Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Kernel Density Estimation: Examining the Effect of Kernel Function, Bandwidth, and Grid Cell Size in Mapping Earthquake-Prone in Indonesia

Ari Fadli

arifadli@mail.ugm.ac.id

Universitas Gadjah Mada https://orcid.org/0000-0002-9551-2580

Tri Kuntoro Priyambodo

Universitas Gadjah Mada Fakultas Matematika dan Ilmu Pengetahuan Alam

Agfianto Eko Putra

Universitas Gadjah Mada Fakultas Matematika dan Ilmu Pengetahuan Alam

Wiwit Suryanto

Universitas Gadjah Mada Fakultas Matematika dan Ilmu Pengetahuan Alam https://orcid.org/0000-0002-6275-7880

Research Article

Keywords: Kernel Density Estimation, Earthquake Mapping, Earthquake Prediction, Hotspot Analysis

Posted Date: March 31st, 2026

DOI: https://doi.org/10.21203/rs.3.rs-5264767/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Kernel Density Estimation: Examining the Effect of Kernel Function, Bandwidth, and Grid Cell Size in Mapping Earthquake-Prone in Indonesia

Ari Fadli[1,3†], Tri Kuntoro Priyambodo[1†], Agfianto Eko Putra[1†], Wiwit Suryanto[2*†]

[1]Department of Computer Science and Electronics, Universitas Gadjah Mada, Sekip Utara, Bulaksumur, Yogyakarta, 55281, Sleman, Indonesia.
[2]Department of Physics, Universitas Gadjah Mada, Sekip Utara, Bulaksumur, Yogyakarta, 55281, Sleman, Indoensia.
[3]Electrical Engineering, Universitas Jenderal Soedirman, Jl. Raya Mayjen Sungkono No.KM 5, Purbalingga, 53371, Central Java, Indonesia.

*Corresponding author(s). E-mail(s): ws@ugm.ac.id;
Contributing authors: arifadli@mail.ugm.ac.id; mastri@ugm.ac.id; agfi@ugm.ac.id;
†These authors contributed equally to this work.

Abstract

Identifying earthquake-prone areas is critical for disaster mitigation to reduce casualties and economic losses. This study applies Kernel Density Estimation (KDE) to analyze and rank earthquake-prone regions, focusing on variations in kernel functions, bandwidths, and grid sizes. The Indonesian Earthquake Catalog (1964–2023) is used as a case study. The results indicate that different kernel functions have unique strengths. The Epanechnikov kernel provides an even density distribution, particularly in low and medium categories, while the Gaussian kernel captures high concentrations effectively, especially in high and extreme categories. The Biweight kernel performs well in medium and high categories but less effectively identifies extreme density concentrations. Grid size also significantly impacts results; smaller grids (0.25 x 0.25) reveal detailed density patterns but may overemphasize localized concentrations, whereas larger grids (5 x 5) are suited for macro-scale analyses but can obscure finer variations. Bandwidth selection significantly affects density estimates. Smaller bandwidths (0.1) spread density widely, resulting in many grids in the low category but few in the medium, high, and extreme categories. Medium bandwidths (0.3) increase the proportion

of medium and high categories, while larger bandwidths (0.5) produce the highest proportion of grids in the medium category, though extreme values remain limited. These variations demonstrate how bandwidth choices influence the balance between localized detail and broader distribution patterns. KDE effectively identifies earthquake-prone areas with varying densities and cluster significance, providing essential insights to prioritize disaster mitigation efforts and improve spatial planning in earthquake-prone regions. The combination of Gaussian kernel, 0.5 bandwidth, and 1x1 degree grid yields the greatest results in completely mapping earthquake risk

Keywords: Kernel Density Estimation, Earthquake Mapping, Earthquake Prediction, Hotspot Analysis

1 Introduction

Indonesia is extremely vulnerable to many natural and man-made disasters, including landslides, earthquakes, volcanoes, tsunamis, droughts, and forest fires Achmad (2023). With 11,660 earthquake vibrations recorded in the Meteorology, Climatology, and Geophysics Agency (BMKG) database in 2019, Indonesia is classified as earthquake-prone due to its location at the meeting point of three major tectonic plates: the Indo-Australian, Eurasian, and Pacific plates Kurniati et al (2021)

Indonesia has experienced earthquakes for a very long time. According to the distribution map of global earthquake sources, Indonesia is situated in an area prone to earthquakes. Since 2013, there has also been a tendency for the quantity and intensity of earthquakes in Indonesia to rise. There was a notable increase in 2018, with 11,920 earthquakes (2,763 in Lombok and 1,438 in Palu). On the other hand, 11,588 earthquakes of varying magnitude occurred in 2019. Before 2020, Indonesia experienced between 4,000 and 7,000 earthquakes annually. From 2008 to 2020, there were 163 destructive earthquakes and 8,258 incidents overall. Earthquakes were felt 757 times, with a seismic scale of >5 SR occurring 242 times, and 11 were destructive earthquakes Achmad (2023)

Various analytical techniques that can be used to characterize disaster-prone areas have developed from the existing scientific literature. The approach is generally used, namely methods that rely on analyzing historical data events. Although no single technique has emerged as "the best" in identifying accident-prone areas, recent research shows that Kernel Density Estimation (KDE) outperforms other approaches Srikanth and Srikanth (2020); Chainey et al (2008); Baranyai and Sipos (2022).

The KDE method has been used in several research domains to identify accident-prone locations between wildlife and vehicles Bíl et al (2016). Furthermore, this method has also been used to identify accident-prone locations experienced by pedestrians Pulugurtha et al (2007); Blazquez and Celis (2013) and its use has reached the stage of spatial and temporal analysis of accidents that occur on highways Achu et al (2019); Anderson (2009); Hashimoto et al (2016). Meanwhile, researchers have used several methods to identify areas with high earthquake risk in identifying areas prone to earthquake disasters. The method used in identifying areas with high seismic risk

was first introduced by mathematician I.M. Gelfand et al. in 1972. The approach was developed later and is called EPA (Area Prowan Gempa) Gelfand et al (1972); Soloviev et al (2014); Gvishiani et al (2020). The KDE method is also used to obtain spatial features of fault density, which are used for earthquake prediction based on spatiotemporal analysis Yousefzadeh et al (2021); Sharma et al (2022).

Furthermore, the clustering method with K-Medoids and K-Means is used to identify groups of areas with the same level of earthquake risk Rifa et al (2020); Senduk et al (2020). Meanwhile, other researchers conducted research using the K-means algorithm and DBSCAN algorithm. This study aims to classify the distribution of earthquakes in Indonesia in 2019 based on the values of magnitude, depth, and position Jufriansah et al (2021). Earthquake Analysis in Indonesia Using the Clustering Method aims for areas with high earthquake-prone areas in Indonesia Prasetio et al (2023).

The method in this study emphasizes data collection in the form of rock types, land slopes, seismic data in the form of soil acceleration in the bedrock, and areas located in Identification of earthquake-prone areas Barlian et al (2022). Conducting Spatial-Temporal Analysis and Identification of Earthquake-Prone Areas Using ST-DBCSAN and K-Means Clustering Samsudin et al (2024).

Based on several previous studies and looking at the potential of KDE's ability to provide good results in identifying areas prone to traffic accidents, the use of KDE has never been proposed in identifying areas according to the level of earthquake risk. We propose a methodology for identifying areas according to the level of earthquake risk using the KDE approach. This study has several parts: Section 2 explains our proposed method and how we evaluate it; Section 3 shows our findings; Section 4 shows our discussion; and Section 4 concludes our study and the future scope for the next study.

2 Research Method

2.1 Proposed Methodology

The research stages used can be seen in Figure 1. Earthquake incident data from 1964 to 2023 was collected from publicly available data sources, namely the United States Geological Survey (USGS). The boundaries of Indonesia used in this study are made in GeoJSON format. These data are used together to map the distribution of earthquake events, as shown in Figure 2.

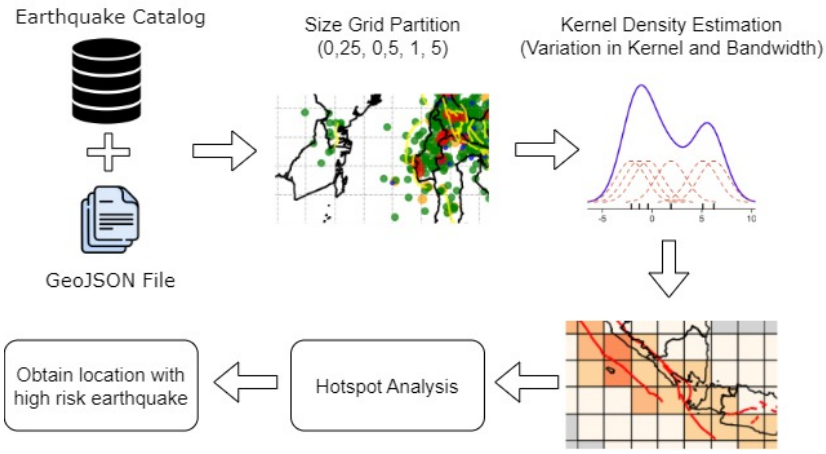


Fig. 1 Research Stages

Furthermore, the density map of earthquake-prone areas is obtained by applying Kernel Density Estimation (KDE). In this process, variations of kernel functions, such as Gaussian, Epanechnikov, and Biweight, and variations of bandwidth and size of area division with a certain grid size are used to produce a more accurate density map of earthquake events. The application of KDE allows for the identification of areas with a high concentration of earthquake events, thus providing a clearer picture of areas potentially prone to earthquakes in the future.

