

Kernel Density Estimation for Seismic Hazard Mapping in Indonesia: Influence of Kernel Function, Bandwidth Size, and Grid Resolution

Ari Fadli^{1,3}, Tri Kuntoro Priyambodo¹, Agfianto Eko Putra¹ and Wiwit Suryanto^{2,*}

¹Department of Computer Science and Electronics, Universitas Gadjah Mada, Indonesia

²Department of Physics, Universitas Gadjah Mada, Indonesia

³Electrical Engineering, Universitas Jenderal Soedirman, Indonesia

(*Corresponding author's e-mail: ws@ugm.ac.id)

Received: 28 February 2025, Revised: 9 April 2025, Accepted: 16 April 2025, Published: 20 June 2025

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 in the context of seismic hazard mapping, focusing on variations in kernel functions, bandwidth sizes, and grid resolutions. 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 resolution also significantly impacts results; smaller grids $0.25^{\circ} \times 0.25^{\circ}$ reveal detailed density patterns but may overemphasize localized concentrations, whereas larger grids $5^{\circ} \times 5^{\circ}$ are suited for macro-scale analyses but can obscure finer variations. Bandwidth size selection significantly affects density estimates. Smaller bandwidths (0.1) spread density widely, resulting in many grids in the low category but fewer 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 for disaster mitigation and spatial planning. The Gaussian kernel, 0.5 bandwidth, and $1^{\circ} \times 1^{\circ}$ grid combination yields the most significant results in mapping earthquake risk. However, the study has some limitations, including sensitivity to dataset completeness, parameter selection, and the smoothing effect of KDE that may underrepresent low-frequency events, particularly in sparse-data regions, which may introduce uncertainties in density estimations. Future studies could explore adaptive bandwidth selection and refined spatial resolution to enhance the robustness of KDE-based earthquake hazard mapping.

Keywords: Kernel Density Estimation, Seismic Hazard Mapping, Hot Spot Analysis, Earthquake Density Distribution, Grid-Based Spatial Analysis, Kernel Function Evaluation, Bandwidth Sensitivity

Introduction

Indonesia is highly vulnerable to natural hazards such as landslides, volcanic eruptions, tsunamis, earthquakes, and human-induced disasters like forest fires [1-3]. In particular, earthquakes pose a significant threat due to the country's location at the convergence of three major tectonic plates: the Indo-Australian,

Eurasian, and Pacific [4]. According to the Meteorology, Climatology, and Geophysics Agency (BMKG), Indonesia recorded over 11,000 seismic events annually in recent years, including 11,920 earthquakes in 2018 and 11,588 in 2019, with several destructive events [5]. These patterns indicate a steady

increase in both the frequency and intensity of earthquakes since 2013.

Researchers have developed various analytical methods to identify and map earthquake-prone areas in response to the rising risk of seismic disasters. These methods typically rely on statistical analysis of historical event data. Among these, Kernel Density Estimation (KDE) has emerged as a promising technique for spatial risk modeling due to its non-parametric nature and flexibility in identifying earthquake hot spots without assuming any underlying distribution [6-8]. KDE requires fewer assumptions than Bayesian probability models or machine learning-based seismic risk assessments. It provides a more intuitive spatial representation of seismic hazards, especially in regions where data availability is uneven or sparse.

Although this study focuses on KDE, it is essential to note the relevance of other mathematical models that enhance spatial risk analysis. For instance, the stability analysis methods proposed by Boonsatit *et al.* [9], Rajchakit *et al.* [10-12] offer theoretical foundations for improving consistency in dynamic environments, such as those impacted by seismic variability. Although the current study does not apply these models, they represent promising avenues for future work that may enhance KDE's spatial representations.

In addition to spatial modeling, prior studies have emphasized the significance of seismic parameters like b-value and z-value in interpreting earthquake patterns. For example, Hisyam *et al.* identified that a drop in the b-value often precedes major seismic events, while the z-value can highlight quiescent periods before an earthquake [13]. These findings support integrating all temporal analyses to understand comprehensively.

KDE has been widely applied across domains, particularly in traffic safety research, to locate accident-prone zones involving wildlife Bil *et al.* [14] and pedestrians Blazquez and Celis [15-16] as well as to conduct spatiotemporal analyses of highway incidents [17-19]. Its potential in earthquake analysis is also well-documented. Since its conceptual origin in the Earthquake-Prone Area (EPA) model introduced by Soloviev *et al.* and Gvishiani *et al.* [20-21]. KDE has enabled researchers to analyze fault density and spatial seismicity patterns for predictive modeling [22-23].

Given its non-parametric nature, Kernel Density Estimation (KDE) offers a robust and flexible approach for modeling spatial distributions without assuming any specific data distribution shape. The flexibility and robustness of KDE make it particularly valuable in earthquake risk mapping, where seismic events often exhibit complex and irregular spatial patterns. KDE produces a continuous density surface that identifies subtle hot spots, which discrete clustering methods might overlook. Its ability to provide localized density estimates also supports integrating heterogeneous data such as fault line proximity, geological structure, and historical seismicity, enhancing the precision of earthquake-prone area delineation [24-25].

Although machine learning approaches and Bayesian models have demonstrated success in seismic risk assessments, they often demand large labeled datasets and substantial parameter calibration. These requirements are not always feasible in the Indonesian context, making KDE an attractive, data-driven alternative. In addition to being computationally efficient, KDE produces visually intuitive outputs that are easy for stakeholders, such as disaster mitigation agencies and urban planners, to interpret and apply.

Several studies have employed clustering techniques, including K-Medoids and K-Means, to enhance the classification of earthquake-prone areas [26-27]. Meanwhile, other researchers have been researching using the K-means and DBSCAN algorithms. This study aims to classify the distribution of earthquakes in Indonesia in 2019 based on the values of magnitude, depth, and position [28]. Earthquake Analysis in Indonesia Using the Clustering Method aims to target areas with high earthquake-prone regions of Indonesia [29]. 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 the Identification of earthquake-prone regions [30]. Conducting Spatial-Temporal Analysis and Identification of Earthquake-Prone Areas Using ST-DBCSAN and K-Means Clustering [31]. However, unlike previous studies that mainly focus on clustering magnitude or depth, this research integrates KDE with clustering methods to develop a classification of earthquake risk levels across Indonesia.

Although numerous studies have demonstrated KDE's effectiveness in identifying accident-prone areas, particularly in traffic analysis, no prior research has systematically employed KDE to classify regions by their level of earthquake risk. This gap in the literature forms the basis of our study, which proposes a novel methodology for earthquake risk classification using KDE. This paper is structured as follows: Section 2 outlines the materials and methods, Section 3 presents the results, Section 4 discusses the results, and Section 5 concludes with a summary and future research directions.

Materials and methods

Proposed methodology

Figure 1 presents a visual overview of the research workflow, including data collection, preprocessing, spatial analysis using Kernel Density Estimation (KDE), and interpretation of seismic risk patterns. As part of this process, we collected

earthquake incident data from 1964 to 2023 from the United States Geological Survey (USGS). We also created the Indonesian boundaries used in this study in GeoJSON format. In addition, active fault line data were obtained from the Pusat Studi Gempa Nasional (PuSGeN) in shapefile (SHP) format for spatial overlay and analysis. This study defines Indonesia's boundaries using the GeoJSON format. It incorporates fault line data from the Pusat Studi Gempa Nasional (PuSGeN) to map the spatial distribution of earthquake events, as shown in **Figure 2**. Kernel Density Estimation (KDE) was applied using variations in kernel functions, bandwidth sizes, and grid resolutions to generate more accurate seismic density maps. These variations were designed to assess their influence on the resulting spatial patterns of seismic hazard. The KDE-based approach facilitates the Identification of high-density earthquake zones, thereby offering clearer insights into regions potentially vulnerable to future seismic activity.

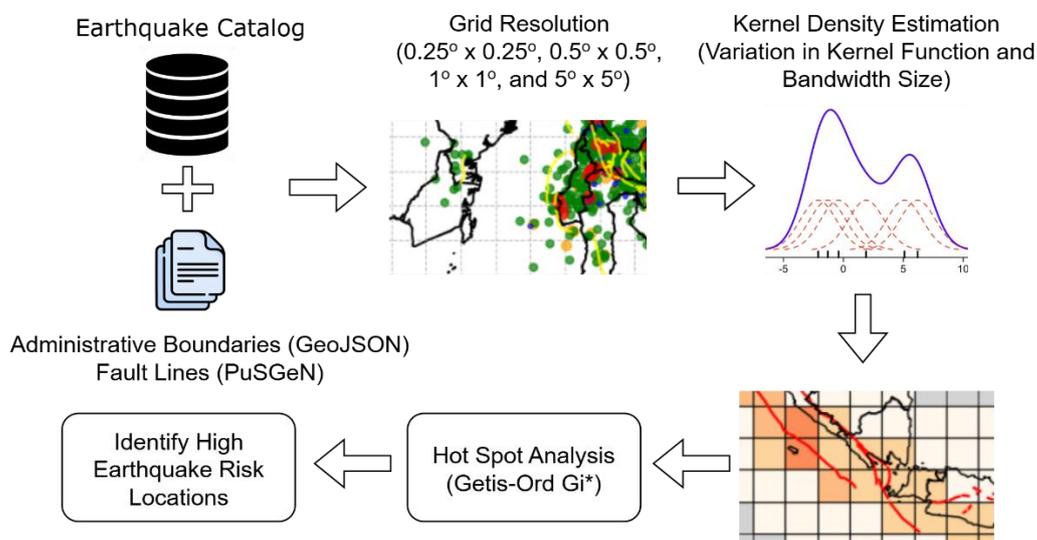


Figure 1 Research stages.

To ensure the robustness of KDE-based seismic risk mapping, this study implemented a comprehensive sensitivity analysis using multiple combinations of bandwidth sizes (0.1, 0.3, and 0.5), grid resolutions ($0.25^\circ \times 0.25^\circ$, $0.5^\circ \times 0.5^\circ$, $1^\circ \times 1^\circ$, and $5^\circ \times 5^\circ$), and kernel functions (Epanechnikov, Gaussian, and Biweight). These parameters were systematically varied to explore and identify the most optimal configuration for accurately delineating earthquake-prone areas in Indonesia.

