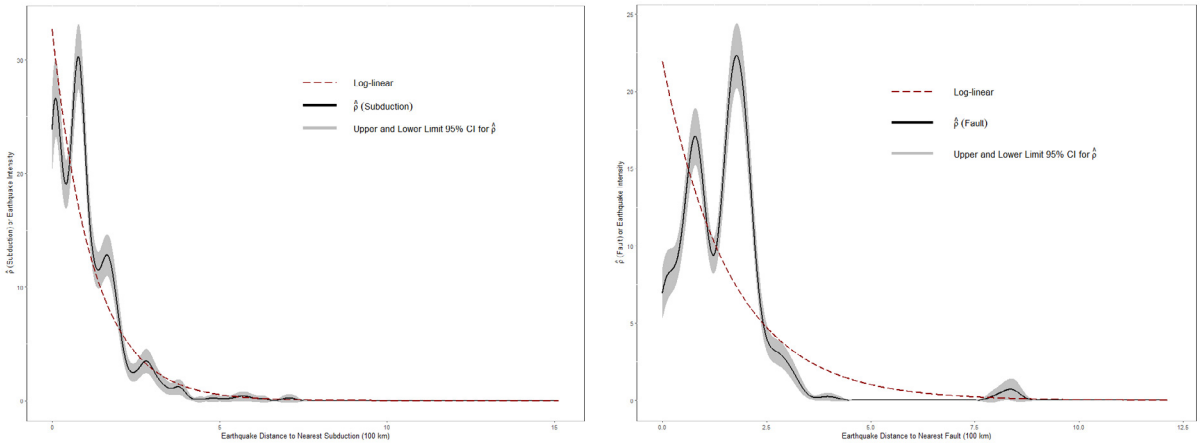


FIGURE 1. Earthquake distribution map in Sumatra 2004-2018 (black dot is the location of the earthquake, grey line is subduction, and orange line is a fault) with an area of $|B| = [104.52;\ 121{,}18] \times [-6.86;\ 7.38]$ 100 km^2



(a) (b)

FIGURE 2. Earthquake density with the contour plot of the distance of the earthquake to (a) subductions and (b) faults (white dots are earthquake locations)



(a) (b)

FIGURE 3. Nonparametric estimation of earthquake intensity as a function of covariates distance from earthquake to (a) subductions and (b) faults

TABLE 1. Covariates description

Variable	Description
$z_1(\boldsymbol{u})$	Earthquake distance to nearest fault (100 km)
$z_2(\boldsymbol{u})$	Earthquake distance to nearest subduction (100 km)

SPATIAL POINT PROCESS

Spatial point process $\mathbf{X}$ is a random mechanism whose result is a point pattern. Spatial point pattern $\mathbf{x}$ is a dataset which contains a set of object or event locations observed in a certain observation window B, $B \subseteq R^2$ written $\mathbf{x} = \{x_1, ..., x_n\}$ (Baddeley, Rubak & Turner 2015). In this study, $\mathbf{x}$ is the occurrence of an earthquake. Given $\lambda(\boldsymbol{u}) : B \to [0, \infty)$ as the intensity function of the point process $\mathbf{X}$, then $\lambda(\boldsymbol{u})$ is the intensity function of a point process related to the expected number of points in an area.

$$\mu(B) = E\left[N(B)\right] = \int_B \lambda(\boldsymbol{u})\, d\boldsymbol{u},\ B \subseteq R^2 \quad . \tag{1}$$

The second order moment for the point process is related to calculating the relationship between points. The second order of a point process in a region B is $\mu^{(2)}$ as shown in Equation (2). $\lambda^{(2)}(\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{\psi})\, d\boldsymbol{u}\, d\boldsymbol{v}$ is the probability of observing a pair of points in two very small regions with centers $\boldsymbol{u}$ and $\boldsymbol{v}$, volumes $d\boldsymbol{u}$ and $d\boldsymbol{v}$, $\mathbf{1}$ is an indicator function which is given a value of 1 if $\boldsymbol{u}$ and $\boldsymbol{v}$ are elements of B (Møller & Waagepetersen 2004).

$$\mu^{(2)}(B) = \int_B \int_B \mathbf{1}\{(\boldsymbol{u}, \boldsymbol{v}) \in B\} \lambda^{(2)}(\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{\psi})\, d\boldsymbol{u}\, d\boldsymbol{v} \tag{2}$$

To find out the distribution pattern of the point process, we can consider the pair correlation function or the K-function (Baddeley, Rubak & Turner 2015).

CAUCHY CLUSTER PROCESS

Suppose $\mathbf{C}$ is a Poisson point process (mainshocks process) with intensity $\kappa 0$. Conditional on $\mathbf{C}$, $\mathbf{X}_c, c \in \mathbf{C}$ is an independent Poisson process (offspring processes) with a log-linear intensity function $\lambda_c(\boldsymbol{u})$ to detect the spatial trend due to P spatial covariates (Choiruddin et al. 2021)

$$\lambda_c(\boldsymbol{u}; \boldsymbol{\theta}) = \exp\left(\zeta + \sum_{p=1}^{P} \theta_p z_p(\boldsymbol{u})\right) k(\boldsymbol{u} - \boldsymbol{c}; \omega), \tag{3}$$

where θ_p is a coefficient of the parametric model; $k(\boldsymbol{u})$ is the probability density function of the distance distribution between the aftershocks and the mainshocks process which is mutually independent and has a bivariate Cauchy distribution with a scale parameter ω given by

$$k(\boldsymbol{u}; \omega) = \left(2\pi\omega^2\right)^{-1} \left(1 + \left(\|\boldsymbol{u}\| / \omega\right)^2\right)^{-3/2}. \tag{4}$$

In this study, we extend the log-linear model in Equation (3) and replace it with the log additive model presented in Equation (5)

$$\lambda_c(\boldsymbol{u}; \boldsymbol{\theta}) = \exp\left(\zeta + \sum_{p=1}^{P} f_p\left(z_p(\boldsymbol{u})\right)\right) k(\boldsymbol{u} - \boldsymbol{c}; \omega), \tag{5}$$

where $f_p\left(z_p(\boldsymbol{u})\right)$ is a smooth function which can be represented by basis expansions detailed herewith. Therefore, $\mathbf{X} = \cup_{c \in \mathbf{C}} \mathbf{X}_c$ is the Cauchy cluster process with the intensity function, pair correlation function, and K-function in Equations (6)-(8) (Ghorbani 2013).