The Getis-Ord Gi* method is used in the hotspot analysis approach to strengthen the analysis. This approach helps identify and group areas based on earthquake risk levels, ranging from low to extreme. By combining KDE and Getis-Ord Gi* analysis, this study has succeeded in mapping the most at-risk areas, which can be used as a reference in disaster mitigation efforts and risk reduction policy planning in earthquake-prone areas in Indonesia.

2.2 Study Area and Data

The study area focuses on Indonesia because it is one of the most earthquake-prone countries in the world. Indonesia is located in the Pacific Ring of Fire, an area with high seismic activity due to the movement of tectonic plates. Three large plates, namely the Indo-Australian, Eurasian, and Pacific, meet and continue to move actively around Indonesia, causing pressure on the earth's crust, which often leads to the release of energy in the form of earthquakes.

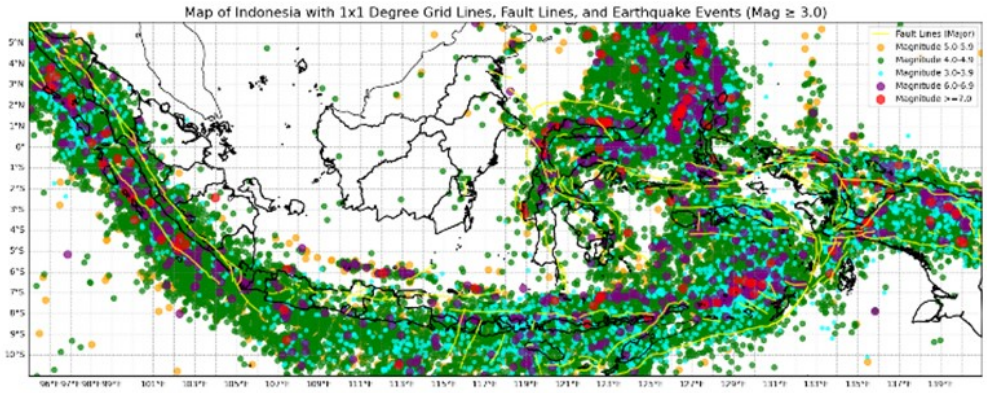


Fig. 2 Geographical coordinates and seismic intensity of earthquakes with a magnitude exceeding 3 on the Richter scale between 1964 and 2023

To support this research, earthquake data from 1964 to 2023 has been collected from a trusted public source, namely the United States Geological Survey (USGS). This data contains detailed information on the earthquake's location, magnitude, and depth. This complete data collection is very important to understand Indonesia's characteristics and seismic patterns, an area prone to earthquakes due to its position in an active tectonic area.

The historical data collected covers almost six decades, providing a strong basis for analyzing earthquake trends and frequencies. This data can map the geographic distribution and intensity of earthquakes throughout Indonesia. This information helps provide deeper insight into how earthquakes are distributed and which areas are most likely to experience earthquake risk in the future.

In addition, using this very long and detailed data allows researchers to conduct a comprehensive risk analysis. The density map of earthquake-prone areas generated from this data provides guidance to formulate more effective disaster mitigation policies. By understanding historical earthquake patterns, preventive measures in vulnerable areas can be better planned to minimize risks to Indonesia's residents and infrastructure.

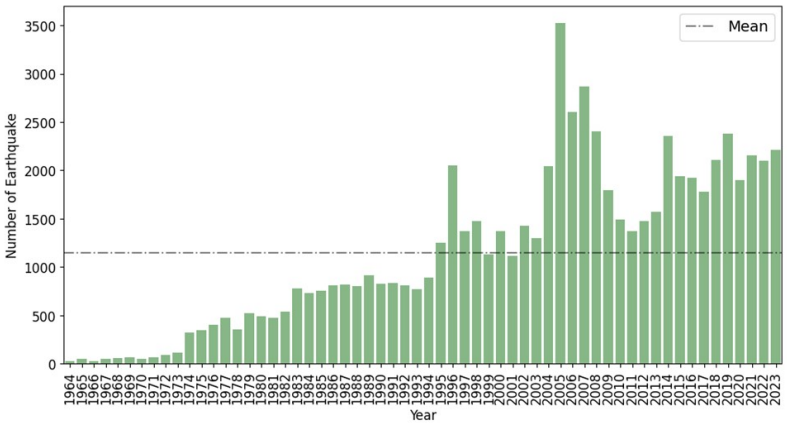


Fig. 3 Histogram displaying the frequency of earthquakes per year from 1964 to 2023

This data set contains information such as the location of the earthquake (longitude and latitude), time of the earthquake, magnitude, depth, type of magnitude, and others. This data set is then grouped based on the magnitude range, as shown in Table 1. The spatial distribution of earthquake incidents based on the magnitude range is shown in Figure 2. Figure 3 shows a Histogram displaying the frequency of earthquakes per year from 1964 to 2023, providing a clear visual representation of earthquake activity. This enables a better understanding of patterns or trends in seismic events, highlighting any noticeable increases or decreases in the frequency of occurrences.

Table 1 Number of Earthquake Event by Magnitude Range

No	Magnitude Range	Total Number of Events
1	3.0-3.9	2737
2	4.0-4.9	46370
3	5.0-5.9	10113
4	6.0-6.9	731
5	>=7.0	83

2.3 Kernel Density Estimation

An independent and identically distributed sample of n observations from a population X with an unknown probability distribution function $f(x)$ can be represented by the series $[x_1, x_2, \ldots, x_n]$. The function $K(x_i, t)$ in Equation 1, also known as a kernel function, is assigned to each i-th sample data point x_i by the kernel estimate $\hat{f}(x)$ of the original $f(x)$ Silverman (1998).

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right) \tag{1}$$

157 K(x,t) is nonnegative and bounded for all x and t

$$0 \leq K(x,t) < \infty \quad \text{for all real x,t}$$

$$\int_{-\infty}^{\infty} K(x,t)\, dt = 1 \tag{2}$$

158 Equation 2 ensures the required normalization of kernel density estimate in
159 Equation 1.

$$\hat{f}(x) = \frac{1}{n} \sum_{i=1}^{n} \int_{-\infty}^{\infty} K(x,t) dt = 1 \tag{3}$$

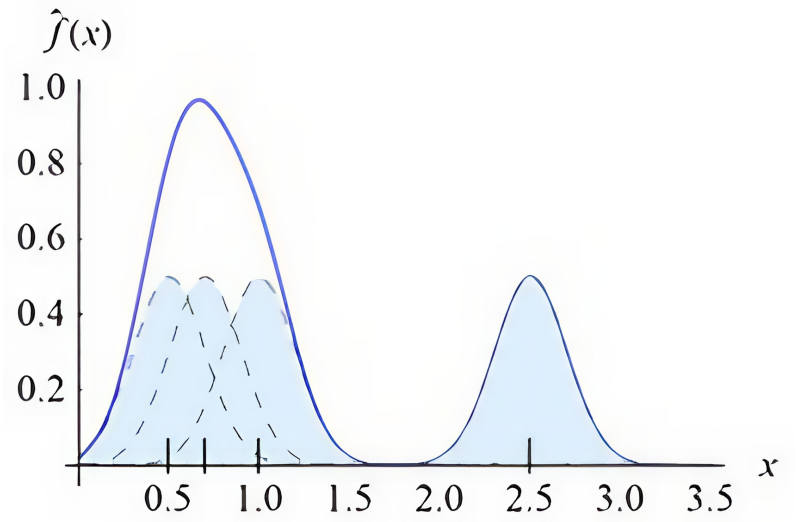


Fig. 4 Construction of kernel density estimator (1) (continuous line) with a symmetric kernel (dashed lines) for a 4-element sample (vertical segments)

160 Stated differently, the kernel converts x_i (point) location into an interval symmetri-
161 cally or asymmetrically centered around x_i in Equation 3. While asymmetric functions
162 have been employed more frequently recently, symmetric kernel functions are still used
163 in most actual applications for kernel estimation Jin and Kawczak (2003). Figure 4
164 and Figure 5 illustrate the idea of kernel estimation for both cases Węglarczyk (2018).
165 Figure 5 illustrates that the form of a symmetric kernel remains constant across
166 all sample points, but Figure 6 demonstrates that the shape of an asymmetric kernel

varies based on the location of the points. The most commonly used form of the kernel function can be expressed thanks to the symmetry property.