Bandwidth selection is particularly critical in KDE, as it governs the degree of smoothness in the resulting density surface. A smaller bandwidth may capture fine-grained seismic clusters but risks introducing noise, while a larger bandwidth provides a more generalized trend that may obscure critical local variations. Grid resolution also plays a role in determining spatial granularity, with finer grids offering higher resolution at the cost of computational complexity. Meanwhile, the choice of kernel function

affects the shape and spread of the estimated density surface, where different kernels may emphasize distinct characteristics of the underlying data distribution.

Therefore, this study systematically tested a range of parameter combinations to determine the optimal configuration that yields the most accurate and interpretable representation of earthquake-prone areas. This sensitivity analysis enhances the validity and reliability of the resulting density maps by ensuring that the observed seismic risk patterns genuinely reflect underlying spatial trends rather than being artifacts of arbitrary parameter selection.

The Getis-Ord G_i^* method is used in the hot spot analysis approach to strengthen the analysis. This approach helps identify and group areas based on earthquake risk levels, ranging from low to extreme. This study successfully maps the most at-risk regions by combining KDE and Getis-Ord G_i^* analysis, offering a reference for disaster mitigation and formulating risk reduction policies in the earthquake-prone areas of Indonesia.

Study area and data

The study area focuses on Indonesia because it is one of the most earthquake-prone countries in the world. Indonesia occupies a position along the Pacific Ring of Fire, where intense seismic activity results from tectonic plate interactions. 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.

To support this research, we have collected earthquake data from 1964 to 2023 from the United States Geological Survey (USGS). This data contains detailed information on the earthquake's location, magnitude, and depth, as recorded in the seismic events. This complete data collection is significant to understanding Indonesia's characteristics and seismic patterns, an area prone to earthquakes due to its position in an active tectonic area.

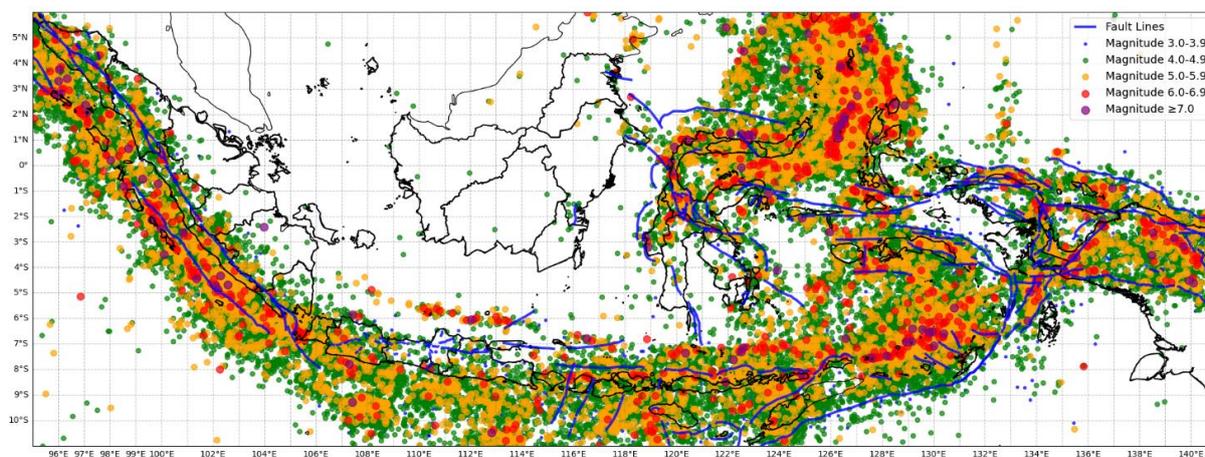


Figure 2 Geographical distribution of earthquake epicenters and seismic intensity patterns across the Indonesian region recorded between 1964 and 2023.

The historical data collected covers almost 6 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 facilitates a deeper understanding of earthquake distribution and the areas most likely to be exposed to seismic risk.

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 guides the formulation of more effective disaster mitigation policies. Understanding historical earthquake patterns enables more effective planning of preventive measures in vulnerable areas, aiming to minimize risks to Indonesia's population and infrastructure.

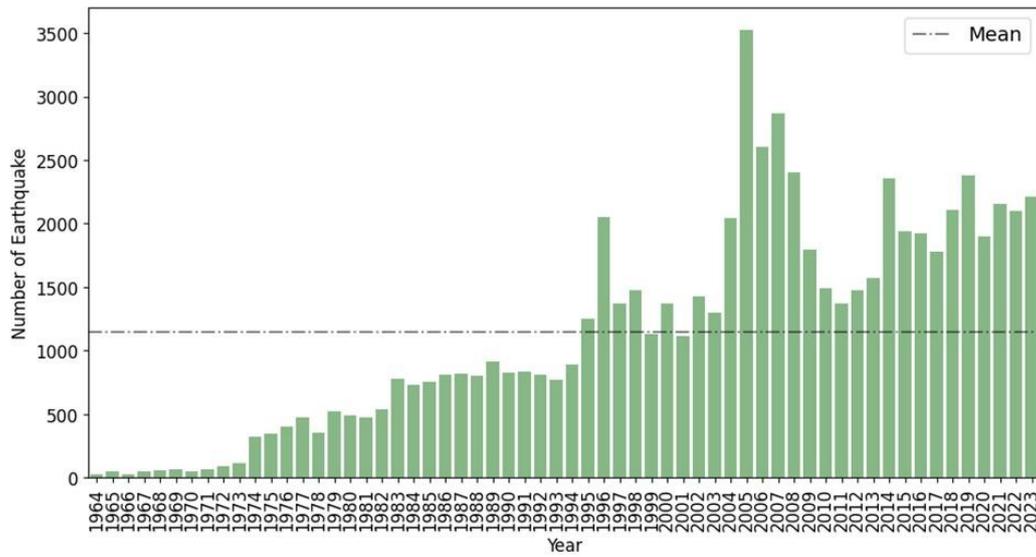


Figure 3 Histogram displaying the frequency of earthquakes per year from 1964 to 2023.

This dataset includes information such as the earthquake's location (longitude and latitude), occurrence time, magnitude, depth, magnitude type, and other relevant details. The dataset is then categorized based on magnitude range, as presented in **Table 1**. **Figure 3** displays a histogram of annual

earthquake frequency from 1964 to 2023. This histogram provides a clear visual representation of seismic activity, facilitating a better understanding of trends and patterns, including any significant increases or decreases in earthquake occurrences over time.

Table 1 Number of earthquake events 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

Kernel density estimation

Statisticians typically represent an independent and identically distributed sample of n observations from a population X with an unknown probability distribution function $f(x)$, generally defined as a set of random variables. $[x_1, x_2, \dots, x_n]$. The function $K(x_i, x)$ in Eq. (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 $f(x)$ [32].

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

$K(x_i, x)$ is nonnegative and bounded for all x_i and x
 $0 \leq K(x_i, x) < \infty$ for all real x_i, x

Eq. (2) ensures the required normalization of kernel density estimate in Eq. (1).

$$\int_{-\infty}^{\infty} K(x_i, x) = K\left(\frac{x - x_i}{h}\right) \tag{2}$$

The kernel function transforms the location of each observation. x_i into an interval that is either symmetrically or asymmetrically centered around x_i , as expressed in Eq. (3). While recent studies increasingly utilize asymmetric kernel functions, symmetric kernels

remain predominant in most practical applications of kernel density estimation (KDE) [33]. **Figure 4** and **Figure 5** illustrate this concept, where the symmetric kernel maintains a constant shape across all sample

points (**Figure 5**). In contrast, the asymmetric kernel's shape varies according to the location of the data points (**Figure 6**) [34]

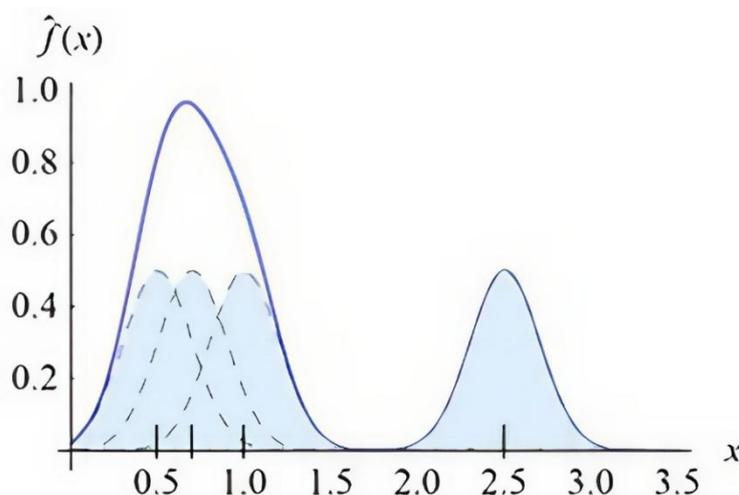


Figure 4 Construction of kernel density estimator (1) (continuous line) with a symmetric kernel (dashed lines) for a 4-element sample (vertical segments).

$$\int_{-\infty}^{\infty} K\left(\frac{x - x_i}{h}\right) dt = 1 \tag{3}$$

The symmetry property of the kernel function allows for a more standardized expression of KDE. The smoothing parameter h , bandwidth, or window width, is crucial in determining the extent of smoothing applied to the sample (**Figure 6**). In the case of symmetric kernels, the specific shape of the kernel function $K(\cdot)$ has relatively little influence on the general form of the estimator [32], [34]. However, the choice of h is crucial: a bandwidth that is too small may lead to overfitting, capturing noise, and insignificant fluctuations in the data, while a bandwidth that is too large may excessively smooth the distribution, potentially obscuring critical structural features such as multimodality. Thus, selecting an appropriate bandwidth involves a careful trade-off between bias and variance in the estimator.