$$\begin{aligned} \lambda(\boldsymbol{u}; \boldsymbol{\theta}) &= \kappa \exp\left(\zeta + \sum_{p=1}^{P} f_p\left(z_p(\boldsymbol{u})\right)\right) \\ &= \exp\left(\theta_0 + \sum_{p=1}^{P} \sum_{j=1}^{K} \theta_{p,j} b_{p,j}\left(z_p(\boldsymbol{u})\right)\right) \end{aligned} \tag{6}$$

$$g(\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{\psi}) = 1 + \frac{1}{8\pi\kappa\omega^2} \left(1 + \left(\|\boldsymbol{u} - \boldsymbol{v}\|^2 / 4\omega^2\right)\right)^{-3/2} \tag{7}$$

$$K(\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{\psi}) = \pi \|\boldsymbol{u} - \boldsymbol{v}\|^2 + \frac{1}{\kappa} \left(1 - \frac{1}{\sqrt{1 + \left(\|\boldsymbol{u} - \boldsymbol{v}\|^2 / 4\omega^2\right)}}\right) \tag{8}$$

The smooth functions $f_p(z_p(\boldsymbol{u}))$ are represented by the basis functions $b_{p,j}$ and coefficients of the basis $\theta_{p,j}$, $j = 1, ..., K$, $p = 1, ..., P$ (Pedersen et al. 2018). $\theta_0 = \zeta + \log\kappa$ is an intercept parameter; κ is the intensity of the mainshocks; and ω represents a scaling parameter related to the deviation of the distance between aftershocks to the mainshocks. $\boldsymbol{\psi} = (\boldsymbol{\theta}^T, \kappa, \omega)^T$ is the parameter to be estimated. We use cubic regression spline which is one-dimensional smoothing as the basis function. The optimum number of knots k ($K = k - 1$) is selected based on the minimum AIC (Akaike Information Criterion).

PARAMETER ESTIMATION

We use a two-step procedure to estimate parameters $\boldsymbol{\psi} = (\boldsymbol{\theta}^T, \kappa, \omega)^T$, where the first step is to estimate $\boldsymbol{\theta}$ by first-order composite likelihood, and the second step is to estimate κ and ω by second-order composite likelihood. The first step procedure is linked to parameter estimation in the Generalized Additive Models (GAMs) detailed in the next section.

GENERALIZED ADDITIVE MODEL (GAM)

The composite log-likelihood function to estimate $\boldsymbol{\theta}$ is as follows.

$$\log L(\boldsymbol{u}; \boldsymbol{\theta}) = \sum_{i=1}^{n} \log\left(\lambda(\boldsymbol{u}_i; \boldsymbol{\theta})\right) - \int_D \lambda(\boldsymbol{u}; \boldsymbol{\theta}) d\boldsymbol{u} \tag{9}$$

Using a numerical quadrature approach $\int_B \lambda(\boldsymbol{u};\boldsymbol{\theta})\,d\boldsymbol{u} \approx \sum_{i=1}^{n+d} \lambda(\boldsymbol{u}_i;\boldsymbol{\theta}) w_i$ with n number of data points, d number of dummy points, $y_i = \frac{I_i}{w_i}$, $I_i = 1$ for $\boldsymbol{u}_i$ is data points and $I_i = 0$ if $\boldsymbol{u}_i$ is dummy point, and w_i is quadrature weight, then equation 9 reduces to

$$\log L(\boldsymbol{u};\boldsymbol{\theta}) \approx \sum_{i=1}^{n+d} w_i \left(y_i \log\left(\lambda(\boldsymbol{u}_i;\boldsymbol{\theta})\right) - \lambda(\boldsymbol{u}_i;\boldsymbol{\theta}) \right). \quad (10)$$

The log-likelihood form in Equation (10) is equivalent to the log-likelihood of the weighted Poisson regression. Poisson regression modeling is obtained from the link function which explains the relationship of the expected value of y_i with covariates. The log-linear model assumes a linear relationship between intensity and covariates, so to be more flexible in modeling, this study uses the Generalized Additive Model. GAM is the development of GLM with additive predictors involving smooth functions and has a structure in Equation (11),

$$g(\mu) = f_1(z_1(\boldsymbol{u})) + f_2(z_2(\boldsymbol{u})) + f_3(z_3(\boldsymbol{u})) + \ldots \quad (11)$$

where $f_p(z_p(\boldsymbol{u}))$ is the smooth functions of geological covariates $z_p(\boldsymbol{u})$ which are estimated nonparametrically, $g(\mu)$ is the link function, $\mu = E(Y)$, and Y has an exponential family distribution (Wood 2017). The log-likelihood function with the intensity function $\lambda(\boldsymbol{u}_i;\boldsymbol{\theta})$ using GAM can be seen in Equation (12).

$$\lambda(\boldsymbol{u}_i;\boldsymbol{\theta}) = \theta_0 + \sum_{p=1}^{P} f_p(z_p(\boldsymbol{u}_i))$$

$$\log L(\boldsymbol{u};\boldsymbol{\theta}) \approx \sum_{i=1}^{n+d} w_i \left(y_i \left(\theta_0 + \sum_{p=1}^{P} f_p(z_p(\boldsymbol{u}_i)) \right) - \exp \left(\theta_0 + \sum_{p=1}^{P} f_p(z_p(\boldsymbol{u}_i)) \right) \right) \quad (12)$$

Each $f_p(z_p(\boldsymbol{u}))$, for $p = 1,2,...,P$, is represented by the basis function $b_{p,j}$ multiplied by the basis coefficient $\theta_{p,j}$ where j $1,2,...,K$ is the basis size which determines the maximum complexity of $f_p(z_p(\boldsymbol{u}))$, written in Equation (13).

$$f_p(z_p(\boldsymbol{u})) = \sum_{j=1}^{K} \theta_{p,j} b_{p,j}(z_p(\boldsymbol{u})) \quad , p = 1,2,...,P \quad (13)$$