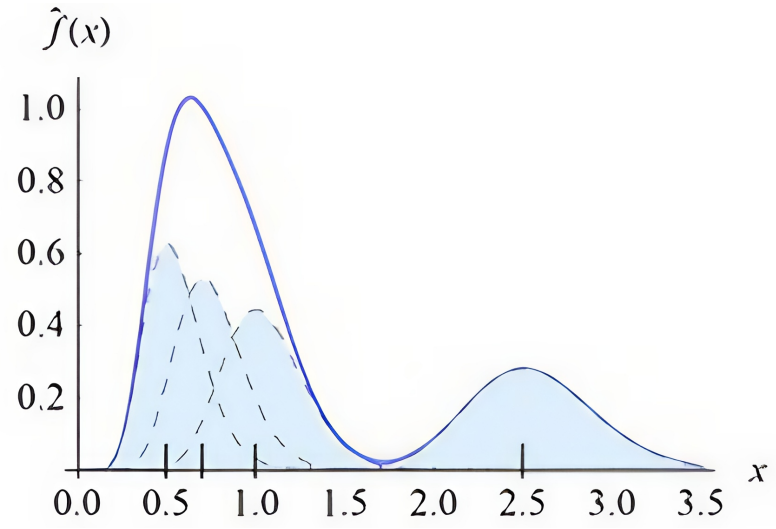


Fig. 5 Construction of kernel density estimator (1) (continuous line) with an asymmetric kernel (dashed lines) for the same 4- element sample as in Figure 4

Parameter h, the smoothing parameter, is the window width or bandwidth governing the amount of smoothing applied to the sample (Figure 6). For symmetrical kernel functions, the choice of the shape of the kernel function K(.) has rather little effect on the shape of the estimator Silverman (1998); Węglarczyk (2018), whereas as Figure 6 shows, the influence of the smoothing parameter h is critical because it determines the amount of smoothing.

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right) \tag{4}$$

A too-small value of h may cause the estimator to show insignificant details, while a too-large value of h causes over-smoothing of the information contained in the sample, which, in consequence, may mask some of the important characteristics, e.g. multimodality, of f(x) (Figure 6). A certain compromise is needed Many types of kernel functions can be found in the relevant literature. Examples of symmetric kernels are presented in Table 2. The Epanechnikov kernel is known for its efficiency in minimizing estimation errors. This kernel is parabolic in shape, meaning it gives more weight to data points closer to the center of the distribution and less weight as they move away from the center. This kernel is one of the most widely used kernels because of its balance between simplicity and accuracy.

The Biweight Kernel, also called the quartic kernel, has a smoother shape compared to the Epanechnikov kernel. It assigns weights to data points in a way that produces a

smooth curve, making it a good choice when you need a higher degree of smoothness in your density estimation. The Triangular Kernel has a simple triangular shape, where data points closer to the center receive more weight, and the weight decreases linearly as they move away from the center. This kernel is straightforward to compute, making it useful for simpler models.

Table 2 Kernel Definitions

Kernel	Definition
Epanechnikov	$K(t) = \begin{cases} \frac{3}{4\sqrt{5}}(1 - \frac{1}{5}t^2), & \text{for } \lvert t\rvert < \sqrt{5} \\ 0, & \text{for } \lvert t\rvert \geq \sqrt{5} \end{cases}$
Biweight	$K(t) = \begin{cases} \frac{15}{16}(1 - t^2)^2, & \text{for } \lvert t\rvert < 1 \\ 0, & \text{for } \lvert t\rvert \geq 1 \end{cases}$
Triangular	$K(t) = \begin{cases} 1 - \lvert t\rvert, & \text{for } \lvert t\rvert < 1 \\ 0, & \text{for } \lvert t\rvert \geq 1 \end{cases}$
Gaussian	$K(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$
Rectangular	$K(t) = \begin{cases} \frac{1}{2}, & \text{for } \lvert t\rvert < 1 \\ 0, & \text{for } \lvert t\rvert \geq 1 \end{cases}$

The Gaussian Kernel is very common because of its smooth, bell-shaped curve. Unlike other kernels, it has no cutoff points and expands without bound, making it ideal for capturing smooth and continuous distributions. These kernels are widely used in statistics and machine learning due to their versatility and effectiveness.

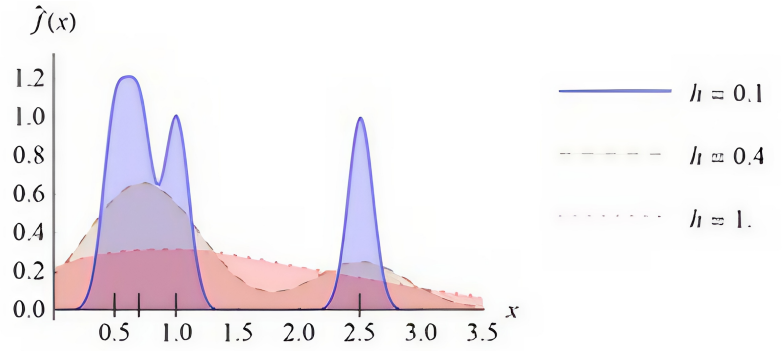


Fig. 6 The value of the smoothing parameter h influences the shape of the resulting kernel density. The 4-element samples (vertical segments) are the same as in Figure 4 and Figure 5

Rectangular (Uniform) Kernels will give equal weight to all data points within a certain range and no weight to points outside that range. They are simple and computationally efficient but lack the smoothness of other kernels. They are best used when simplicity and speed are more important than smoothness. Each kernel type has its advantages, and the choice depends on the specific needs of the analysis—whether

you need smoothness, computational efficiency, or greater accuracy in capturing data patterns.

2.4 Kernel Density Estimation Properties

Figure 7 illustrates Kernel Density Estimation and several parameters that can be considered in generating density estimates. These parameters include grid cell size, interpolation method, and bandwidth Hart and Zandbergen (2014).

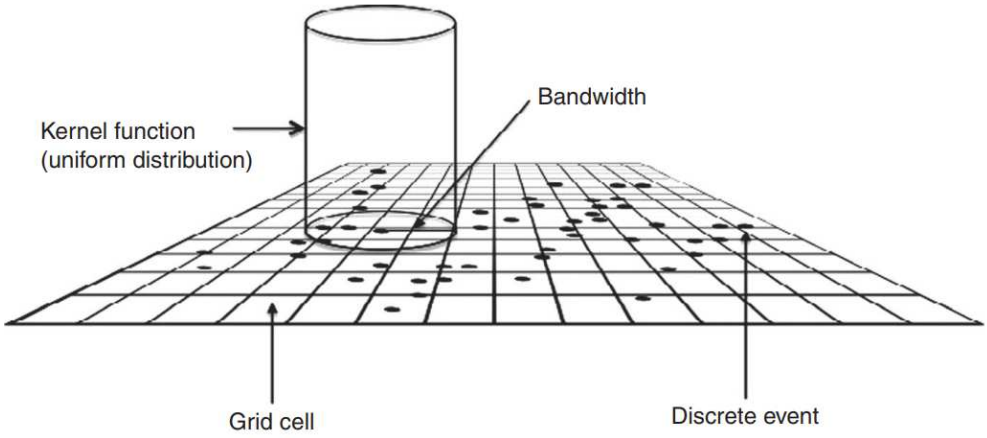


Fig. 7 An illustration of kernel density estimation

Current literature offers some guidelines for determining grid cell size, with grid sizes consisting of cells that are approximately 100 x 100 feet Chainey and Ratcliffe (2005). Two additional parameters must be set before a KDE map is produced: the interpolation method and the search radius (Figure 7). The normal distribution kernel, the uniform (flat) distribution, the quartic (spherical) distribution, and the triangular (linear) distribution are examples of different interpolation methods that can be used in crime analysis Levine (2017). When a kernel function is placed over the grid cell center points, the number of crime incidents within the bandwidth of the function is used to determine the density estimate assigned to each cell.

The values are weighted if a nonuniform kernel function is selected but not when a uniform (flat) distribution is selected. Although multiple kernel functions can be used in the interpolation process, some GIS applications limit the number of interpolation methods available or require the user to develop syntax-based routines if they are to be used. Current studies do not guide which interpolation method is most appropriate for mapping prospective crime hotspots. On the other hand, bandwidth selection has received relatively more academic attention. For example, Bailey and Gatrell suggested a bandwidth of 0.68 times the number of points raised to the power of -0.2, scaled to

the area of the study area. This value can be adjusted by multiplying it by the square root of the study area size Bailey et al (1997).

2.5 Hotspot Analysis

In conducting hotspot analysis, the concept of spatial autocorrelation is used. Spatial autocorrelation can be analyzed from two different perspectives: global and local analysis Mathur (2015). Indicators of global spatial autocorrelation include the Gi*/GetisOrd method Getis and Ord (1992) and Moran Lokal I Anselin (1995). This study uses GetisOrd Gi*. Getis and Ord introduced the Gi* statistical method to identify statistically significant spatial clusters with high values (hot spots) and low values (cold spots). The Gi* statistic will be expressed in Equation 5 Songchitruksa and Zeng (2010).

$$G_i(d) = \frac{\sum_j W_{ij}(d)x_j}{\sum_j x_j} \tag{5}$$

where:

- $Gi*$ is a statistic that describes the spatial dependency of incident i over all n accident locations, d is the distance threshold
- wij is spatial weight, which quantifies the relationship between two accident locations, i and j, based on the Fixed Distance Band
- xj is the number of accidents at location j,

Hotspot Analysis is used to perform the statistical significance test Gi*. This analysis aims to check whether high-intensity values surround high-intensity values and determine whether it is a hotspot. Meanwhile, if low-intensity values surround low-intensity values, the cluster is known as a cold spot. The output of Gi* statistics is the Gi*z and Gi*p score values. A high positive Gi*z score corresponds to a cluster with high values, known as a hotspot. A smaller negative Gi*z score implies a clustering of low values called a cold spot. Gi*z close to zero indicates a random distribution. The Hot Spot Analysis is a tool to calculate the Getis-Ord Gi* Statistics for each feature in the dataset.