Researchers have developed and documented various kernel functions in the literature to accommodate these considerations. **Table 2** summarizes examples of symmetric kernels commonly applied in KDE. Researchers recognize the Epanechnikov kernel's optimality in minimizing the

mean integrated squared error. Its parabolic shape assigns greater weight to observations near the center of the distribution and gradually less to those further away. Practitioners widely utilize it in practice due to its balance between efficiency and computational simplicity.

Other symmetric kernels offer varying levels of smoothness and computational characteristics. The Biweight kernel (also referred to as the quartic kernel) provides a smoother density estimate than the Epanechnikov by assigning weights more gradually, making it suitable for analyses requiring higher continuity. In contrast, the Triangular kernel applies linearly decreasing weights from the center and is favored by practitioners for its ease of implementation and low computational cost, especially in simpler modeling contexts. The Gaussian kernel, known for its smooth and continuous bell-shaped curve, has infinite support and is particularly effective for modeling distributions that require continuity without strict boundaries.

Meanwhile, the Rectangular (Uniform) kernel assigns equal weight to all observations within a fixed interval and zero weight outside it. While this approach lacks the smoothness provided by other kernels, it is highly efficient and often used in applications where

computational simplicity is prioritized over estimator smoothness. Ultimately, the choice of kernel depends on the analysis’s specific objectives, whether

emphasizing smoothness, accuracy, interpretability, or computational performance.

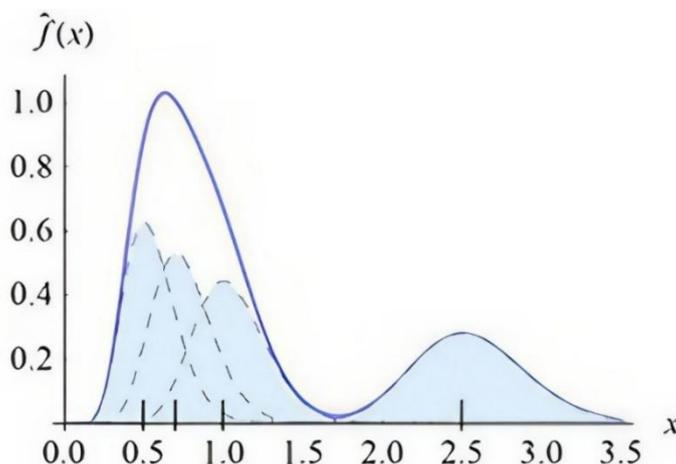


Figure 5 Construction of kernel density estimator (1) (continuous line) with an asymmetric kernel(dashed lines) for the same 4-element sample as in **Figure 4**.

Overall, **Table 2** summarizes the most commonly used symmetric kernel functions along with their respective characteristics, offering a practical reference for selecting an appropriate kernel in KDE applications. Each kernel provides a distinct trade-off between smoothness, computational efficiency, and sensitivity to data structure. The decision to use a particular kernel should align with the objectives of the

analysis, whether emphasizing smoothness, accuracy, interpretability, or computational performance. In the context of seismic hazard mapping, where accurate detection of earthquake-prone zones is crucial, the selection of a suitable kernel can significantly improve the reliability and interpretability of the resulting density estimates.

Table 2 Kernel definitions.

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

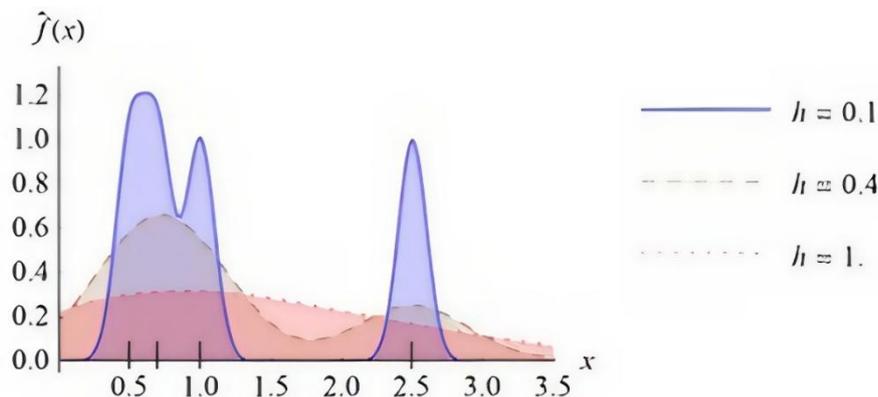


Figure 6 The smoothing parameter h 's value influences the shape of the resulting kernel density. The 4-element samples (vertical segments) are the same as in **Figures 4** and **5**.

Kernel density estimation properties

Figure 7 illustrates Kernel Density Estimation and highlights key parameters researchers typically consider when generating density estimates. These parameters include grid cell size, interpolation method, and bandwidth [35]. Current literature offers some guidelines for determining grid cell size, with grid sizes consisting of cells that are approximately 100×100 feet [36]. Before generating a KDE map, users must set 2 additional parameters: the interpolation method and the

search radius **Figure 7**. The standard 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 [37]. When analysts place a kernel function over the center points of grid cells, they use the number of crime incidents within the function's bandwidth to determine the density estimate assigned to each cell.

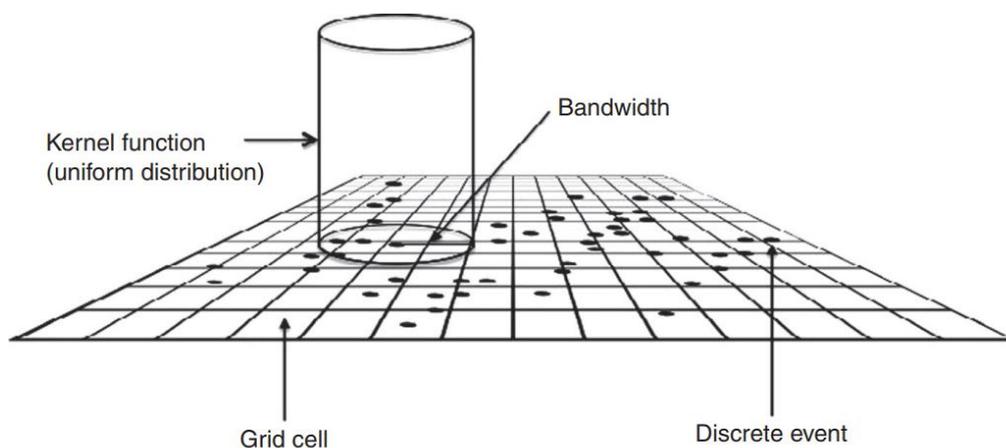


Figure 7 An illustration of kernel density estimation.

The values are weighted if a nonuniform kernel function is selected but not when a uniform (flat) distribution is selected. Although analysts may apply various kernel functions during the interpolation process, some GIS applications limit the available interpolation methods or require users to create syntax-based routines to use additional functions. Current studies do not guide which interpolation method is

most appropriate for mapping prospective crime hot spots. 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. The user can adjust this value by multiplying it by the square root of the study area size [38].

Effectiveness of KDE in hazard mapping

Kernel Density Estimation (KDE) is employed in this study to estimate the probability density function of earthquake occurrences using the formulation presented in Eq. (1). In this formulation, h denotes the bandwidth parameter controlling the density estimate's smoothness. At the same time, K is the kernel function, such as Gaussian, Epanechnikov, Box, or Triangle, that determines the contribution of each data point x_i . By applying KDE to historical earthquake data, researchers can generate continuous spatial density surfaces to visualize regions of elevated seismic activity more precisely.

The application of KDE in spatial analysis has shown strong performance across various domains. For instance, in a study titled A Case Study on Kernel Density Estimation and Hotspot Analysis Methods in Traffic Safety Management, KDE was employed to identify hazardous road segments by analyzing the spatial distribution of traffic accidents. By converting point-based incident data into continuous density surfaces, KDE facilitated the detection of accident hot spots, enabling more targeted and evidence-based interventions. This application underscores the strength of KDE in capturing latent spatial patterns that may not be visible through traditional aggregation or zoning techniques [6]. Similarly, it demonstrated that KDE consistently outperformed alternative hot spot mapping techniques such as spatial ellipses, nearest neighbor analysis, and thematic mapping in predicting future crime locations. These studies highlight KDE's effectiveness in capturing latent spatial patterns and producing smooth, continuous surfaces that enhance interpretability [7].

Furthermore, KDE was applied using the same mathematical formulation described in Eq. (1) to estimate the probability density function of accident-related variables. Through this approach, researchers could identify locations with high concentrations of accidents and define boundaries of critical sections. The Identification of high-risk locations and the definition of essential boundaries facilitated the prioritization of preventive measures and infrastructure improvements. The study illustrates that the KDE method is not only effective in the context of road accidents but also holds significant potential for disaster risk reduction, including earthquake-related

hazard mapping, especially when incorporating relevant factors such as Annual Average Daily Traffic (AADT) [8].

These prior applications of KDE in diverse domains, ranging from traffic safety Srikanth and Srikanth [6] and crime prediction Chainney *et al.* [7] to infrastructure planning for disaster risk reduction Boonsatit and Sipos [8] collectively demonstrate its robustness in identifying spatial hot spots based on historical event data. The methodological consistency across these studies, all utilizing the same KDE formulation, reinforces its adaptability and reliability in various contexts. Given its proven capability to detect high-risk zones and represent spatial phenomena continuously, KDE emerges as a highly suitable and powerful tool for seismic hazard mapping. By similarly leveraging historical earthquake data, KDE can aid in identifying earthquake-prone regions with greater precision, ultimately supporting more informed mitigation strategies and regional planning initiatives.

Hot spot analysis

Hot spot analysis is a spatial statistical method that identifies statistically significant clusters of high values (hot spots) and low values (cold spots) based on spatial autocorrelation. Researchers typically analyze spatial autocorrelation from global and local perspectives [39]. Common global indicators include the Getis-Ord G_i^* statistic [40] and Moran's I [41].