A large basis size causes overfitting, so a smoothing penalty is used (a penalty matrix S that depends on the basis function) to prevent excess wiggliness. Penalties are added to the log-likelihood model and control smoothness via the smoothing parameter γ. The optimum smoothing parameter is selected by minimum Generalized Cross Validation (GCV) $v_g = \frac{(n+d)\|\boldsymbol{y} - \hat{\boldsymbol{\mu}}\|^2}{\left[(n+d) - tr(\boldsymbol{A})\right]^2}$, where $\boldsymbol{A} = \boldsymbol{b}(\mathbf{z}(\boldsymbol{u}))\left(\boldsymbol{b}^T(\mathbf{z}(\boldsymbol{u}))\boldsymbol{b}(\mathbf{z}(\boldsymbol{u})) + \boldsymbol{S}\right)^{-1} \boldsymbol{b}^T(\mathbf{z}(\boldsymbol{u}))$, $(n+d)$ is total number of quadrature points. Parameter θ_p is estimated by maximizing the penalized log-likelihood function in Equation (14) using penalized iteratively re-weighted least squares (PIRLS), given a link function $\boldsymbol{\eta} = g(\boldsymbol{\mu}) = \log(\lambda(\boldsymbol{u};\boldsymbol{\theta}))$ and an initial value $H^{[0]} = g'(\hat{\mu}_i^{[0]})(y_i - \hat{\mu}_i^{[0]}) + \hat{\eta}_i^{[0]}$, $W^{[0]} = diag\left(1 \Big/ \left(\mathrm{var}(y_i) g'(\hat{\mu}_i^{[0]})^2\right)\right)$, then $\hat{\boldsymbol{\theta}}$ becomes in Equation (15) (Wood 2017). The penalty measure of $f_p(z_p(\boldsymbol{u}))$ is shown in Equation (16).

$$l_p(\boldsymbol{\theta}) = \log L(\boldsymbol{u};\boldsymbol{\theta}) - \frac{1}{2} \sum_{p=1}^{P} \gamma_p \boldsymbol{\theta}_p^T \boldsymbol{S}_p \boldsymbol{\theta}_p \quad (14)$$

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \left\{ \left\| \sqrt{W} H - \sqrt{W} \boldsymbol{b}(\mathbf{z}(\boldsymbol{u}))\boldsymbol{\theta} \right\|^2 + \sum_{p=1}^{P} \gamma_p \boldsymbol{\theta}_p^T \boldsymbol{S}_p \boldsymbol{\theta}_p \right\} \quad (15)$$

$$\begin{aligned} \int \left[f_p''(z_p(\boldsymbol{u})) \right]^2 d(z_p(\boldsymbol{u})) &= \int \sum_{j=1}^{k} \theta_{p,j} b_{p,j}''(z_p(\boldsymbol{u})) d(z_p(\boldsymbol{u})) \\ &= \int \left(\boldsymbol{\theta}_p^T \boldsymbol{b}^T(\mathbf{z}_p(\boldsymbol{u})) \boldsymbol{b}(\mathbf{z}_p(\boldsymbol{u})) \boldsymbol{\theta}_p \right) d(\mathbf{z}_p(\boldsymbol{u})) \\ &= \boldsymbol{\theta}_p^T \boldsymbol{S}_p \boldsymbol{\theta}_p \\ \boldsymbol{S}_p &= \left[\int \boldsymbol{b}^T(\mathbf{z}_p(\boldsymbol{u})) \boldsymbol{b}(\mathbf{z}_p(\boldsymbol{u})) d(\mathbf{z}_p(\boldsymbol{u})) \right] \end{aligned} \quad (16)$$

For covariates such as distance, a cubic regression spline can be used which is one-dimensional smoothing. The cubic regression spline function can implicitly be written as $f(z(\boldsymbol{u})) = \sum_{j=1}^{K} \theta_j b_j(z(\boldsymbol{u})) = \sum_{j=1}^{k-1} \theta_j b_j(z(\boldsymbol{u}))$. Basis function $b(z(\boldsymbol{u}))$ with k *knots*, $z^1(\boldsymbol{u}),...,z^k(\boldsymbol{u})$, $\theta_j = b(z^j(\boldsymbol{u}))$, and $\delta_j = b''(z^j(\boldsymbol{u}))$ can be expressed in Equation (17).

$$\begin{aligned} b(z(\boldsymbol{u})) = {} & a_j^- z(\boldsymbol{u}) \theta_j + a_j^+ z(\boldsymbol{u}) \theta_{j+1} + c_j^- z(\boldsymbol{u}) \delta_j + \\ & c_j^+ z(\boldsymbol{u}) \delta_{j+1} \;,\; z^j(\boldsymbol{u}) \le z(\boldsymbol{u}) \le z^{j+1}(\boldsymbol{u}) \end{aligned} \quad (17)$$

$z^k(\boldsymbol{u})$ is the location of the knot point. a_j^-, a_j^+, c_j^- and c_j^+ are defined in Equation (18).

$$a_j^- = \left(z^{j+1}(\boldsymbol{u}) - z(\boldsymbol{u})\right)\Big/ h_j;\ a_j^+ = \left(z(\boldsymbol{u}) - z^j(\boldsymbol{u})\right)\Big/ h_j$$
$$c_j^- = \left(\left(z^{j+1}(\boldsymbol{u}) - z(\boldsymbol{u})\right)^3 \Big/ h_j - h_j\left(z^{j+1}(\boldsymbol{u}) - z(\boldsymbol{u})\right)\right)\Big/ 6; \quad (18)$$
$$h_j = z^{j+1}(\boldsymbol{u}) - z^j(\boldsymbol{u})$$
$$c_j^+ = \left(\left(z(\boldsymbol{u}) - z^j(\boldsymbol{u})\right)^3 \Big/ h_j - h_j\left(z(\boldsymbol{u}) - z^j(\boldsymbol{u})\right)\right)\Big/ 6$$

The penalty matrix for the cubic spline regression basis is $\boldsymbol{S} = \boldsymbol{D}^T \boldsymbol{B}^- \boldsymbol{D}$, $\boldsymbol{D}$ is the upper tridiagonal matrix and $\boldsymbol{B}$ is the symmetric tridiagonal matrix (Wood 2017).