The resulting z-score and p-value show where the spatial cluster of high or low-value features is. This tool looks at each feature in the context of neighboring features. A feature with a high value is interesting but may not be a statistically significant Hot Spot. To be a statistically significant Hot Spot, a feature will have a high value and be surrounded by other features with high values. For positive Z-score values, the larger the positive Z-score, the more intense the clustering, thus forming a Hot Spot.

For negative Z-scores, the smaller the negative Z-score, the more intense the clustering, thus forming a Cold Spot. Hot Spot Analysis (Getis-Ord Gi*) is included in the Mapping Clusters Toolset to locally identify spatial clusters of high values (hot spots) and low values (cold spots) that are statistically significant.

3 Results

3.1 Kernel function effect

This section will describe the research results on applying the Kernel Density Estimation method based on the kernel function in mapping earthquake-prone areas. Figure 8, Figure 9, and Figure 10 sequentially show the results of mapping earthquake-prone areas in Indonesia using the Kernel Density Estimation method using the Epanechnikov, Gaussian, and Biweight kernel functions.

Figure 8 shows the mapping results using the Epanechnikov kernel function. Based on the research, the kernel can efficiently calculate density estimates more sharply in the center of the data distribution but decreases more quickly at the edges. This mapping shows a density distribution that focuses on areas with high earthquake activity, such as around active tectonic plates in Indonesia.

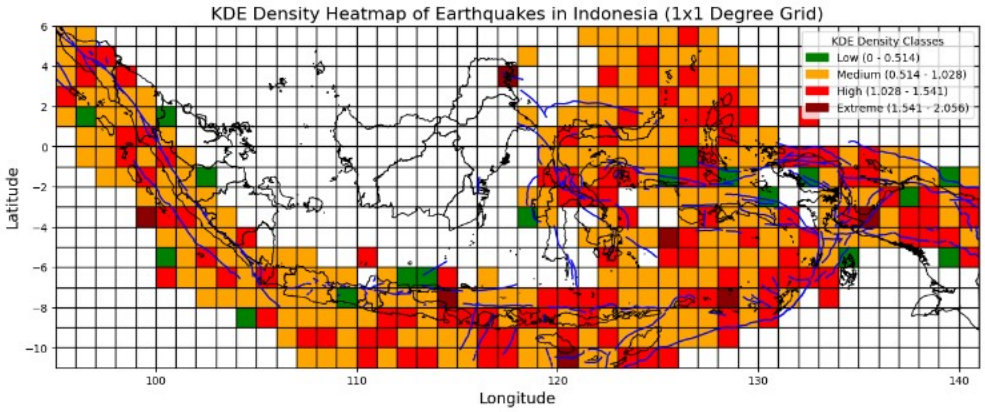


Fig. 8 Earthquake Prone Mapping Using Epanechnikov Kernel

Figure 9 shows the mapping of earthquake-prone areas using the Gaussian kernel function. Based on research conducted, this kernel tends to provide a density distribution that spreads over a wider area. So it can provide a more comprehensive understanding of areas with potential earthquake risk.

Table 3 Distribution Area Risk Level Earthquake Prone by Kernel Function

Kernel Function (BW = 0.5 Grid 1x1)	**Number of Grid**			
	Low	**Medium**	**High**	**Extreme**
Epanechnikov	18	216	114	7
Gaussian	197	167	25	7
Biweight	251	97	45	3

Furthermore, Table 3 compares the number of earthquake-prone areas in Indonesia, which are differentiated based on the level of risk of earthquake events. Table 3

compares the number of areas based on the earthquake risk category. The results show that most areas are classified as low and medium risk, but several areas are included in the high and extreme categories. Thus, Table 3 becomes an important basis for understanding the distribution of earthquake risk and planning more appropriate mitigation policies.

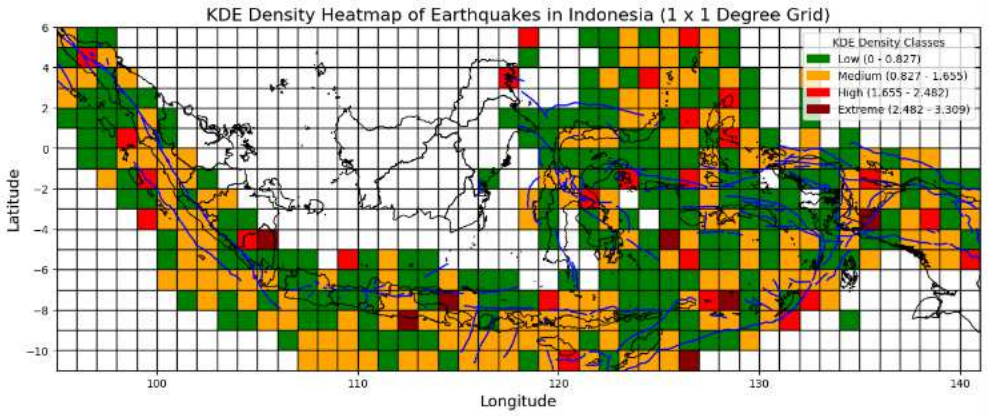


Fig. 9 Earthquake Prone Mapping Using Gaussian Kernel

Table 4 shows the results in the form of density value limits according to the level of risk in sequence, namely low, medium, high, and extreme. This table presents the density value limits for each risk level, ranging from low, medium, high, to extreme. These limits are calculated based on the distribution of earthquake density obtained from the kernel function used and serve as a quantitative reference in classifying areas based on risk. With these limits, mapping of earthquake-prone areas can be done consistently, and each area can be evaluated objectively to determine the level of alertness required.

Table 4 Kernel Function and Density Value Range Threshold

Kernel Function	**Density Value Range Threshold**			
	Low	**Medium**	**High**	**Extreme**
Epanechnikov	0.514	1.028	1.541	2.056
Gaussian	0.827	1.654	2.482	3.309
Biweight	0.644	1.288	1.932	2.576

Figure 10 provides a visualization of the mapping results using the Biweight kernel function. This kernel has properties similar to Epanechnikov but produces a wider distribution in the central density area. Thus, this mapping identifies areas that consistently show significant earthquake density.

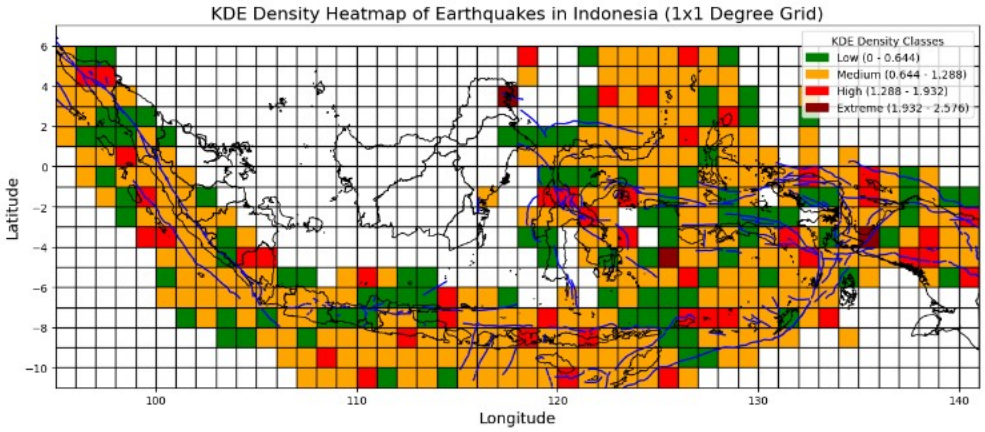


Fig. 10 Earthquake Prone Mapping Using Biweight Kernel

Table 5 shows the results of the top five earthquake-prone areas checked for statistical significance based on their z-scores and then ranked based on their estimated density based on the kernel function. Based on Table 5, each region is sorted based on the estimated density value generated. The results clearly show the areas with the highest earthquake potential. Statistical significance calculated using z-score allows researchers to assess how much deviation from the average of the density values generated.

Table 5 KDE Density and GiZScore for Various Locations with Different Kernel Functions

No	Place	KDE Density	GiZScore
Epanechnikov	1 km NNW of Tarakan, Indonesia	2.056	-0.662
	Banda Sea	1.661	0.759
	91 km S of Waingapu, Indonesia	1.639	0.351
	Papua, Indonesia	1.618	0.989
	45 km NE of Panji, Indonesia	1.603	1.01
Gaussian	Papua, Indonesia	3.309	0.868
	Banda Sea	3.198	0.656
	148 km SSE of Viqueque, Timor Leste	2.847	-0.11
	Kepulauan Barat Daya, Indonesia	2.749	1.222
	43 km SSW of Dampit, Indonesia	2.703	0.7
Biweight	1 km NNW of Tarakan, Indonesia	2.576	-0.666
	Papua, Indonesia	1.957	0.97
	Banda Sea	1.941	0.689
	45 km NE of Panji, Indonesia	1.837	0.995
	Kepulauan Barat Daya, Indonesia	1.806	1.222

In addition, Table 5 is a tool to identify areas that require more attention in disaster mitigation efforts. By knowing the top five areas that have high risk, policy makers

and decision makers can focus more on planning and implementing appropriate mitigation strategies, as well as allocating resources more efficiently to improve earthquake preparedness.