This study adopts the Getis-Ord G_i^* method to detect local spatial clusters. Getis and Ord introduced the G_i^* statistic to determine whether high (or low) values cluster spatially in a statistically significant way. Eq. (4) defines the G_i^* statistic [42].

$$G_i(d) = \frac{\sum_j W_{ij}(d)x_j}{\sum_j x_j} \quad (4)$$

Where: G_i^* Measures the degree of spatial association for location i within a distance threshold d . $W_{ij}(d)$ is the spatial weight between locations i and j (often using a fixed distance band). x_j Denotes the number of incidents (e.g., accidents) at location j .

Hot spot analysis is used to assess the statistical significance of these spatial clusters. The analysis output includes each feature's G_i^* z-score and G_i^* p-

value. A high positive z-score indicates clustering of high values (hot spots). At the same time, a low negative z-score reveals clustering of low values (cold spots). A z-score near zero suggests a random spatial pattern.

The Hot spot Analysis tool (Getis-Ord G_i^*) evaluates each feature in the context of its neighbors. A feature is considered a statistically significant hot spot only when it has a high value and its neighboring features also have high values. The higher the positive z-score, the more intense the clustering of high values. Similarly, the lower the negative z-score, the more intense the clustering of low values. This analysis is implemented in spatial GIS environments under the

Mapping Clusters Toolset, enabling the local Identification of spatial clusters that are statistically meaningful.

Results

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 - 10** sequentially presents the spatial distribution of seismic hazard zones in Indonesia, derived using the Kernel Density Estimation (KDE) method with the Epanechnikov, Gaussian, and Biweight kernel functions.

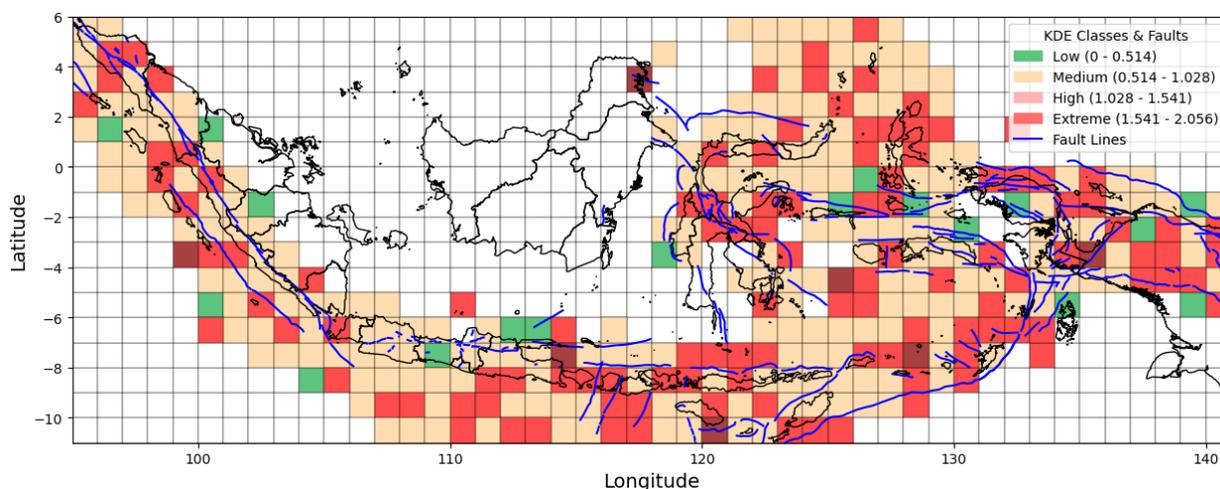


Figure 8 Map showing the spatial distribution of seismic hazard zones in Indonesia, derived from Kernel Density Estimation (KDE) using the Epanechnikov kernel function.

Figure 8 shows the mapping results using the Epanechnikov kernel function. Based on the research, this kernel can efficiently estimate density values by assigning greater weight to observations near the center and decreasing more rapidly toward the edges. As a result, the density distribution appears more localized and sharply concentrated in regions with intense seismic activity. This mapping highlights several critical zones along Indonesia's tectonic boundaries, indicating the Epanechnikov kernel's effectiveness in emphasizing core seismic regions while minimizing noise in less active areas.

Table 3 shows the number of grid representing earthquake-prone areas in Indonesia, categorized by KDE classes under different kernel functions, using a fixed bandwidth of 0.5 and a $1 \times 1^\circ$ grid resolution. The

results show that the analysis classifies most areas as low and medium risk but places several areas in the high and extreme categories. This indicates that while seismic activity is widespread, the intensity and potential impact vary significantly across regions. Notably, each kernel function produces different classification patterns, reflecting each method's sensitivity in identifying spatial concentrations of earthquake risk. For example, the Gaussian kernel tends to identify more low-risk areas, while the Epanechnikov and Biweight kernels detect more concentrated zones of higher risk. Thus, **Table 3** becomes an essential basis for understanding the distribution of earthquake risk and planning more appropriate and region-specific mitigation policies.

Table 3 Number of grids in each KDE Classes for different kernel functions, using a bandwidth of 0.5 and a 1×1 grid resolution.

Kernel Function (BW = 0.5, Grid 1°×1°)	KDE Classes			
	Low	Medium	High	Extreme
Epanechnikov	18	216	114	7
Gaussian	197	167	25	7
Biweight	251	97	45	3

Table 4 presents the threshold density values delineating the low, medium, high, and extreme seismic-risk classes for three kernel functions (Epanechnikov, Gaussian, and Biweight), calculated using a bandwidth of 0.5 and a 1 × 1° grid resolution. Researchers estimate these limits based on the

distribution of earthquake density derived from the applied kernel function and use them as a quantitative reference for classifying areas by risk. These limits enable researchers to consistently map earthquake-prone regions and objectively assess each area's required alertness level.

Table 4 KDE thresholds value for classifying density levels using three different kernel functions, using a bandwidth of 0.5 and a 1×1 grid resolution.

Kernel Function (BW = 0.5, Grid 1°×1°)	KDE Threshold Value			
	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 9 presents Indonesia's spatial distribution of seismic hazard zones, generated using Kernel Density Estimation (KDE) with the Gaussian kernel function. Based on research conducted, this kernel tends to provide a density distribution that spreads over a wider area, enabling a more comprehensive representation of seismic risk. The Gaussian kernel's ability to produce smooth and continuous density

surfaces allows for more explicit Identification of seismic hot spots across Indonesia, especially in regions with complex tectonic structures. Areas with high and extreme density values are notably concentrated along major tectonic zones, such as the Sumatra Fault and Java Trench, which reflects the effectiveness of this kernel in highlighting critical seismic regions.

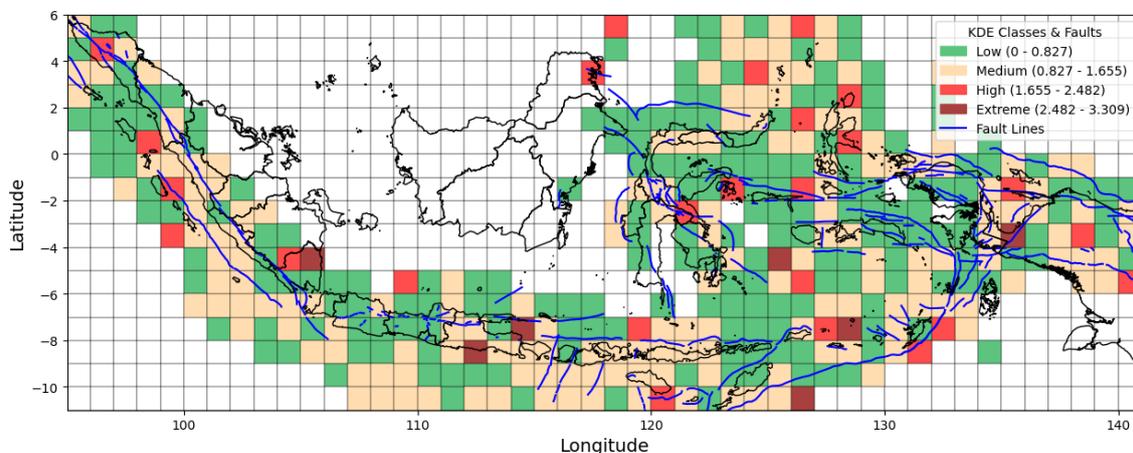


Figure 9 Map showing the spatial distribution of seismic hazard zones in Indonesia, derived from Kernel Density Estimation (KDE) using the Gaussian kernel function.

Figure 10 provides a visualization of the mapping results using the Biweight kernel function. This kernel shares characteristics similar to those of the Epanechnikov kernel but generates a broader and smoother distribution within the central density region. As a result, it can highlight areas with consistent seismic activity while maintaining a balance between central emphasis and peripheral sensitivity. The density mapping reveals prominent seismic zones, particularly in regions with frequent tectonic activity, indicating the kernel's suitability for capturing sustained earthquake-

prone areas with moderate smoothing. **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 the z-score allows researchers to assess how much deviation from the average density values is generated.