$$D_{j,j} = 1/h_j \ ;\ j = 1,...,k-2$$
$$D_{j,j+1} = -1/h_j - 1/h_{j+1} \ ;\ j = 1,...,k-2 \quad (19)$$
$$D_{j,j+2} = 1/h_{j+1} \ ;\ j = 1,...,k-2$$

$$B_{j,j} = \left(h_j + h_{j+1}\right)/3 \ ;\ j = 1,...,k-2$$
$$B_{j,j+1} = B_{j+1,j} = h_{j+1}/6 \quad ;j = 1,...,k-3$$

The cubic spline regression function can be written in matrix form, i.e., $\boldsymbol{f} = \boldsymbol{H}\theta$ with, $\boldsymbol{f} = \left(f\left(z_1(\boldsymbol{u})\right),...,f\left(z_{n+d}(\boldsymbol{u})\right)\right)^T$, $\boldsymbol{\theta} = \left(\theta_1,...,\theta_{n+d}\right)^T$, $H_{i,j} = b_j\left(z_i(\boldsymbol{u})\right)$, $j = 1,...,k-1$, $i = 1,...,$ $n+d$, is the total number of quadrature points.
The estimated parameter $\boldsymbol{\theta}$ in $\boldsymbol{f}$ is obtained by minimizing Equation (20).

$$\left(\boldsymbol{y} - \boldsymbol{H\theta}\right)^T \left(\boldsymbol{y} - \boldsymbol{H\theta}\right) + \gamma \boldsymbol{\theta}^T \boldsymbol{S\theta} \quad (20)$$

The condition for Equation (20) to be minimum is that the first derivative of Equation (20) is equal to zero, so the estimated value for the $\hat{\boldsymbol{\theta}}$ parameter is shown in Equation (21).

$$\hat{\boldsymbol{\theta}} = \left(\boldsymbol{H}^T\boldsymbol{H} + \hat{\gamma}\boldsymbol{S}\right)^{-1}\boldsymbol{H}^T\boldsymbol{y} \quad (21)$$

$$\hat{f}\left(z(\boldsymbol{u})\right) = \boldsymbol{H}\left(\boldsymbol{H}^T\boldsymbol{H} + \hat{\gamma}\boldsymbol{S}\right)^{-1}\boldsymbol{H}^T\boldsymbol{y} \quad (22)$$

In **mgcv**, model terms like *s(x,bs="cr",k=15)* use this basis (*k* defaults to 10 if not supplied).

SECOND-ORDER COMPOSITE LIKELIHOOD

Second-order composite likelihood is a parameter estimation method to get a cluster parameter estimator κ and ω. The second-order composite likelihood function is constructed from all pairs of data points $\boldsymbol{u}$ and $\boldsymbol{v}$ shown in Equation (23), where $\boldsymbol{u} \in \mathbf{X}$, $\boldsymbol{v} \in \mathbf{X}$, and $\boldsymbol{u} \neq \boldsymbol{v}$.

$$\log CL(\boldsymbol{\psi}) = \sum_{\boldsymbol{u}\in\mathbf{X}}\sum_{\substack{\boldsymbol{v}\in\mathbf{X}\\ \boldsymbol{u}\neq\boldsymbol{v}}} w(\boldsymbol{u},\boldsymbol{v})$$
$$\left(\log\lambda^{(2)}(\boldsymbol{u},\boldsymbol{v};\boldsymbol{\psi}) - \log\int_B\int_B w(\boldsymbol{u},\boldsymbol{v})\lambda^{(2)}(\boldsymbol{u},\boldsymbol{v};\boldsymbol{\psi})\,d\boldsymbol{u}d\boldsymbol{v}\right)$$
$$w(\boldsymbol{u},\boldsymbol{v}) = \mathbf{1}\left\{\|\boldsymbol{u}-\boldsymbol{v}\| \le R\right\} \quad (23)$$

w is a function of weight, $R > 0$ is the upper limit of the correlation distance from the model. The result of the derivation of Equation (23) concerning $\boldsymbol{\psi}$ is as follows.

$$\frac{\partial}{\partial\boldsymbol{\psi}}\log CL(\boldsymbol{\psi}) = \sum_{\boldsymbol{u}\in\mathbf{X}}\sum_{\substack{\boldsymbol{v}\in\mathbf{X}\\ \boldsymbol{u}\neq\boldsymbol{v}}} w(\boldsymbol{u},\boldsymbol{v})\frac{\kappa_2(\boldsymbol{u},\boldsymbol{v};\boldsymbol{\psi})}{\lambda^{(2)}(\boldsymbol{u},\boldsymbol{v};\boldsymbol{\psi})} - \sum_{\boldsymbol{u}\in\mathbf{X}}\sum_{\substack{\boldsymbol{v}\in\mathbf{X}\\ \boldsymbol{u}\neq\boldsymbol{v}}} w(\boldsymbol{u},\boldsymbol{v})\frac{\kappa_2(w)}{\lambda^{(2)}(w)} \quad (24)$$
$$\kappa_2(\boldsymbol{u},\boldsymbol{v};\boldsymbol{\psi}) = \frac{\partial}{\partial\boldsymbol{\psi}}\lambda^{(2)}(\boldsymbol{u},\boldsymbol{v};\boldsymbol{\psi})$$

MODEL ASSESSMENT

The best model criteria in this study used the envelope test and AIC (Akaike Information Criterion). The envelope is the critical limit of the K-function statistical test which validates the suitability of the point pattern data to the point process model. The K-function plot of the original data points will be compared with the results of the Cauchy processes model simulation. A model is said to be good if the original K-function data point plot is in the K-function envelope interval. In addition, the best model is selected based on the smallest AIC value using the AIC formula in Equation (25).

$$AIC = -2L_{\max} + 2q \quad (25)$$

$L_{\max}$ is the maximum value of Equation (23) and q is the sum of the effective degrees of freedom (EDF) and the number of cluster parameters. The total effective degrees of freedom of the smooth is $tr\left(\left(\boldsymbol{b}^T(\mathbf{z}(\boldsymbol{u}))W\boldsymbol{b}(\mathbf{z}(\boldsymbol{u})) + \boldsymbol{S}_\gamma\right)^{-1}\boldsymbol{b}^T(\mathbf{z}(\boldsymbol{u}))W\boldsymbol{b}(\mathbf{z}(\boldsymbol{u}))\right)$, where W is the final weights used in IRLS iteration; $\boldsymbol{b}(\mathbf{z}(\boldsymbol{u}))$ is the model matrix for the smooth term; and $\boldsymbol{S}_\gamma$ is the block diagonal matrix of $\gamma_p\boldsymbol{S}_p$.