3.2 Bandwidth function effect

This section will describe the research results on applying the Kernel Density Estimation method based on the bandwidth size from the Gaussian kernel function in mapping earthquake-prone areas. Figure 11, Figure 12, and Figure 13 sequentially show the results of mapping earthquake prone areas in Indonesia using the Gaussian function Kernel Density Estimation method width bandwidth size 0.1, 0.3, and 0.5.

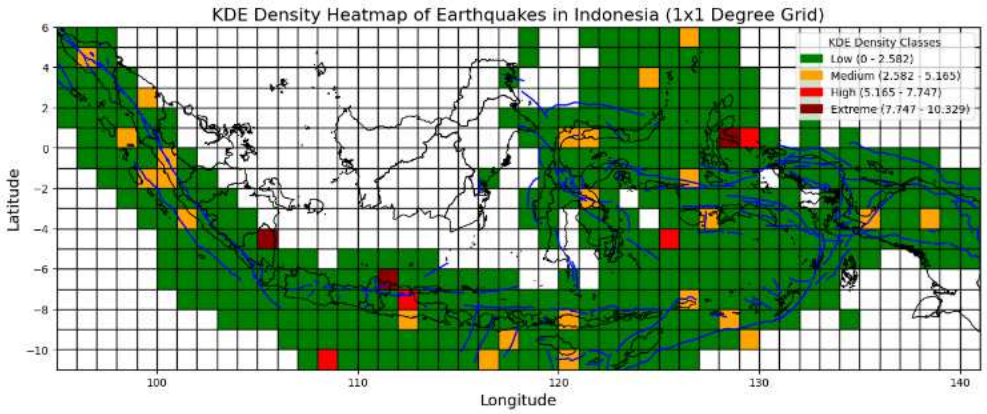


Fig. 11 Earthquake Prone Mapping Using Epanechnikov Kernel

Figure 11 shows the results of mapping earthquake-prone areas with a bandwidth size of 0.1. With a small bandwidth, mapping earthquake-prone areas tends to produce density estimates in areas with very high earthquake activity. In addition, the use of small bandwidth results in noise in the data, which has the potential to provide a less stable picture of the distribution of earthquake risk.

Furthermore, Figure 12 results from mapping earthquake-prone areas with a bandwidth size of 0.3. These bandwidth mapping results show a more even density distribution and help understand the potential risk in a wider area. While Figure 13 shows the mapping results using a bandwidth size of 0.5.

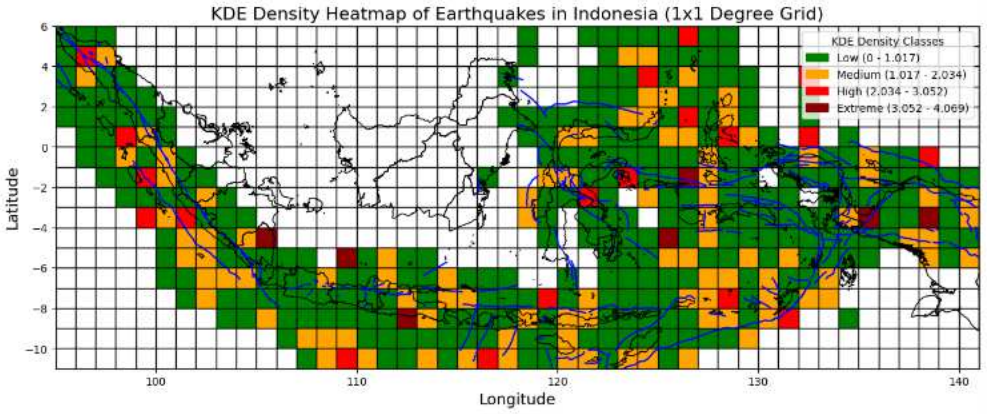


Fig. 12 Earthquake Prone Mapping Using Gaussian Kernel

Furthermore, Table 6 compares the number of earthquake-prone areas in Indonesia, which are differentiated based on the level of risk of earthquake events. Table 7 shows the results in the form of density value limits according to the level of risk in sequence, namely low, medium, high, and extreme.

Table 6 Bandwidth Size (Gaussian, Grid 1x1) and Number of Grid

Bandwidth Size (Gaussian, Grid 1x1)	Number of Grid			
	Low	Medium	High	Extreme
0.1	365	24	4	3
0.3	273	97	19	7
0.5	197	167	25	7

This bandwidth value produces a mapping showing the density distribution with a more comprehensive representation of the distribution of earthquake risk. Although this mapping can reveal general trends well, the use of large bandwidths provides the possibility of eliminating important detailed information related to areas with high earthquake activity.

Table 7 Bandwidth Size (Gaussian, Grid 1x1) and Density Value Range Threshold

Bandwidth Size (Gaussian, Grid 1x1)	Density Value Range Threshold			
	Low	Medium	High	Extreme
0.1	2.582	5.164	7.747	10.329
0.3	1.017	2.034	3.051	4.068
0.5	0.827	1.654	2.482	3.309

On the other hand, Table 7 shows the density value limits for each risk category, sorted from low, medium, high, and extreme. These limits provide important context for assessing how serious the risks are in each category. With this table, it is hoped

that stakeholders can make more targeted decisions and base preventive measures on relevant data in dealing with the threat of earthquakes in Indonesia.

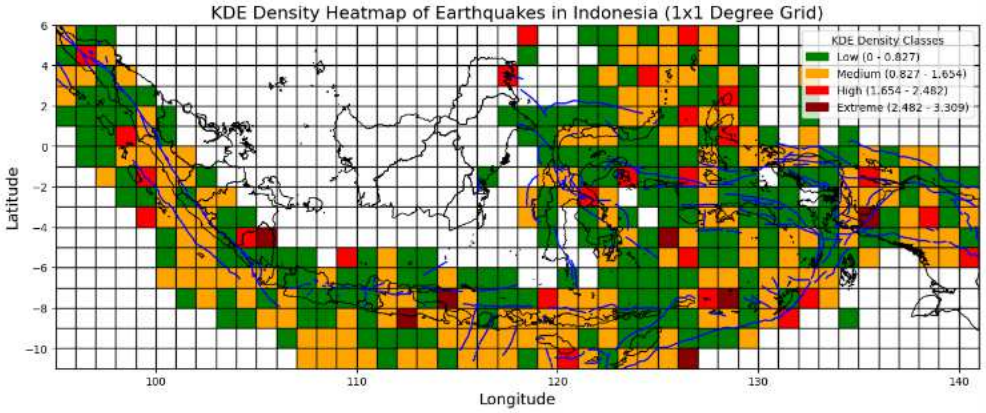


Fig. 13 Earthquake Prone Mapping Using Biweight Kernel

Table 8 shows the results of the top five earthquake-prone areas checked for statistical significance based on their z-scores and then ranked based on their estimated density using Gaussian kernel KDE based on bandwidth size. Table 8 presents the results of the analysis of the top five earthquake-prone areas that have been examined for statistical significance using z-scores, followed by ranking based on density estimates generated from the Kernel Density Estimation (KDE) method with a Gaussian kernel function and a specified bandwidth size. This table serves to identify areas with the most significant earthquake risk in Indonesia, paying particular attention to areas that show statistically high earthquake activity density.

Table 8 Bandwidth Size, Place, KDE Density, and GiZScore

Bandwidth Size	Place	KDE Density	GiZScore
Bandwidth = 0.1	18 km SE of Kotabumi, Indonesia	10.329	-0.329
	121 km ESE of Sofifi, Indonesia	10.101	0.819
	10 km NNW of Blora, Indonesia	7.999	0.538
	South of Java, Indonesia	7.258	0.325
	0 km SE of Gampengrejo, Indonesia	6.856	1.323
Bandwidth = 0.3	130 km NW of Batang, Indonesia	4.068	-0.282
	Banda Sea	4.033	0.503
	Papua, Indonesia	3.963	0.734
	18 km SE of Kotabumi, Indonesia	3.585	0.175
	Pulau-Pulau Sula, Indonesia	3.513	0.444
Bandwidth = 0.5	Papua, Indonesia	3.309	0.867
	Banda Sea	3.198	0.656
	148 km SSE of Viqueque, Timor Leste	2.848	-0.110
	Kepulauan Barat Daya, Indonesia	2.750	1.222
	43 km SSW of Dampit, Indonesia	2.703	0.701

Through this analysis, each area in the table is ranked based on the density estimate value generated from the KDE, which reflects a greater potential risk. Statistical significance calculated with z-scores allows researchers to assess the reliability and strength of the data so that the listed areas not only have high density but are also areas worthy of attention for disaster risk mitigation.

3.3 Grid Size function effect

This section will describe the research results on applying the Kernel Density Estimation method based on the grid size from the Gaussian kernel function in mapping earthquake-prone areas. Figure 14, Figure 15, Figure 16, and Figure 17 sequentially show the results of mapping earthquake prone areas in Indonesia using the Gaussian function Kernel Density Estimation method width grid size 0.25 x 0.25, 0.5 x 0.5, and 1 x 1, and 5 x 5.