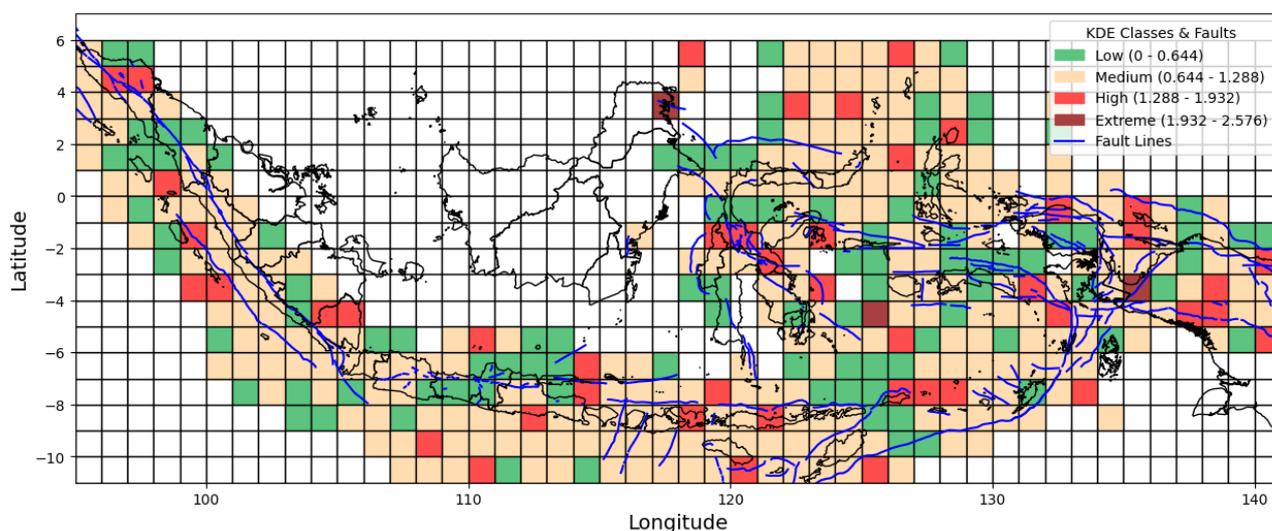


Figure 10 Map showing the spatial distribution of seismic hazard zones in Indonesia, derived from Kernel Density Estimation (KDE) using the Biweight kernel function.

In addition, **Table 5** is a tool for identifying areas that require more attention in disaster mitigation efforts. By knowing the top five areas that are high risk, policymakers and decision-makers can focus more on planning and implementing appropriate mitigation strategies and allocating resources more efficiently to improve earthquake preparedness. This targeted approach enables more effective risk reduction by prioritizing regions with high seismic activity and statistical significance. It also facilitates the development of localized early warning systems, critical infrastructure reinforcement, and seismic risk integration into spatial and land-use planning. Furthermore, the insights provided by **Table 5** can

support evidence-based decision-making, helping to enhance the resilience of vulnerable communities and reduce potential losses from future earthquake events.

The Identification of these high-risk zones also provides a foundation for cross-sectoral collaboration between government agencies, local authorities, and disaster management organizations. Incorporating the KDE-based results into national disaster risk reduction frameworks may lead to more systematic preparedness efforts, particularly in provinces with limited response capacity.

Table 5 Top five earthquake-prone areas in Indonesia ranked by statistical significance (G_i^* z-scores) and estimated density value across varying kernel functions.

Kernel Function	Location	Density Value	G_i^* z-score
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	1.010
Gaussian	Papua, Indonesia	3.309	0.868
	Banda Sea	3.198	0.656
	Kepulauan Barat Daya, Indonesia	2.749	1.222
	43 km SSW of Dampit, Indonesia	2.703	0.700
	18 km SE of Kotabumi, Indonesia	2.519	0.350
Biweight	1 km NNW of Tarakan, Indonesia	2.576	-0.666
	Papua, Indonesia	1.957	0.970
	Banda Sea	1.941	0.689
	45 km NE of Panji, Indonesia	1.837	0.995
	Kepulauan Barat Daya, Indonesia	1.806	1.222

Bandwidth size 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 - 13** sequentially shows the results of mapping earthquake-prone areas in Indonesia using the Gaussian

Kernel Density Estimation method with bandwidth sizes of 0.1, 0.3, and 0.5.

Figure 11 illustrates Indonesia's spatial distribution of seismic hazard zones, derived from Kernel Density Estimation (KDE) using a Gaussian kernel with a bandwidth of 0.1. Applying such a small bandwidth leads to highly localized density estimates, predominantly highlighting regions with intense seismic activity. While this enhances the detection of

concentrated seismic hot spots, it also introduces considerable noise into the analysis, potentially resulting in an unstable and fragmented representation of the overall seismic risk distribution. Consequently, the visualization may overemphasize isolated events and underrepresent broader regional trends, reducing its effectiveness for comprehensive hazard assessment and decision-making.

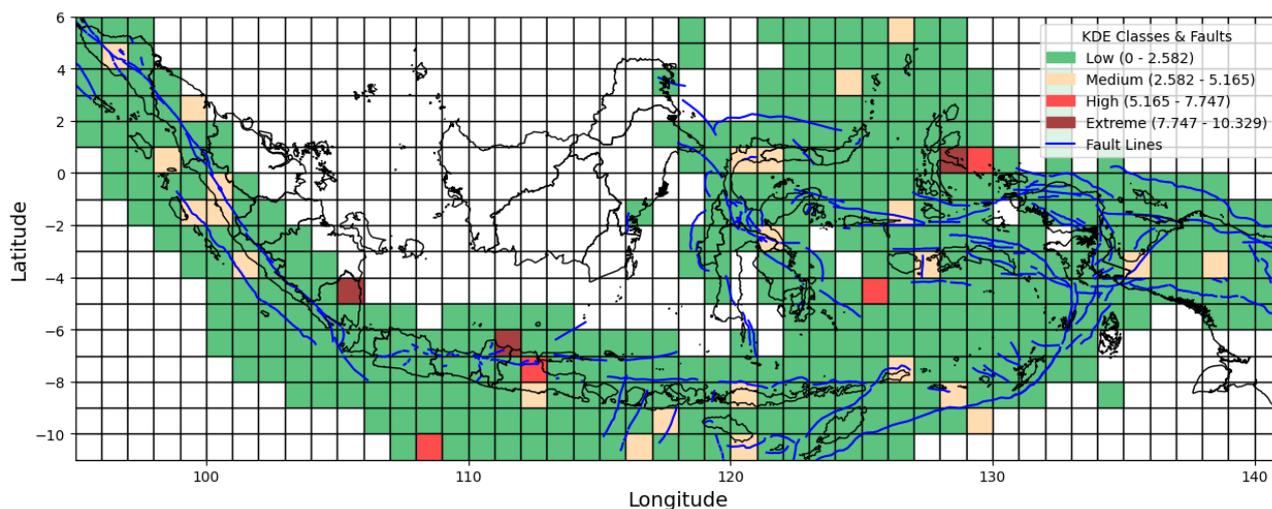


Figure 11 Map showing the spatial distribution of seismic hazard zones in Indonesia, derived from Kernel Density Estimation (KDE) using a Gaussian kernel with a bandwidth of 0.1.

Furthermore, **Figure 12** spatial distribution of seismic hazard zones in Indonesia, derived from Kernel Density Estimation (KDE) using a Gaussian kernel with a bandwidth of 0.3, which yields a more balanced and smoothed density distribution across the study area. This configuration reduces local fluctuations and noise compared to the narrower bandwidth, enabling a

more precise visualization of broader seismic trends. The resulting map effectively captures concentrated seismic hot spots and surrounding transitional zones, facilitating a more comprehensive interpretation of earthquake risk potential over a wider geographic area. This makes it particularly useful for regional-scale seismic hazard assessment and spatial planning.

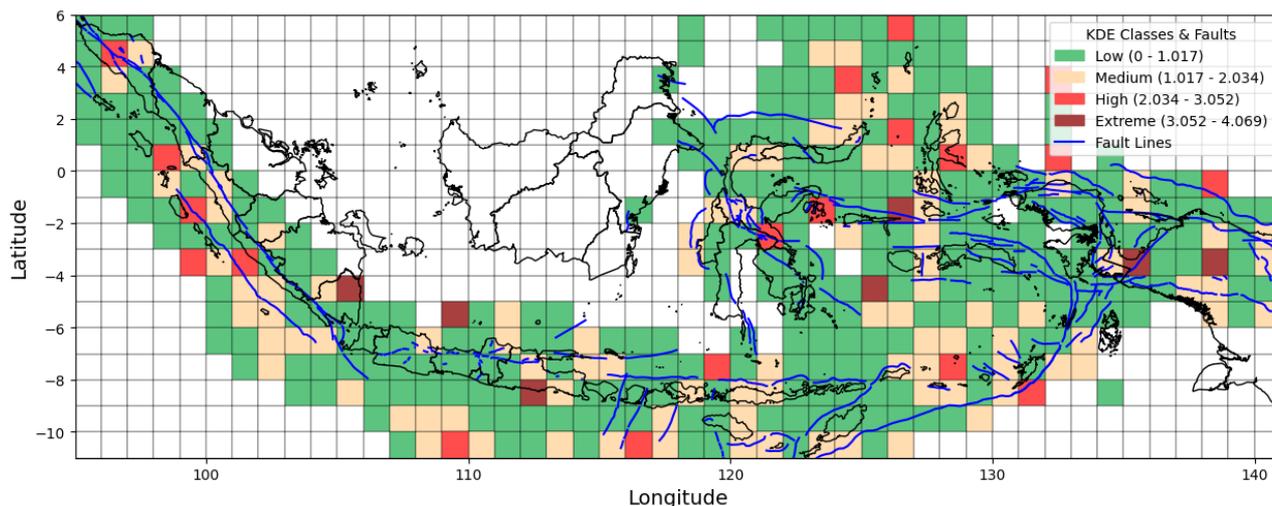


Figure 12 Map showing the spatial distribution of seismic hazard zones in Indonesia, derived from Kernel Density Estimation (KDE) using a Gaussian kernel with a bandwidth of 0.3.

Table 6 shows the number of grids representing earthquake-prone areas in Indonesia, categorized by KDE classes with different bandwidth sizes of 0.1, 0.3, and 0.5 for a 1°×1° grid. This table highlights how the choice of bandwidth influences the distribution of earthquake risk areas, with larger bandwidths yielding more comprehensive representations of risk. However, larger bandwidths may also obscure important details, particularly in areas with high earthquake activity. As the bandwidth increases, the number of areas classified as low-risk decreases, while the number of areas in the

medium and high-risk categories increases. This suggests that broader smoothing captures more regional seismicity patterns but reduces local sensitivity. For instance, a bandwidth of 0.1 identifies only a small number of medium to extreme risk areas, while a bandwidth of 0.5 reveals a more balanced distribution across all risk levels. These findings emphasize the importance of selecting an appropriate bandwidth when using KDE to support accurate seismic hazard mapping and informed mitigation planning.