RESULTS AND DISCUSSION

In this study, we detect spatial trends and clustering effects of earthquake distribution in Sumatra. Modeling and parameter estimation are done computationally using the **spatstat** and **mgcv** packages on R. Based on

the results of the chi-square test and the inhomogeneous K-function plot in (Choiruddin, Susanto & Metrikasari 2021)'s study, the distribution of earthquakes in Sumatra is inhomogeneous and clustered when the radius is less than 200 km. Thus, the Cauchy cluster point process model was chosen to model the earthquake distribution, because it performed best. Parameter estimation consists of cluster parameters and geological factor parameters, which are estimated using a two-step estimation. Research by Choiruddin, Susanto and Metrikasari (2021) uses a log-linear intensity model to estimate the parameters of geological variables.

This study aims to compare the performance of the Cauchy cluster process model with log-linear intensity and log-additive intensity. We model earthquake intensity $\lambda(\boldsymbol{u};\boldsymbol{\theta})$, which is the propensity for earthquakes to occur at a certain distance from subduction and fault. First, we use a log-linear intensity model, and the results are presented in Table 2. The distance of an earthquake to the nearest subduction and fault has a significant effect on the intensity of earthquakes in Sumatra with an inverse effect on the risk of an earthquake. The effect of the subduction zone on the intensity of the earthquake is greater than the fault effect. For every additional 100 km of the distance from a location to the nearest subduction zone, the risk of an earthquake at that location decreases exp(-0.95)=0.39 or it can be said that the closer a location is to the subduction zone, the risk of an earthquake at that location increases by 1/exp(-0.95)=2.56.
The earthquake intensity model using the log-linear model is as follows.

$$\hat{\lambda}\left(\boldsymbol{u};\hat{\boldsymbol{\theta}}\right)=\exp\left(4.79-0.64\left(z_1\left(\boldsymbol{u}\right)\right)-0.95\left(z_2\left(\boldsymbol{u}\right)\right)\right)$$

Using the values of $\hat{\kappa}$ and $\hat{\omega}$ in Table 2, $|B|=237.33$ $(100\text{km})^2$ the estimated number of mainshocks is 126 with a standard deviation of aftershocks around the mainshocks of 13 km.

Then, we use the log-additive intensity model due to the non-parametric estimation of earthquake intensity in Figure 3 being far from the log-linear model. The log-additive model can analyze complex non-linear relationships flexibly through smooth functions. We assume that the log-additive intensity model is as follows, with $f_p\left(z_p\left(\boldsymbol{u}\right)\right)$ is represent the basis function i.e., the cubic regression spline, for each covariate $z_p\left(\boldsymbol{u}\right)$ $p=1,2$.

$$\lambda(\boldsymbol{u};\boldsymbol{\theta})=\exp\left(\theta_0+f_1\left(z_1\left(\boldsymbol{u}\right)\right)+f_2\left(z_2\left(\boldsymbol{u}\right)\right)\right)$$

The optimum number of knots and smoothing parameters are selected based on the smallest AIC value. The Cauchy cluster process model with intensity using GAM which produces the smallest AIC value (considering the modeling results using the number of knots by default from the R, which is ten and considering the location of knots based on peaks in nonparametric estimates) is a model using three knots for each geological factor. Using the GCV method (default in **mgcv**), the optimum smoothing parameters estimation is $\gamma=0.08677$ for subduction and $\gamma=0.00118$ for fault. Then, by considering the estimation results of smoothing parameters using GCV, we try to use another smoothing parameter value i.e., $\gamma=0,\ 0.001,\ 0.01,\ 0.1,\ 1$. We choose the smoothing parameter which produces the smallest AIC value in the Cauchy cluster process model. The best model is to use the smoothing parameters, $\gamma=0$ for subduction and $\gamma=0.1$ for fault.
The locations of the knot points for the distance of the earthquake to the fault are:

$$z^1\left(\boldsymbol{u}\right)=0.00329;\ z^2\left(\boldsymbol{u}\right)=2.68399;\ \ z^3\left(\boldsymbol{u}\right)=12.20169.$$

Details of the location knot points for the subduction distance covariates are:

$$z^1\left(\boldsymbol{u}\right)=0.00031;\ \ z^2\left(\boldsymbol{u}\right)=3.16416;\ \ z^3\left(\boldsymbol{u}\right)=15.13639.$$

The results of modeling the distribution of earthquakes in Sumatra using the Cauchy cluster process with GAM on the intensity function are shown in Table 3. Based on the values of $\hat{\kappa}$, $\hat{\omega}$, and the area of the observation window (237.33 $(100\text{ km})^2$), the estimated number of mainshocks is 114 with a standard deviation of aftershocks around the mainshocks of 14 km. The earthquake intensity function model formed is as follows,

$$\hat{\lambda}\left(\boldsymbol{u};\hat{\boldsymbol{\theta}}\right)=\exp\left(-4.47-0.88b_{1,1}\left(z_1\left(\boldsymbol{u}\right)\right)-28.87b_{1,2}\left(z_1\left(\boldsymbol{u}\right)\right)-2.21b_{2,1}\left(z_2\left(\boldsymbol{u}\right)\right)-17.54b_{2,2}\left(z_2\left(\boldsymbol{u}\right)\right)\right)$$

The basis function on the covariate of the earthquake distance to the nearest fault which refers to Equation (17) is as follows.

$$\hat{f}_1\left(z_1\left(\boldsymbol{u}\right)\right)=-0.88b_{1,1}\left(z_1\left(\boldsymbol{u}\right)\right)-28.87b_{1,2}\left(z_1\left(\boldsymbol{u}\right)\right);$$

for $0.00329\le z_1\left(u\right)\le 2.68399$;

$$b_{1,1}\left(z_1\left(\boldsymbol{u}\right)\right)=\frac{\left(2.68399-z_1\left(\boldsymbol{u}\right)\right)}{2.68071}f\left(0.00329\right)+\frac{\left(z_1\left(\boldsymbol{u}\right)-0.00329\right)}{2.68071}f\left(2.68399\right)+$$

$$\frac{\left[\frac{\left(2.68399-z_1(\boldsymbol{u})\right)^3}{2.68071}-(2.68071)\left(2.68399-z_1(\boldsymbol{u})\right)\right]}{6}f''(0.00329)+$$
$$\frac{\left[\frac{\left(z_1(\boldsymbol{u})-0.00329\right)^3}{(2.68071)}-(2.68071)\left(z_1(\boldsymbol{u})-0.00329\right)\right]}{6}f''(2.68399)$$