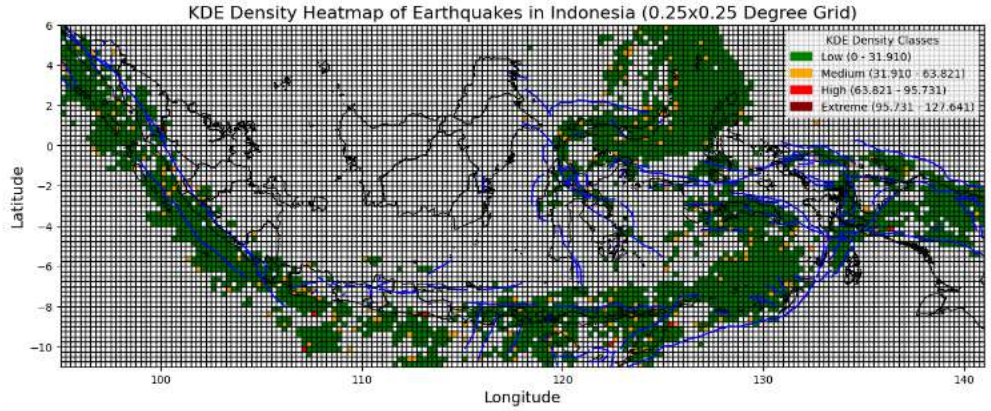


Fig. 14 Earthquake Prone Mapping Using Epanechnikov Kernel

The variation in grid size has a significant impact on the detail and interpretation of the mapping results. In Figure 14, a grid size of 0.25 x 0.25 produces a very detailed mapping, allowing the identification of small areas with high potential earthquake risk. Furthermore, Figure 15, with a grid of 0.5 x 0.5, provides a clearer and more representative picture of high-risk areas. In Figure 16, a grid size of 1 x 1 presents a more general risk distribution but reduces the ability to detect significant local variations. Finally, Figure 17 with a grid of 5 x 5 provides a broader view but at the risk of obscuring important details related to earthquake activity in certain areas.

Table 9 compares the number of earthquake-prone areas in Indonesia, which are differentiated based on the level of risk of earthquake events. Table 9 compares the number of earthquake-prone areas in Indonesia based on the level of earthquake risk. This table provides an overview of the distribution of areas in risk categories, ranging from low to extreme, which allows researchers and policymakers to identify areas that require special attention in disaster mitigation planning.

Table 9 Grid Size (Gaussian, BW = 0.5) and Number of Grid in Each Density Category

Grid Size (Gaussian, BW = 0.5)	Number of Grid			
	Low	Medium	High	Extreme
0.25 x 0.25	2486	195	7	1
0.5 x 0.5	855	294	21	3
1 x 1	155	123	43	6
5 x 5	20	3	2	4

Table 10 shows the results in the form of density value limits according to the level of risk in sequence, namely low, medium, high, and extreme. In addition, Table 10 displays the sequence density value limits for each risk level: low, medium, high, and extreme. These limits help in understanding the characteristics and intensity of each area according to the risk category so that more appropriate interventions can be designed. These two tables are important tools in formulating the mitigation steps needed to deal with potential earthquake threats in Indonesia.

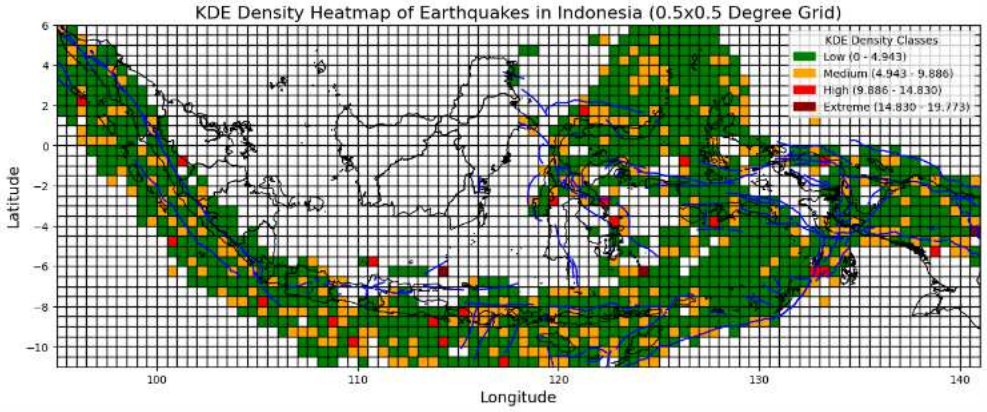


Fig. 15 Earthquake Prone Mapping Using Gaussian Kernel

Table 11 shows the results of the top five earthquake-prone areas checked for statistical significance based on their z-scores and then ranked based on their estimated density using Gaussian kernel KDE based on bandwidth size.

Table 10 Grid Size (Gaussian, Grid 1x1) and Density Value Range Threshold

Grid Size (Gaussian, Grid 1x1)	Density Value Range Threshold			
	Low	Medium	High	Extreme
0.25 x 0.25	31.910	63.821	95.731	127.641
0.5 x 0.5	4.943	9.886	14.829	19.773
1 x 1	0.827	1.654	2.482	3.309
5 x 5	0.032	0.064	0.096	0.129

Table 11 shows the analysis results of the top five earthquake-prone areas, and their statistical significance was assessed using z-scores. These areas are then ranked based on the density estimates generated through Kernel Density Estimation (KDE) with a Gaussian kernel function using a specified bandwidth size. This ranking not only provides an overview of the areas most at risk. With high statistical significance, this table emphasizes the importance of special attention to these areas in disaster mitigation planning.

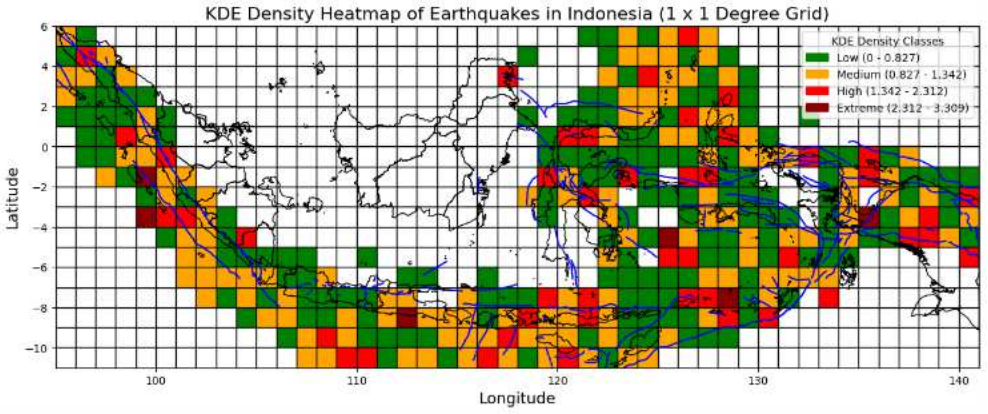


Fig. 16 Earthquake Prone Mapping Using Biweight Kernel

The information presented in Table 11 enables researchers and policymakers to prioritize resources and interventions in areas most vulnerable to earthquakes so that preventive measures can be implemented more effectively.

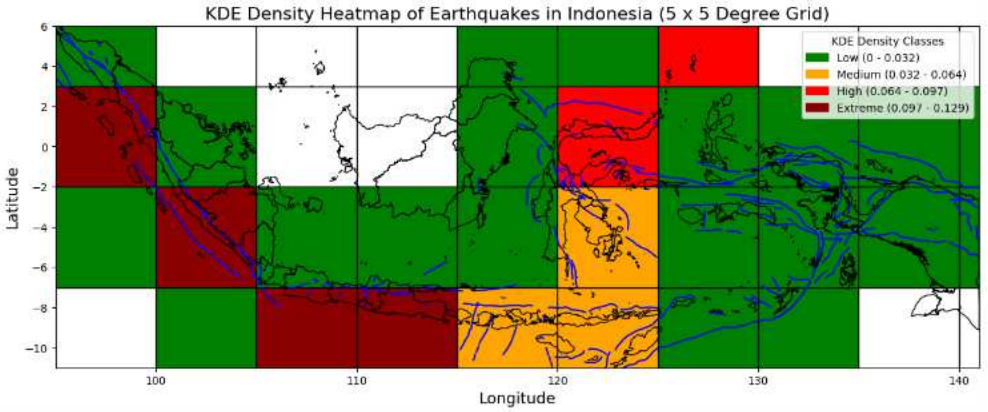


Fig. 17 Earthquake Prone Mapping Using Biweight Kernel

In addition, the results of this table are used to increase public awareness of earthquake risks in Indonesia, encouraging participation in preparedness and mitigation programs. By understanding density and risk patterns, it is hoped that more responsive and sustainable policies can be produced to deal with potential earthquake threats in Indonesia.

Table 11 Grid Size, Place, KDE Density, and GiZScore

Grid Size	Place	KDE_Density	GiZScore
0.25 x 0.25	100 km SE of Tobelo, Indonesia	127.641	0.637
	129 km WNW of Kupang, Indonesia	88.046	0.634
	south of Java, Indonesia	87.135	1.212
	101 km ESE of Padangsidempuan, Indonesia	76.584	1.120
	21 km WSW of Kencong, Indonesia	75.003	-0.155
0.5 x 0.5	101 km NNE of Sumenep, Indonesia	19.773	-0.603
	180 km SSW of Vanimo, Papua New Guinea	18.408	0.831
	216 km SE of Katabu, Indonesia	14.962	0.170
	215 km SSE of Taliwang, Indonesia	14.356	0.434
	274 km WNW of Kiunga, Papua New Guinea	13.984	0.149
1 x 1	Papua, Indonesia	3.309	0.868
	Banda Sea	3.198	0.656
	148 km SSE of Viqueque, Timor Leste	2.848	-0.110
	Kepulauan Barat Daya, Indonesia	2.750	1.222
	43 km SSW of Dampit, Indonesia	2.703	0.701
5 x 5	Southwest of Sumatra, Indonesia	0.129	0.379
	Southern Sumatra, Indonesia	0.118	0.497
	Nias region, Indonesia	0.109	-0.052
	South of Java, Indonesia	0.108	0.496
	South of Java, Indonesia	0.099	0.233

4 Discussions

4.1 Kernel function effect

The Epanechnikov kernel mostly classifies areas as areas with medium earthquake-prone levels (216 grids), followed by high (114 grids), at the low level, very few (18 grids), and in the extreme category, there are only seven grids. Furthermore, the Gaussian Kernel provides more even results, namely with classified results in areas with low (197 grids) and medium (167 grids) categories, while in the high (25 grids) and extreme (7 grids) categories. While the biweight kernel tends to produce more areas classified in the low category (251 grids), indicating a lower density distribution in general, and the medium (97 grids) and high (45 grids) categories show a narrower distribution compared to other kernel functions.