Table 6 Number of grids in each KDE Class for three bandwidth sizes, using a Gaussian kernel and a 1°×1° grid resolution.

Bandwidth size (Gaussian, Grid 1°×1°)	KDE Classes			
	Low	Medium	High	Extreme
0.1	365	24	4	3
0.3	273	97	19	7
0.5	197	167	25	7

Table 7 provides the corresponding density value thresholds for each risk category. These thresholds help understand the severity of risks across different areas, as they define the boundaries separating low, medium, high, and extreme risk zones. The thresholds enable a more objective assessment of seismic hazard levels by offering a clear numerical basis for classification.

Using these values, stakeholders can make more informed decisions and implement targeted preventive measures to address earthquake risks in Indonesia. In this way, the thresholds support spatial analysis and serve as a practical guide for developing effective disaster risk reduction strategies.

Table 7 KDE thresholds value for classifying density levels using three bandwidth sizes, a Gaussian kernel, and a 1°×1° grid resolution.

Bandwidth size (Gaussian, Grid 1°×1°)	KDE Threshold Value			
	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, **Figure 13** illustrates the results of earthquake-prone area mapping using a bandwidth size of 0.5. This larger bandwidth produces a more generalized and smoothed density distribution, effectively minimizing local variability and noise. As a result, broader regional trends become more apparent, aiding in identifying large-scale seismic patterns.

While the finer details and localized high-density clusters may be less distinct than smaller bandwidths, this approach is particularly advantageous for macro-level seismic hazard analysis. It supports strategic planning and policy-making by emphasizing wider zones of consistent seismic risk rather than isolated hot spots.

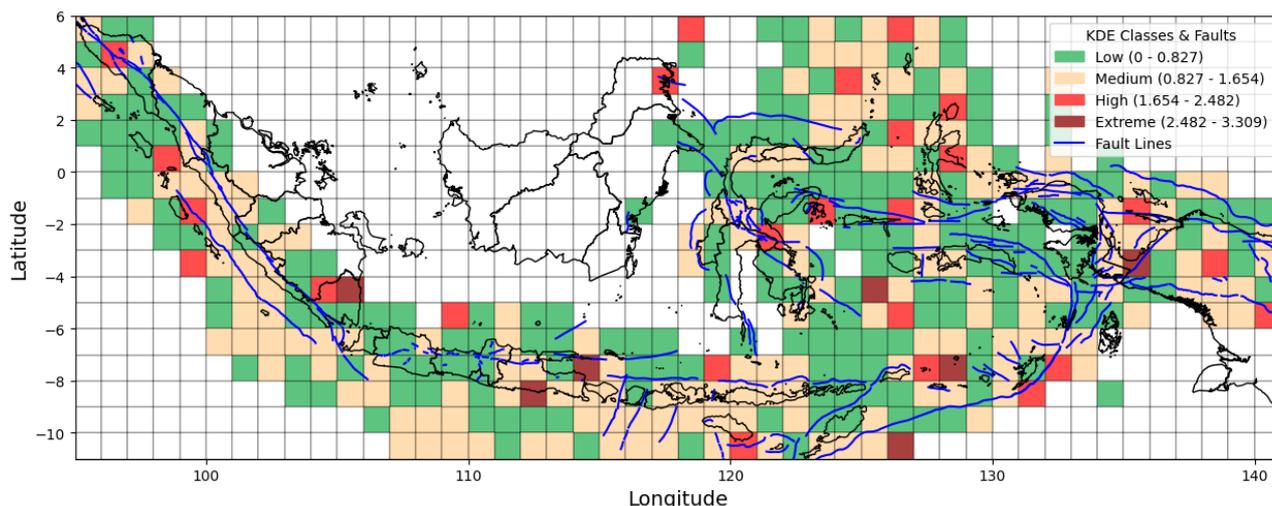


Figure 13 Map showing the spatial distribution of seismic hazard zones in Indonesia, derived from Kernel Density Estimation (KDE). The KDE was computed using a Gaussian kernel with a bandwidth of 0.5.

Table 8 presents the top five earthquake-prone areas identified based on statistical significance using z-scores and ranked by density estimates generated through the Kernel Density Estimation (KDE) method with a Gaussian kernel and specific bandwidth. This analysis highlights areas with both high seismic

activity and statistical relevance, offering a focused view of regions that warrant priority in disaster risk mitigation. By combining density values and z-scores, the table ensures that the selected areas are not only densely affected but also statistically significant regarding earthquake risk.

Table 8 Top five earthquake-prone areas in Indonesia ranked by statistical significance (G_i^* z-scores) and estimated density using Gaussian Kernel Density Estimation (KDE) across varying bandwidth sizes.

Bandwidth Size	Location	Density Value	G_i^* z-score
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
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
0.5	Papua, Indonesia	3.309	0.867
	Banda Sea	3.198	0.656
	Kepulauan Barat Daya, Indonesia	2.750	1.222
	43 km SSW of Dampit, Indonesia	2.703	0.701
	18 km SE of Kotabumi, Indonesia	2.519	0.350

Grid resolution effect

This section will describe the research results on applying the Kernel Density Estimation method based on the grid resolution from the Gaussian kernel function in mapping earthquake-prone areas. **Figure 14 - 17** sequentially shows the results of mapping earthquake-prone areas in Indonesia using the Gaussian kernel function with grid resolutions of $0.25 \times 0.25^\circ$, $0.5^\circ \times 0.5^\circ$, $1^\circ \times 1^\circ$, and $5^\circ \times 5^\circ$.

The variation in grid resolution significantly impacts the detail and interpretation of the mapping results. In **Figure 14**, a grid resolution of $0.25^\circ \times 0.25^\circ$

produces a very detailed mapping, allowing the Identification of small areas with high potential earthquake risk. Furthermore, **Figure 15**, with a grid of $0.5^\circ \times 0.5^\circ$, provides a more transparent and representative picture of the high-risk regions. In **Figure 16**, a grid resolution of $1^\circ \times 1^\circ$ presents a more general risk distribution but reduces the ability to detect significant local variations. Finally, **Figure 17**, with a grid of $5^\circ \times 5^\circ$, provides a broader view but at the risk of obscuring essential details related to earthquake activity in certain areas.

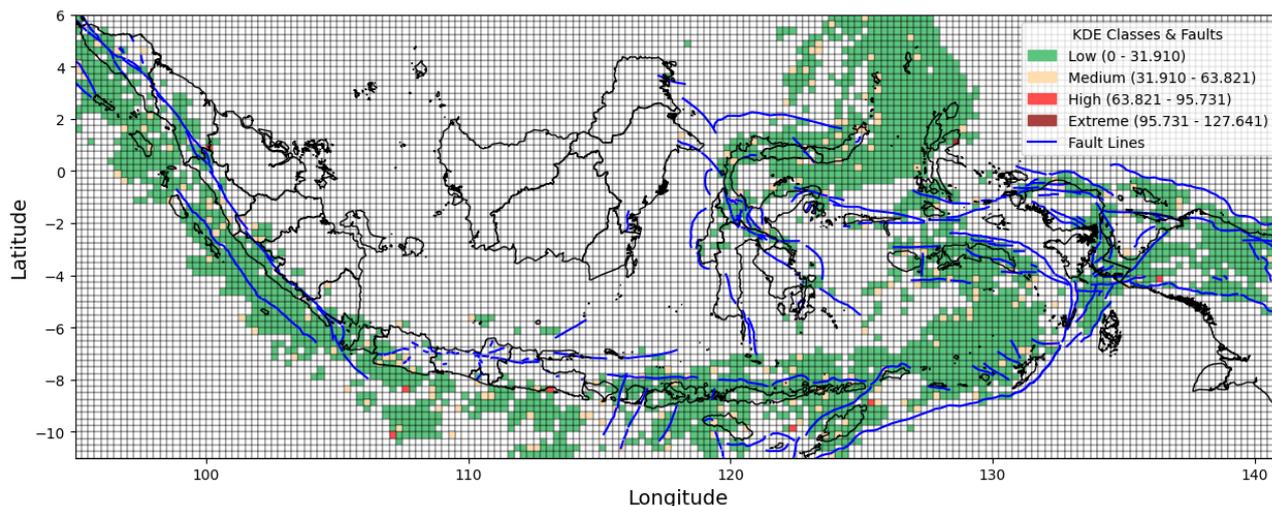


Figure 14 Map showing the spatial distribution of seismic hazard zones in Indonesia, derived from Kernel Density Estimation (KDE) using a Gaussian kernel with a grid resolution of $0.25^\circ \times 0.25^\circ$.

Figure 14 presents the earthquake density map of Indonesia using a $0.25^\circ \times 0.25^\circ$ grid. The map appears highly detailed, showing a distribution of points with varying levels of seismic density. Dark red areas indicate zones with extreme density, while green represents regions with low density. The outlines and geological structures seem to align with the earthquake distribution patterns. With this finer grid resolution, the map reveals small areas with potentially high seismic

activity that may not have been visible on maps with coarser resolution. **Table 9** compares the number of earthquake-prone regions 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 Number of grids in each KDE Class based on different spatial resolutions using the Gaussian kernel with a bandwidth of 0.5

Grid Resolution (Gaussian, BW = 0.5)	KDE Classes			
	Low	Medium	High	Extreme
$0.25^\circ \times 0.25^\circ$	2486	195	7	1
$0.5^\circ \times 0.5^\circ$	855	294	21	3
$1^\circ \times 1^\circ$	155	123	43	6
$5^\circ \times 5^\circ$	20	3	2	4

Figure 15 displays Indonesia's spatial distribution of seismic hazard zones based on KDE using a Gaussian kernel and a $0.5^\circ \times 0.5^\circ$ grid resolution. This map displays a moderate spatial detail, balancing between granularity and clarity. Areas with extreme density are still clearly visible, marked in dark red,

while vast green zones indicate low-density regions. The patterns appear more generalized than the finer grid, yet the high-risk zones remain distinguishable. The visualization helps highlight broader seismic patterns while still maintaining the ability to identify localized risk concentrations.