$f(0.00329)=-0.9911650$; $f(2.68399)=0.080639775$;
$f''(0.00329)=0$; $f''(2.68399)=-0.1031343.$

for $2.68399\le z_1(u)\le 12.20169$;

$$b_{1,2}\left(z_1(\boldsymbol{u})\right)=\frac{\left(12.20169-z_1(\boldsymbol{u})\right)}{9.51769}f(2.68399)+$$
$$\frac{\left(z_1(\boldsymbol{u})-2.68399\right)}{9.51769}f(12.20169)+$$
$$\frac{\left[\frac{\left(12.20169-z_1(\boldsymbol{u})\right)^3}{9.51769}-(9.51769)\left(12.20169-z_1(\boldsymbol{u})\right)\right]}{6}f''(2.68399)+$$
$$\frac{\left[\frac{\left(z_1(\boldsymbol{u})-2.68399\right)^3}{(9.51769)}-(9.51769)\left(z_1(\boldsymbol{u})-2.68399\right)\right]}{6}f''(12.20169)$$

$f(2.68399)=-0.1053056$; $f(12.20169)=0.9879381$
$f''(2.68399)=0.0274945$; $f''(12.20169)=0$

While the basis function for the subduction covariate is as follows.

$$\hat{f}_2\left(z_2(\boldsymbol{u})\right)=-2.21b_{2,1}\left(z_2(\boldsymbol{u})\right)-17.54b_{2,2}\left(z_2(\boldsymbol{u})\right);$$
for $0.00031\le z_2(u)\le 3.16416$;

$$b_{2,1}\left(z_2(\boldsymbol{u})\right)=\frac{\left(3.16416-z_2(\boldsymbol{u})\right)}{3.16385}f(0.00031)+$$
$$\frac{\left(z_2(\boldsymbol{u})-0.00031\right)}{3.16385}f(3.16416)+$$
$$\frac{\left[\frac{\left(3.16416-z_2(\boldsymbol{u})\right)^3}{3.16385}-(3.16385)\left(3.16416-z_2(\boldsymbol{u})\right)\right]}{6}f''(0.00031)+$$
$$\frac{\left[\frac{\left(z_2(\boldsymbol{u})-0.00031\right)^3}{(3.16385)}-(3.16385)\left(z_2(\boldsymbol{u})-0.00031\right)\right]}{6}f''(3.16416)$$

$f(0.00031)=-0.985831$; $f(3.16416)=0.1175359$
$f''(0.00031)=0$; $f''(3.16416)=-0.07304853.$

for $3.16416\le z_2(\boldsymbol{u})\le 15.13639$

$$b_{2,2}\left(z_2(\boldsymbol{u})\right)=\frac{\left(15.13639-z_2(\boldsymbol{u})\right)}{11.97223}f(3.16416)+$$
$$\frac{\left(z_2(\boldsymbol{u})-3.16416\right)}{11.97223}f(15.13639)+$$
$$\frac{\left[\frac{\left(15.13639-z_2(\boldsymbol{u})\right)^3}{11.97223}-(11.97223)\left(15.13639-z_2(\boldsymbol{u})\right)\right]}{6}f''(3.16416)+$$
$$\frac{\left[\frac{\left(z_2(\boldsymbol{u})-3.16416\right)^3}{(11.97223)}-(11.97223)\left(z_2(\boldsymbol{u})-3.16416\right)\right]}{6}f''(15.13639)$$

$f(15.13639)=0.9837697$; $f(3.16416)=-0.1196773$
$f''(15.13639)=0$; $f''(3.16416)=0.01738952.$

The goodness of the Cauchy cluster process model in modeling earthquakes in Sumatra can be seen from the AIC value and the K-function envelope plot. The AIC value of the Cauchy cluster model is presented in Table 4. The AIC value in the Cauchy cluster model with the log-additive intensity model is smaller than the log-linear model. Based on the K-function envelope (Figure 4), the Cauchy cluster process with the intensity function using GAM is the best model because the black line in the GAM model is smaller that comes out of the grey area (Figure 4(b)) than the log-linear model. The black line exits the envelope interval around the values of $60 < r < 180$ km and $r > 320$ km. Thus, the best model for modelling the distribution of earthquakes in Sumatra is the Cauchy cluster process with a log-additive intensity model.

The results of the prediction of earthquake intensity in Sumatra using the Cauchy cluster process with covariates of subduction and fault are shown in Figure 5, in which the brightest colour is an area with very high earthquake intensity. Earthquakes in Sumatra tend to be at high risk in areas which are flanked by subduction zones and faults, namely in the west of the island of Sumatra which borders the Indian Ocean and is predicted to be highest in the northern region of Sumatra. Aceh has the highest risk of earthquakes compared to other areas in Sumatra because the upper area of Sumatra is close to subduction zones and faults such as Weh, Jaboe, Seulawah Agam, Peuet Sagoe, and Geureudong. Intensity prediction

using a log-linear model only captures high intensity in the North Sumatra region where subduction and fault confluence, while the intensity in Nias and Mentawai areas is small. The predicted results with intensity using GAM more clearly capture the effect of subduction and fault on earthquake intensity as seen in Figure 5(b) the Simeule, Nias, and Mentawai areas are brighter than in Figure 5(a).