Furthermore, regarding the density value produced, the Epanechnikov kernel has a lower limit than the Gaussian and biweight kernel for each class. The Gaussian kernel shows a higher threshold value, which allows more grids to fall into the low to medium categories. The Epanechnikov kernel function produces a wide range of KDE values, with the highest value being 2.056 at 1 km NNW of Tarakan, Indonesia. However, other locations show lower KDE values and a negative GiZScore for some locations,

such as Banda Sea (1.661) and Papua, Indonesia (1.618), indicating that its statistical relevance is questionable despite measured density.

The Gaussian kernel, on the other hand, produces higher KDE values overall, with Papua, Indonesia, recording the highest value of 3.309 and a positive GiZScore (0.868), indicating that this location has significant potential for seismic activity. Other locations, such as Banda Sea (3.198) and 148 km SSE of Viqueque, Timor Leste (2.848), also show good KDE values with GiZScores close to zero or positive, indicating better statistical relevance.

Ultimately, the biweight kernel function has a fairly competitive KDE value, with a maximum value of 2,576 at 1 km NNW of Tarakan, Indonesia. However, the GiZScore values for several locations, such as Papua (1,957), show varying results, with several locations showing negative GiZScores

4.2 Bandwidth function effect

The bandwidth size of 0.1 produces the number of earthquake-prone areas in the low category (365 grids), but the number of grids in the medium (24 grids), high (4 grids), and extreme (3 grids) categories is very low. This shows that density estimates tend to be widely spread with a small bandwidth and produce many low-density areas.

Furthermore, a bandwidth of 0.3 with a medium bandwidth size decreases the number of grids in the low category (273 grids), but the number of grids in the medium category increases (97 grids). The high and extreme categories also show an increase in the number of grids (19 and 7 grids, respectively), indicating that density estimates begin to distribute higher values with a larger bandwidth.

The bandwidth size of 0.5 produces the largest number of grids in the medium category (167 grids), indicating a significant shift in distribution. Although the number of grids in the low category (197 grids) decreases, the high (25 grids) and extreme (7 grids) categories show an increase but are still relatively low compared to the number of grids in the other categories.

Bandwidth with width 0.1 produces the highest density location: 18 km SE of Kotabumi, Indonesia (KDE = 10,329). The highest positive significance was 0 km SE of Gampengrejo, Indonesia (GiZScore = 1,323), indicating significant clustering. 18 km SE of Kotabumi, Indonesia, has the highest KDE but a negative GiZScore (-0.329), indicating a spatially insignificant distribution.

Bandwidth Bandwidth with a width of 0.3 produces the highest density location: 130 km NW of Batang, Indonesia (KDE = 4,068). The highest positive significance is in Papua, Indonesia (GiZScore = 0.734), indicating moderately significant clustering. Location with low or negative GiZScore value: 130 km NW of Batang, Indonesia (GiZScore = -0.282), indicating spatially insignificant despite having high density.

A bandwidth of 0.5 yields the location with the highest density: Papua, Indonesia (KDE = 3,309). The highest positive significance: Southwest Islands, Indonesia (GiZScore = 1,222), indicating a significant cluster with a fairly high density. The location with a negative GiZScore: 148 km SSE of Viqueque, Timor Leste (GiZScore = -0.11), indicating the area is not spatially significant

4.3 Grid Size function effect

The 0.25 x 0.25 grid shows a predominance of low density, with only 1 grid in the extreme category. The 0.5 x 0.5 grid shows an increase in the medium and high categories, reflecting a higher variation in density. The 1 x 1 grid shows some areas with higher density (43 grids in the High category), although there are still grids in the low category.

The 5 x 5 grid shows a drastic decrease in the number of grids in the low and medium categories, but some areas are still identified with high density in the extreme category. Choosing the right grid size is critical for density analysis, as smaller grid sizes produce many low-density grids, while larger grid sizes allow for clearer identification of high-density areas. This shows the importance of considering grid size in density analysis to obtain more informative and useful results.

Very High Density at Small Sizes. At the 0.25 x 0.25 grid size, there are very high threshold values for all categories, with the highest value reaching 127,641 in the Extreme category. This indicates that this small area has a very high concentration of density. Density Decrease with Larger Grid Sizes As the grid size increases to 0.5 x 0.5 and 1 x 1, the threshold values for the Low, Medium, High, and Extreme categories decrease significantly. This indicates that as the grid size increases, high-density areas become sparser.

Very Low Density at the Largest Size At the 5 x 5 grid size, the threshold values for all categories are very low, with a maximum value of 0.129 in the Extreme category. This indicates that there are almost no areas of significant density at larger grid sizes.

At the 0.25 x 0.25 Grid Size, the location with the highest density is 100 km SE of Tobelo, Indonesia (KDE = 127,641). The highest positive significance is south of Java, Indonesia (GiZScore = 1.212), indicating a significant cluster with high density. The next 21 km WSW of Kencong, Indonesia, has a negative GiZScore (-0.155), indicating no significant clusters despite a fairly high KDE.

At Grid Size 0.5 x 0.5, The location with the highest density is 101 km NNE of Sumenep, Indonesia (KDE = 19,773). The highest positive significance: 180 km SSW of Vanimo, Papua New Guinea (GiZScore = 0.831), indicating significant clusters despite lower density than the 0.25 x 0.25 grid. 101 km NNE of Sumenep, Indonesia, has a negative GiZScore (-0.603), indicating an insignificant distribution despite having the highest density in this grid size.

Grid Size with size 1 x 1 The location with the highest density is Papua, Indonesia (KDE = 3,309). The highest positive significance is the Southwest Islands, Indonesia (GiZScore = 1.222), which indicates a highly significant cluster. 148 km SSE of Viqueque, Timor Leste has a negative GiZScore (-0.11), indicating a non-significant distribution despite a relatively high KDE. The 5 x 5 grid of locations with the highest density is southwest of Sumatra, Indonesia (KDE = 0.129). No locations show highly significant GiZScore values. The highest value is recorded in Southern Sumatra, Indonesia (GiZScore = 0.497), which still indicates a relatively weak cluster. The Nias region, Indonesia, has a negative GiZScore (-0.052), indicating a non-significant distribution

5 Conclusion and Future Scope

This study examines the role of kernel function, bandwidth width, and grid size in seismic risk mapping in Indonesia. Three kernel functions, namely Epanechnikov, Gaussian, and Biweight, produce seismic risk mapping findings that differ significantly. The Epanechnikov kernel produces the greatest concentration of risk in the medium category, whereas the Gaussian has a more balanced distribution of low and medium risk. Biweight, on the other hand, identifies more grids at low risk and is very sensitive to low-intensity events. Gaussian has the highest threshold value, indicating good sensitivity to tiny deviations in the data, making it an excellent choice for spotting possible earthquake hotspots.

Furthermore, bandwidth has a significant impact on risk mapping resolution and data interpretation quality. A tiny bandwidth (0.1) enables extremely high spatial resolution, but results in a granular and fragmented distribution that is difficult to discern at large scales. A broad bandwidth (0.5) produces a smoother and more balanced risk distribution, allowing for more precise identification of risk zones without sacrificing critical details. A bandwidth of 0.5 proved to be best since it provided a balance of local detail and global coverage, allowing for more consistent detection of different degrees of earthquake danger.

Grid size is a key factor in determining spatial resolution and map accuracy. Small grids (0.25x0.25 degrees) provide the best resolution, but the resulting data is sometimes too detailed for regional study. A big grid (5x5 degrees) produces data aggregation that is excessively coarse and less meaningful. The 1x1 degree grid strikes the ideal mix between resolution and regional coverage, allowing for high-accuracy identification of danger hotspots while preserving the larger spatial context. The combination of Gaussian kernel, 0.5 bandwidth, and 1x1 degree grid yields the greatest results in completely mapping earthquake risk.

This research can be further developed by including new methodologies to improve the accuracy of earthquake event density analysis. One way that can be used is the Local Outlier Factor (LOF) method to discover extreme events, particularly in the context of rare events. LOF considers the local density of the data and can provide a more accurate picture of the event distribution than KDE, which is frequently influenced by bandwidth choices. LOF allows for more precise detection of abnormalities or risk hotspots, which improves the effectiveness of risk mapping.