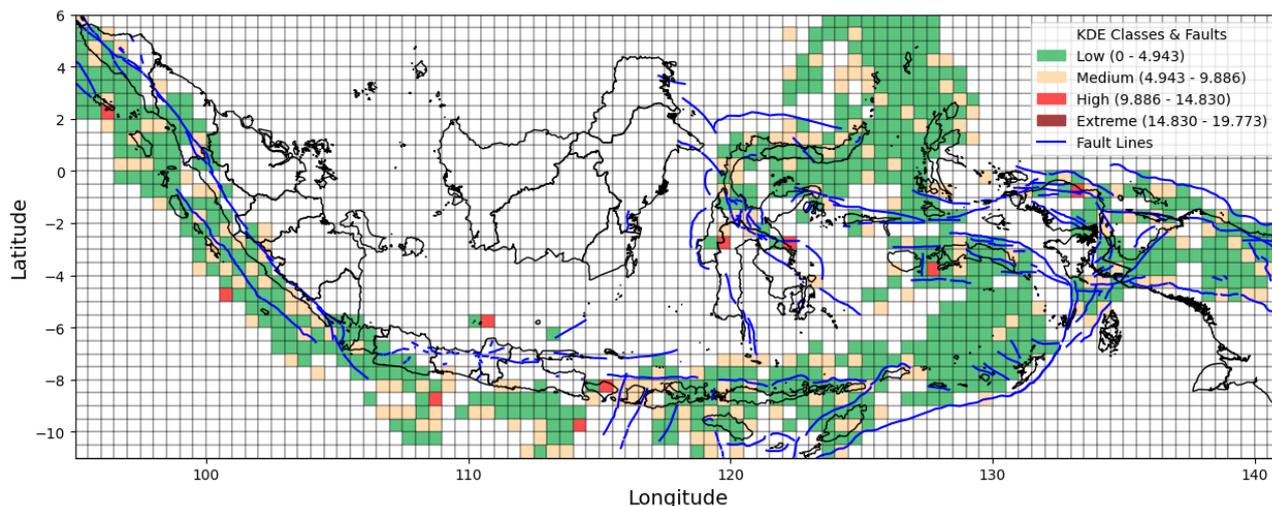


Figure 15 Map showing the spatial distribution of seismic hazard zones in Indonesia, derived from Kernel Density Estimation (KDE) using a Gaussian kernel with a grid resolution of $0.5^\circ \times 0.5^\circ$.

Table 10 presents density value thresholds for each earthquake risk level: low, medium, high, and extreme. These values clarify the intensity and characteristics of seismic risk in different regions. They help researchers interpret spatial patterns and design

targeted interventions. Along with earlier analyses, this table supports more effective earthquake mitigation. It serves as a key reference for risk-based planning in Indonesia.

Table 10 KDE threshold values for classifying density levels using the Gaussian kernel and bandwidth of 0.5 across different grid resolutions.

Grid Resolution (Gaussian, BW = 0.5)	KDE Threshold Value			
	Low	Medium	High	Extreme
$0.25^\circ \times 0.25^\circ$	31.910	63.821	95.731	127.641
$0.5^\circ \times 0.5^\circ$	4.943	9.886	14.829	19.773
$1^\circ \times 1^\circ$	0.827	1.654	2.482	3.309
$5^\circ \times 5^\circ$	0.032	0.064	0.096	0.129

Figure 16 displays Indonesia's spatial distribution of seismic hazard zones based on KDE using a Gaussian kernel and a $1^\circ \times 1^\circ$ grid resolution. With this coarser resolution, the spatial patterns of earthquake density become more generalized, emphasizing regional trends rather than localized variations. Despite the reduced granularity, areas with extreme and high densities remain marked, offering valuable insights into broader seismic activity zones. The color-coded classification helps distinguish different risk levels

across the region, reinforcing the role of this scale in large-area hazard assessments.

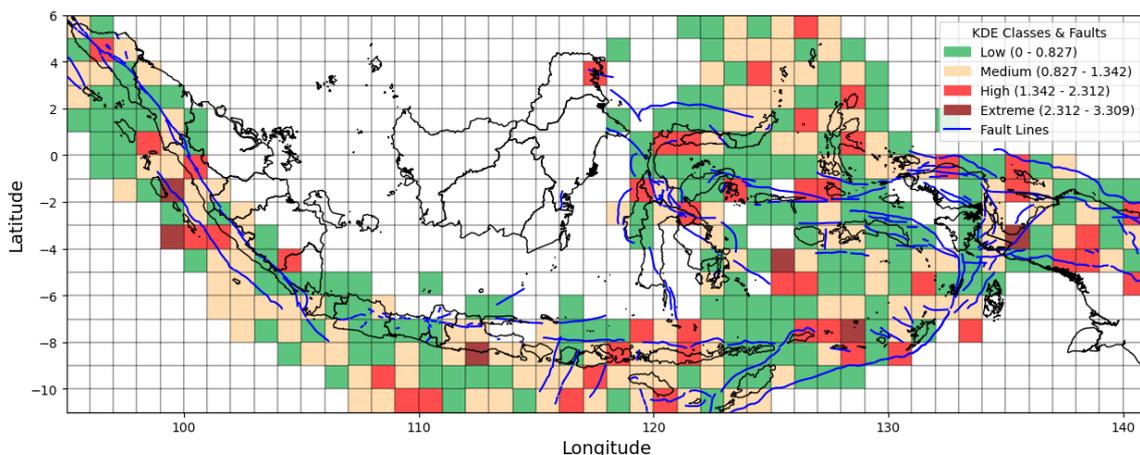


Figure 16 Map showing the spatial distribution of seismic hazard zones in Indonesia, derived from Kernel Density Estimation (KDE) using a Gaussian kernel with a grid resolution of $1^\circ \times 1^\circ$.

Figure 17 displays Indonesia's spatial distribution of seismic hazard zones based on KDE using a Gaussian kernel and a $5^\circ \times 5^\circ$ grid resolution. The representation becomes highly generalized at this coarse resolution, grouping vast geographic regions into singular density values. While this scale reduces

spatial detail, it effectively highlights the most seismically active zones across the country. Areas classified as "extreme" and "high" are still visible, although smaller clusters are no longer distinguishable. This mapping suits broad national-level overviews rather than localized hazard identification.

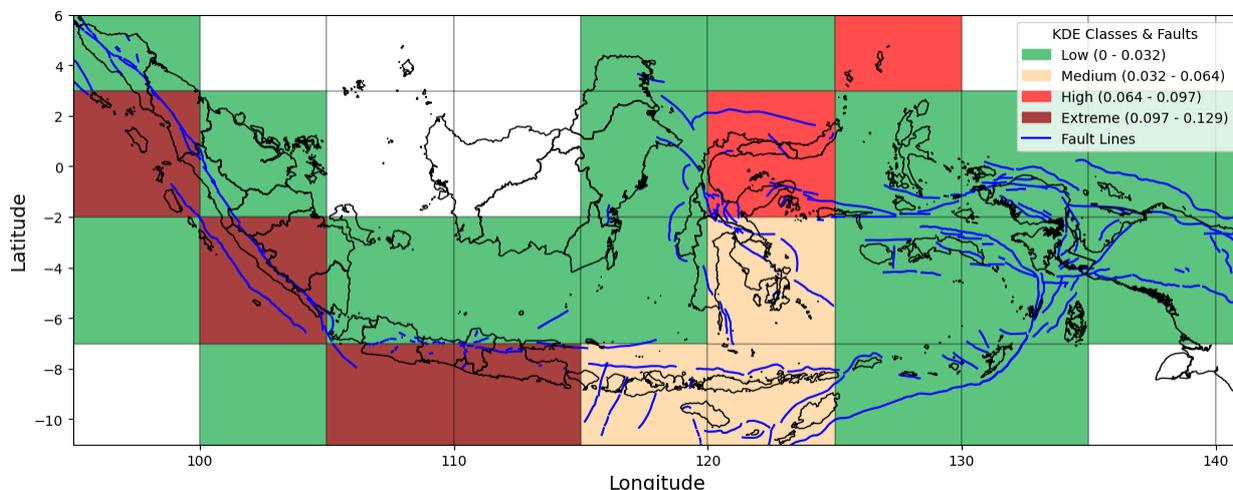


Figure 17 Map showing the spatial distribution of seismic hazard zones in Indonesia, derived from Kernel Density Estimation (KDE) using a Gaussian kernel with a grid resolution of $5^\circ \times 5^\circ$.

Table 11 shows the results of the top five earthquake-prone areas, which were assessed for statistical significance using z-scores and then ranked based on their estimated density values generated through Kernel Density Estimation (KDE) with a Gaussian kernel function and specified bandwidth. This ranking not only provides an overview of the areas most at risk but also emphasizes, through high z-score

values, the importance of giving special attention to these locations in disaster mitigation planning.

The information in **Table 11** enables researchers and policymakers to prioritize resources and interventions in areas most vulnerable to earthquakes to implement preventive measures more effectively. In addition, the **Table 11** results help increase public awareness of earthquake risks in Indonesia and encourage participation in preparation and mitigation

programs. A better understanding of density and risk patterns can guide the development of more responsive

and sustainable policies to address potential earthquake threats in Indonesia.

Table 11 Top five earthquake-prone areas in Indonesia ranked by statistical significance (G_i^* z-scores) and estimated density using Gaussian Kernel Density Estimation (KDE) across varying grid resolutions.