TABLE 2. Parameter estimation in Cauchy cluster process model with log-linear intensity

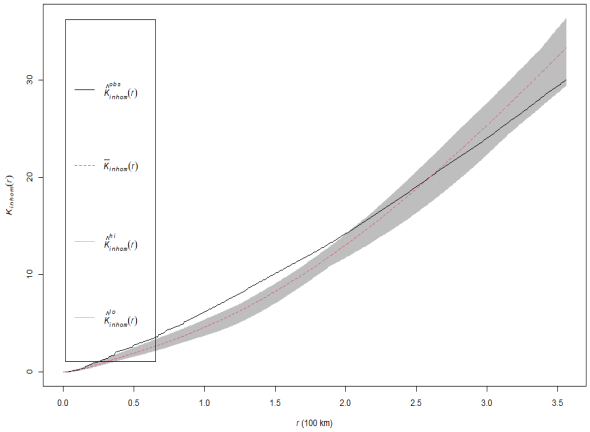
Parameter	Estimate	S.E.	Z-value
$\hat{\kappa}$	0.53		
$\hat{\omega}$	0.13		
Intercept $(\hat{\theta}_0)$	4.79	0.48	9.98
Earthquake distance to fault $(\hat{\theta}_1)$	-0.64	0.16	-3.95
Earthquake distance to subduction $(\hat{\theta}_2)$	-0.95	0.16	-5.87

TABLE 3. Parameter estimation in Cauchy cluster process model (log-additive intensity)

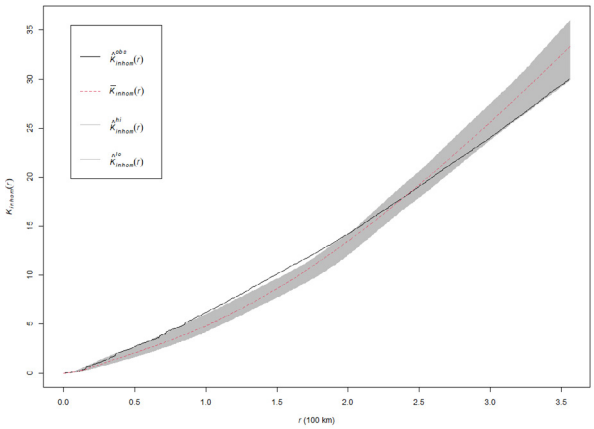
Parameter		Estimate	S.E.	*Z-value*
$\hat{\kappa}$		0.48		
$\hat{\omega}$		0.14		
Intercept	$(\hat{\theta}_0)$	-4.47	2.82	-1.59
Earthquake distance to nearest fault	$(\hat{\theta}_{1,1})$	-0.88	0.82	-1.08
	$(\hat{\theta}_{1,2})$	-28.87	13.79	-2.09
Earthquake distance to nearest subduction	$(\hat{\theta}_{2,1})$	-2.21	0.84	-2.63
	$(\hat{\theta}_{2,2})$	-17.54	17.73	-0.99

TABLE 4. Model assessment for Cauchy cluster process

Intensity function model	L_{max}	AIC
Log-linear	947756	-1895502
Log-additive	955217.7	-1910420

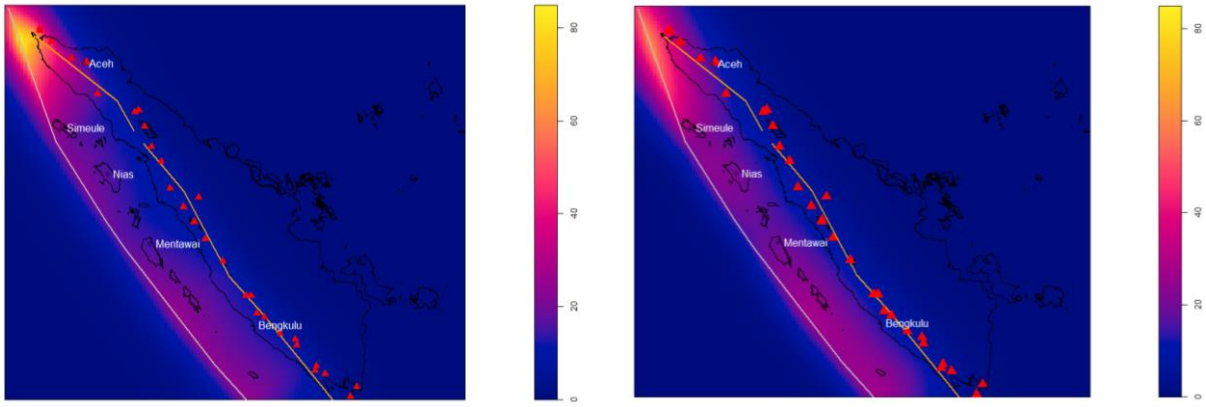


(a)



(b)

FIGURE 4. Envelopes inhomogeneous K-function for Cauchy cluster process with (a) Log-linear intensity model, (b) log-additive intensity model



(a) (b)

FIGURE 5. Earthquake intensity prediction map in Sumatra using Cauchy cluster model with (a) log-linear intensity, (b) log-additive intensity (GAMs approach)

CONCLUSIONS

Visually, nonparametric estimation of earthquake intensity shows that the effect of the geological variables is far from log-linear. As an alternative to avoid the linearity assumption, we use the log-additive model in modeling the effects of subduction and fault on earthquake intensity. Cauchy cluster process model with the log-additive intensity model gives better performance in terms of AIC value and K-function envelopes. The estimated number of mainshocks due to the clustering effect is 114 with aftershocks spreading 14 km around the mainshocks. Areas in Sumatra which are predicted to have a high risk of earthquakes are the areas in the upper part of Sumatra (the provinces of Aceh and North Sumatra), and the western part of Sumatra which borders the Indian Ocean (Mentawai, Nias, and Simeulue) and Bengkulu. Suggestions for further research are to improve the model that considers parametric or nonparametric terms, the appropriate basis function, and interactions between geological variables using tensor product smooths. One would also consider the extension of Epidemic Type Aftershock Sequence (ETAS) which takes into account the effect of geological variables through the intensity with a log-additive form.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge financial support from the Institut Teknologi Sepuluh Nopember for this work, under the project scheme of the Publication Writing and IPR Incentive Program (PPHKI) 2022. We thank the editor and reviewers for their constructive and helpful comments.

REFERENCES

Anwar, S., Yaseen, M. & Mahmood, S.A. 2023. Higher order Gibbs point process modeling of 2005-Kashmir earthquakes. *Modeling Earth Systems and Environment* 9: 1335-1347.

Baddeley, A., Rubak, E. & Turner, R. 2015. *Spatial Point Patterns: Methodology and Applications with R*. Boca Raton: CRC Press.