Furthermore, the density estimate method based on k-nearest neighbors (k-NN) can be an effective alternative. This method provides more flexible density estimation by estimating density using nearest neighbors, which is particularly beneficial in cases when the data is not uniformly distributed. The use of the k-NN approach allows for the management of more complicated spatial variations, resulting in more accurate results when determining earthquake distribution patterns. Combining this strategy with machine learning models or real-time-based AI can also significantly improve earthquake risk prediction and mitigation. . .

Acknowledgements. LPDP funded this research through the Beasiswa Pendidikan Indonesia in the Computer Science Doctoral Study Program, Department of Computer Science and Electronics, Faculty of Mathematics and Natural Sciences, Universitas

Gadjah Mada. I want to thank my supervisor and co-supervisor for their patient guidance, encouragement, and advice throughout my time as his student. I was fortunate to have supervisors and co-supervisors who cared about my situation and who responded to my questions and inquiries in the same timely manner.

Funding. LPDP provided funding through the Beasiswa Pendidikan Indonesia in the Computer Science Doctoral Study Program, Department of Computer Science and Electronics, Faculty of Mathematics and Natural Sciences, Universitas Gadjah Mada.

Declarations

Conflict of interest. The authors state that they have no known competing financial interests or personal relationships that could have influenced the work presented in this study.

Author's Contribution. Author's Contribution All authors contributed to the study's conception and design. Material preparation, data collection, and analysis were performed by [Ari Fadli], [Tri Kuntoro Priyambodo], [Agfianto Eko Putra], and [Wiwit Suryanto]. All authors read and approved the final manuscript.

References

Achmad W (2023) The effectiveness of earthquake disaster management policy in indonesia. Ganaya : Jurnal Ilmu Sosial dan Humaniora 6(2):367–377. https://doi.org/10.37329/ganaya.v6i2.2453, URL https://jayapanguspress.penerbit.org/index.php/ganaya/article/view/2453

Achu AL, Aju CD, Suresh V, et al (2019) Spatio-temporal analysis of road accident incidents and delineation of hotspots using geospatial tools in thrissur district, kerala, india. KN - Journal of Cartography and Geographic Information 69:255–265. https://doi.org/10.1007/s42489-019-00031-1

Anderson TK (2009) Kernel density estimation and k-means clustering to profile road accident hotspots. Accident Analysis Prevention 41:359–364. https://doi.org/10.1016/j.aap.2008.12.014

Anselin L (1995) Local indicators of spatial association—lisa. Geographical Analysis 27:93–115. https://doi.org/10.1111/j.1538-4632.1995.tb00338.x

Bailey TC, Gatrell AC, Crain IK (1997) Interactive spatial data analysis. Cartographica 34:69. URL https://www.proquest.com/scholarly-journals/interactive-spatial-data-analysis/docview/89070032/se-2?accountid=13771, copyright - Copyright University of Toronto Press Spring 1997

Baranyai D, Sipos T (2022) Black-spot analysis in hungary based on kernel density estimation. Sustainability 14:8335. https://doi.org/10.3390/su14148335

Barlian E, Tarigan J, Nasution Z, et al (2022) Identification of earthquake-prone areas as a regional planning study for deli serdang regency. Journal of Physics: Conference Series 2193:012015. https://doi.org/10.1088/1742-6596/2193/1/012015

Blazquez CA, Celis MS (2013) A spatial and temporal analysis of child pedestrian crashes in santiago, chile. Accident Analysis Prevention 50:304–311. https://doi.org/10.1016/j.aap.2012.05.001

Bíl M, Andrášik R, Svoboda T, et al (2016) The kde+ software: a tool for effective identification and ranking of animal-vehicle collision hotspots along networks. Landscape Ecology 31:231–237. https://doi.org/10.1007/s10980-015-0265-6

Chainey S, Ratcliffe J (2005) GIS and Crime Mapping. Wiley, https://doi.org/10.1002/9781118685181

Chainey S, Tompson L, Uhlig S (2008) The utility of hotspot mapping for predicting spatial patterns of crime. Security Journal 21:4–28. https://doi.org/10.1057/palgrave.sj.8350066

Gelfand I, Guberman S, Izvekova M, et al (1972) Criteria of high seismicity, determined by pattern recognition. Tectonophysics 13:415–422. https://doi.org/10.1016/0040-1951(72)90031-5

Getis A, Ord JK (1992) The analysis of spatial association by use of distance statistics. Geographical Analysis 24:189–206. https://doi.org/10.1111/j.1538-4632.1992.tb00261.x

Gvishiani AD, Soloviev AA, Dzeboev BA (2020) Problem of recognition of strong-earthquake-prone areas: a state-of-the-art review. Izvestiya, Physics of the Solid Earth 56:1–23. https://doi.org/10.1134/S1069351320010048

Hart T, Zandbergen P (2014) Kernel density estimation and hotspot mapping. Policing: An International Journal of Police Strategies Management 37:305–323. https://doi.org/10.1108/PIJPSM-04-2013-0039

Hashimoto S, Yoshiki S, Saeki R, et al (2016) Development and application of traffic accident density estimation models using kernel density estimation. Journal of Traffic and Transportation Engineering (English Edition) 3:262–270. https://doi.org/10.1016/j.jtte.2016.01.005

Jin X, Kawczak J (2003) Birnbaum-saunders and lognormal kernel estimators for modelling durations in high frequency financial data. Annals of Economics and Finance 4:103–124. URL https://EconPapers.repec.org/RePEc:cuf:journl:y:2003:v:4:i:1:p:103-124

Jufriansah A, Pramudya Y, Khusnani A, et al (2021) Analysis of earthquake activity in indonesia by clustering method. Journal of Physics: Theories and Applications

5:92. https://doi.org/10.20961/jphystheor-appl.v5i2.59133

Kurniati D, Fauzi MZ, Ripangi R, et al (2021) Klasterisasi daerah rawan gempa bumi di indonesia menggunakan algoritma k-medoids. MALCOM: Indonesian Journal of Machine Learning and Computer Science 1:47–57. https://doi.org/10.57152/malcom.v1i1.61

Levine N (2017) CrimeStat: A Spatial Statistical Program for the Analysis of Crime Incidents, Springer International Publishing, pp 381–388. https://doi.org/10.1007/978-3-319-17885-1_229

Mathur M (2015) Spatial autocorrelation analysis in plant population: An overview. Journal of Applied and Natural Science 7:501–513. https://doi.org/10.31018/jans.v7i1.639

Prasetio A, Effendi MM, M MND (2023) Analisis gempa bumi di indonesia dengan metode clustering. Bulletin of Information Technology (BIT) 4:338–343. https://doi.org/10.47065/bit.v4i3.820

Pulugurtha SS, Krishnakumar VK, Nambisan SS (2007) New methods to identify and rank high pedestrian crash zones: An illustration. Accident Analysis Prevention 39:800–811. https://doi.org/10.1016/j.aap.2006.12.001

Rifa IH, Pratiwi H, Respatiwulan R (2020) Clustering of earthquake risk in indonesia using k-medoids and k-means algorithms. MEDIA STATISTIKA 13:194–205. https://doi.org/10.14710/medstat.13.2.194-205

Samsudin AR, Fudholi DH, Iswari L (2024) Temporal spatial property profiling and identification of earthquake prone areas using st-dbscan and k-means clustering. Jurnal Teknik Informatika (Jutif) 5:917–929. https://doi.org/10.52436/1.jutif.2024.5.3.1293, URL http://jutif.if.unsoed.ac.id/index.php/jurnal/article/view/1293

Senduk FR, Indwiarti I, Nhita F (2020) Clustering of earthquake prone areas in indonesia using k-medoids algorithm. Indonesia Journal on Computing (Indo-JC) 4:65–76. https://doi.org/10.34818/INDOJC.2019.4.3.359, URL https://socj.telkomuniversity.ac.id/ojs/index.php/indojc/article/view/359

Sharma A, Ahuja A, Devi S, et al (2022) Use of spatio-temporal features for earthquake forecasting of imbalanced data. In: 2022 International Conference on Intelligent Innovations in Engineering and Technology (ICIIET). IEEE, pp 178–182, https://doi.org/10.1109/ICIIET55458.2022.9967687

Silverman BW (1998) Density Estimation for Statistics and Data Analysis, 1st edn.

Soloviev AA, Gvishiani AD, Gorshkov AI, et al (2014) Recognition of earthquake-prone areas: Methodology and analysis of the results. Izvestiya, Physics of the Solid Earth 50:151–168. https://doi.org/10.1134/S1069351314020116

649 Songchitruksa P, Zeng X (2010) Getis–ord spatial statistics to identify hot spots by
650 using incident management data. Transportation Research Record: Journal of the
651 Transportation Research Board 2165:42–51. https://doi.org/10.3141/2165-05

652 Srikanth L, Srikanth I (2020) A case study on kernel density estimation and hotspot
653 analysis methods in traffic safety management. In: 2020 International Conference
654 on COMmunication Systems NETworkS (COMSNETS). IEEE, pp 99–104, https:
655 //doi.org/10.1109/COMSNETS48256.2020.9027448

656 Węglarczyk S (2018) Kernel density estimation and its application. ITM Web of
657 Conferences 23:00037. https://doi.org/10.1051/itmconf/20182300037

658 Yousefzadeh M, Hosseini SA, Farnaghi M (2021) Spatiotemporally explicit earthquake
659 prediction using deep neural network. Soil Dynamics and Earthquake Engineering
660 144:106663. https://doi.org/10.1016/j.soildyn.2021.106663