BMKG. 2019. *Katalog Gempabumi Signifikan dan Merusak 1821-2018*. T. Prasetya & Daryono. Pusat Gempabumi dan Tsunami Kedeputian Bidang Geofisika Badan Meteorologi Klimatologi dan Geofisika: Jakarta.

Choiruddin, A., Aisah, Trisnisa, F. & Iriawan, N. 2021. Quantifying the effect of geological factors on distribution of earthquake occurrences by inhomogeneous Cox processes. *Pure and Applied Geophysics* 178(5): 1579-1592.

Choiruddin, A., Susanto, T.Y., Husain, A. & Kartikasari, Y.M. 2023. kppmenet: Combining the kppm and elastic net regularization for inhomogeneous Cox point process with correlated covariates. Under revision.

Choiruddin, A., Susanto, T.Y. & Metrikasari, R. 2021. Two-step estimation for modeling the earthquake occurrences in Sumatra by Neyman-Scott Cox point processes. In *Communications in Computer and Information Science,* edited by Mohamed, A., Yap, B.W., Zain, J.M., Berry, M.W. Singapore: Springer. 1489: 146-159. https://doi.org/10.1007/978-981-16-7334-4_11

D'Angelo, N., Siino, M., D'Alessandro, A. & Adelfio, G. 2022. Local spatial log-Gaussian Cox processes for seismic data. *AStA Advances in Statistical Analysis* 106: 633-671.

Geng, J., Shi, W. & Hu, G. 2021. Bayesian nonparametric nonhomogeneous Poisson process with applications to USGS earthquake data. *Spatial Statistics* 41: 100495.

Ghorbani, M. 2013. Cauchy cluster process. *Metrika* 76(5): 697-706.

Husain, A. & Choiruddin, A. 2021. Poisson and Logistic regressions for inhomogeneous multivariate point processes: A case study in the Barro Colorado Island plot. In *Soft Computing in Data Science. SCDS 2021. Communications in Computer and Information Science*, vol 1489, edited by Mohamed, A., Yap, B.W., Zain, J.M. & Berry, M.W. Singapore: Springer. https://doi.org/10.1007/978-981-16-7334-4_22

Ibrahim. 2019. *BMKG Soft Launching Uji Coba Sistem Peringatan Dini Gempa*. https://www.bmkg.go.id/press-release/?p=bmkg-soft-launching-uji-coba-sistem-peringatan-dini-gempa&tag=press-release&lang=ID.

Iftimi, A., Cronie, O. & Montes, F. 2019. Second-order analysis of marked inhomogeneous spatiotemporal point processes: Applications to earthquake data. *Scandinavian Journal of Statistics* 46(3): 661-685.

Jalilian, A. 2017. Modelling and classification of species abundance: A case study in the Barro Colorado Island plot. *Journal of Applied Statistics* 44(13): 2401-2409.

Møller, J. & Waagepetersen, R.P. 2004. *Statistical Inference and Simulation for Spatial Point Processes*. Boca Raton: CRC Press.

Nasional, P.S.G. 2017. *Peta Sumber dan Bahaya Gempa Indonesia tahun 2017*. Pusat Litbang Perumahan dan Pemukiman, Kementerian Pekerjaan Umum dan Perumahan Rakyat (National Center for Earthquake Studies. Indonesian Seismic Sources and Seismic Hazard Maps 2017. Center for Research.

Ogata, Y. 1988. Statistical models for earthquake occurrences and residual analysis for point processes. *Journal of the American Statistical Association* 83(401): 9-27.

Ogata, Y. 2006. Monitoring of anomaly in the aftershock sequence of the 2005 earthquake of M7. 0 off coast of the western Fukuoka, Japan, by the ETAS model. *Geophysical Research Letters* 33(1).

Pedersen, E.J., Miller, D.L., Simpson, G.L. & Ross, N. 2018. Hierarchical generalized additive models: An introduction with mgcv. *PeerJ* 7:e6876.

Pratiwi, H., Haryanto, W., Subanti, S., Mangku, I.W. & Ferawati, K. 2021. A self-exciting point process with cyclic component, trend component, triggering function, and response function. *AIP Conference Proceedings*. 2329: 060033.

Pratiwi, H., Rini, L.S. & Mangku, I.W. 2018. Marked point process for modelling seismic activity (case study in Sumatra and Java). *Journal of Physics: Conference Series* 1022(1): 12004.

Siino, M., Adelfio, G., Mateu, J., Chiodi, M. & D'Alessandro, A. 2017. Spatial pattern analysis using hybrid models: An application to the Hellenic seismicity. *Stochastic Environmental Research and Risk Assessment* 31(7): 1633-1648.

Sosilawati, Handayani, A., Wahyudi, A.R., Mahendra, Z.A., Massudi, W., Febrianto, S. & Suhendri, N.A. 2017. *Sinkronisasi Program dan Pembiayaan Pembangunan Jangka Pendek 2018-2020. Keterpaduan Pengembangan Kawasan dengan Infrastruktur PUPR Pulau Jawa, Volume 1.* Pusat Pemrograman dan Evaluasi Keterpaduan Infrastruktur Pupr, Badan Pengembangan Infrastruktur Wilayah, Kementerian Pekerjaan Umum dan Perumahan Rakyat.

Trisnisa, F., Metrikasari, R., Rabbanie, R., Sakdiyah, K. & Choiruddin, A. 2019. Model inhomogeneous spatial Cox processes untuk pemetaan risiko gempa bumi di pulau Jawa. *Inferensi* 2(2): 107-111.

Triyono, R. 2015. Ancaman gempabumi di Sumatera tidak hanya bersumber dari Mentawai megathrust. *Geofisika Klas I Padang Panjang*.

Türkyilmaz, K., van Lieshout, M.N.M. & Stein, A. 2013. Comparing the Hawkes and trigger process models for aftershock sequences following the 2005 Kashmir earthquake. *Mathematical Geosciences* 45(2): 149-164.

U.S. Geological Survey (USGS). Search Earthquake Catalog. https://earthquake.usgs.gov/earthquakes/search/.

Wood, S.N. 2017. *Generalized Additive Models: An Introduction with R.* (2nd ed). Boca Raton: CRC Press.

Youngman, B.D. & Economou, T. 2017. Generalised additive point process models for natural hazard occurrence. *Environmetrics* 28(4): e2444.

*Corresponding author; email: choiruddin@its.ac.id