Grid Resolution	Location	Density Value	G_i^* z-score
0.25° × 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° × 0.5°	101 km NNE of Sumenep, Indonesia	19.773	-0.603
	216 km SE of Katabu, Indonesia	14.962	0.170
	215 km SSE of Taliwang, Indonesia	14.356	0.434
	197 km WSW of Bengkulu, Indonesia	13.818	0.353
	33 km WSW of Gambiran Satu, Indonesia	13.327	0.706
1° × 1°	Papua, Indonesia	3.309	0.868
	Banda Sea	3.198	0.656
	Kepulauan Barat Daya, Indonesia	2.750	1.222
	43 km SSW of Dampit, Indonesia	2.703	0.701
	18 km SE of Kotabumi, Indonesia	2.519	0.350
5° × 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

Discussions

Kernel function effect

The Epanechnikov kernel mainly classifies areas as areas with medium earthquake-prone levels (216 grids), followed by high (114 grids), and at the low level, very few (18 grids). 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. Although the biweight kernel tends to produce more areas classified in the low category (251 grids), indicating a lower density distribution in general, 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 kernels 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 density values, with the highest value being 2.056 at 1 km NNW of Tarakan, Indonesia. However, other locations show lower density values and a negative G_i^* z-score for some places, such as the Banda Sea (1.661) and Papua, Indonesia (1.618). This indicates that its statistical relevance is questionable despite the measured density.

The Gaussian kernel, on the other hand, produces higher density values overall, with Papua, Indonesia, recording the highest value of 3.309 and a positive G_i^* z-score (0.868), indicating that this location has significant potential for seismic activity. Other places, such as Banda Sea (3.198) and Kepulauan Barat Daya, Indonesia (2.749), also show good KDE values with G_i^* z-score, indicating better statistical relevance. Ultimately, the biweight kernel function has a reasonably competitive density value, with a maximum value of 2,576 at 1 km NNW of Tarakan, Indonesia. However, the G_i^* z-score values for several locations, such as Papua (1,957), show varying results, with several locations showing negative G_i^* z-score.

These observations highlight that different kernel functions can significantly influence the spatial distribution of density values and their associated G_i^* scores. For instance, the Epanechnikov kernel produces

wider variability in density estimates, often accompanied by downbeat or low G_i^* z-score, even in areas with relatively high-density values such as the Banda Sea and Papua—indicating potential spatial insignificance. In contrast, the Gaussian kernel produces smoother distributions and more consistent G_i^* z-score that align better with high-density values. The biweight kernel, while generating competitive KDE values, shows fluctuations in statistical significance across locations. These discrepancies suggest kernel choice introduces uncertainty in hot spot identification, especially when statistical significance (G_i^*) and density values are misaligned. Therefore, interpreting G_i^* scores must be done cautiously, considering how the kernel function impacts the density surface and the spatial context in which clusters are detected. The variation in results across kernels underscores the sensitivity of G_i^* statistics to parameter selection and the importance of justifying the methodological choices in spatial analysis.

Bandwidth size 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. These results indicate that a small bandwidth (0.1) causes the density estimates to spread widely, resulting in 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 distributing higher values with a larger bandwidth.

The bandwidth size of 0.5 produces the most significant number of grids in the medium category (167 grids), indicating a substantial 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.

The highest density value (10.329) is produced using a bandwidth of 0.1, located 18 km southeast of

Kotabumi, Indonesia. The highest positive G_i^* z-score (1.323) was observed 0 km southeast of Gampengrejo, Indonesia, indicating a spatial cluster with moderate statistical significance, indicating significant clustering. 18 km SE of Kotabumi, Indonesia, has the highest KDE but a negative G_i^* z-score (-0.329), suggesting a spatially insignificant distribution.

Bandwidth with a width of 0.3 produces the highest density location: 130 km NW of Batang, Indonesia (density value = 4.068). The highest positive significance is in Papua, Indonesia (G_i^* z-score = 0.734), indicating moderately significant clustering. The location with a low or negative G_i^* z-score value is 130 km NW of Batang, Indonesia (G_i^* z-score = -0.282), which indicates that the area is spatially insignificant despite having a high density.

When using a bandwidth of 0.5, the highest earthquake density appears in Papua, Indonesia, with a density value of 3.309. The highest positive G_i^* z-score (1.222) is observed in the Southwest Islands, Indonesia, indicating a spatial cluster with moderate significance, indicating a significant cluster with a reasonably high density. Similarly, Kepulauan Barat Daya, Indonesia, exhibits a positive G_i^* z-score of 1.222, indicating a statistically significant cluster associated with relatively high density.

These results highlight how bandwidth parameter variations directly affect the KDE estimates and the statistical significance of spatial clustering (G_i^* z-score). At smaller bandwidths (e.g., 0.1), high-density values may not always correspond to significant hot spots due to the fragmented and localized nature of the density, as seen in Kotabumi, where a negative G_i^* z-score accompanies high density. This inconsistency introduces uncertainty in spatial interpretation, mainly when spatial dependence is weak or inconsistently distributed. As bandwidth increases, the smoothing effect enhances the spatial coherence of high-density areas, resulting in more stable and higher G_i^* z-score but potentially obscuring smaller-scale clusters. These findings emphasize that the G_i^* statistic is highly sensitive to the chosen bandwidth, and its interpretation must consider the underlying spatial configuration and the scale at which density is aggregated. Therefore, selecting bandwidth values should carefully consider both the physical characteristics of seismicity and the

statistical implications to minimize misinterpretation of spatial significance.

Grid resolution effect

The $0.25^\circ \times 0.25^\circ$ grid predominates low density, with only one grid in the extreme category. The $0.5^\circ \times 0.5^\circ$ grid shows an increase in the medium and high categories, reflecting a higher variation in density. The $1^\circ \times 1^\circ$ grid shows some areas with higher density (43 grids in the High category), although there are still grids in the low category. The $5^\circ \times 5^\circ$ grid shows a drastic decrease in grids in the low and medium categories. However, the analysis still identifies some areas with high density in the extreme category. Choosing the correct grid resolution is critical for density analysis, as smaller grid resolutions produce many low-density grids. In comparison, larger grid resolutions allow for more precise identification of high-density areas. The selection of grid resolution is crucial in density analysis to obtain more informative and valuable results.

Very High Density at Small resolution. At the $0.25^\circ \times 0.25^\circ$ grid resolution, there are very high threshold values for all categories, with the highest value reaching 127.641 in the Extreme category. This extremely high value suggests that the small $0.25^\circ \times 0.25^\circ$ grid captures a very concentrated area of seismic activity. Density decreases with larger grid resolution. As the grid resolution increases to $0.5^\circ \times 0.5^\circ$ and $1^\circ \times 1^\circ$, the threshold values for the Low, Medium, High, and Extreme categories decrease significantly. The significant reduction in threshold values across all categories indicates that increasing the grid resolution leads to a sparser distribution of high-density areas.

Very Low Density at the largest size at the $5^\circ \times 5^\circ$ grid resolution, the threshold values for all categories are very low, with a maximum value of 0.129 in the Extreme category. The low threshold values at the $5^\circ \times 5^\circ$ grid resolution suggest that areas of significant density are nearly absent at larger spatial scales. At the $0.25^\circ \times 0.25^\circ$ grid resolution, the location with the highest density is 100 km SE of Tobelo, Indonesia (density value = 127.641). The highest positive significance is south of Java, Indonesia (G_i^* z-score = 1.212), indicating a significant cluster with high density. The following 21 km WSW of Kencong,

Indonesia, has a negative G_i^* z-score (-0.155), indicating no significant clusters despite a reasonably high density.

At a grid resolution of $0.5^\circ \times 0.5^\circ$, the highest density is observed 101 km NNE of Sumenep, Indonesia, with a density value of 19.773. The highest positive G_i^* z-score (0.706) is found 33 km west-southwest of Gambiran Satu, Indonesia, indicating a weak but relatively higher spatial clustering, indicating significant clusters despite lower density than the $0.25^\circ \times 0.25^\circ$ grid. 101 km NNE of Sumenep, Indonesia, has a negative G_i^* z-score (-0.603), indicating an insignificant distribution despite having the highest density in this grid resolution.

Grid resolution $1^\circ \times 1^\circ$. The region of Papua, Indonesia, exhibits the highest observed density (3.309), suggesting a prominent cluster of seismic events. The highest positive G_i^* z-score (1.222) is observed in the Southwest Islands, Indonesia, indicating a statistically significant high cluster. 43 km SSW of Dampit, Indonesia, has a positive G_i^* z-score (0.701), suggesting a substantial distribution with a relatively high density. The highest seismic density was observed in the $5^\circ \times 5^\circ$ grid cell situated southwest of Sumatra, Indonesia, with a recorded density value of 0.129. No locations show highly significant G_i^* z-score values. The highest G_i^* z-score (0.497) is observed in Southern Sumatra, Indonesia, which still represents a relatively weak spatial cluster. Indonesia's Nias region exhibits a negative G_i^* z-score (-0.052), indicating a statistically non-significant spatial distribution.

These findings underscore how the grid resolution selection directly impacts the statistical behavior of G_i^* z-scores. At finer resolutions (e.g., $0.25^\circ \times 0.25^\circ$), specific locations exhibit extremely high density values yet relatively low or even negative G_i^* z-score, such as in Kencong and Sumenep. This pattern suggests that high local density does not always translate to statistically significant hot spots, mainly when spatial variability is high, and neighboring grids show low density. Conversely, larger grid resolution reduces noise and stabilizes G_i^* values, but at the cost of spatial precision. The varying results across grid resolution reflect the sensitivity of hot spot detection to spatial aggregation and spatial weights, introducing

uncertainty in statistical interpretation. Therefore, when interpreting G_i^* z-score, it is essential to consider these limitations and the influence of parameter settings to avoid misclassification of risk areas and maintain spatial inference's integrity.

Data duality and potential biases in the earthquake catalog

The accuracy and reliability of seismic hazard analysis are strongly dependent on the quality and completeness of the underlying earthquake database. While the government agency (BMKG) maintains a comprehensive record of seismic events in Indonesia, several potential biases may influence spatial density analysis.

First, temporal discrepancies may occur due to technological advancements in earthquake detection. Small-magnitude earthquakes may have gone unrecorded earlier, whereas modern instruments can now capture them. This difference in detection sensitivity over time can lead to temporal bias, especially when applying Kernel Density Estimation (KDE) to analyze long-term spatial distributions.

Second, spatial reporting errors may occur due to the uneven distribution of seismic stations across the Indonesian archipelago. Even when actual seismic activity is similar, regions with a higher density of seismic stations, such as Java or Sumatra, are likely to report more earthquake events. This unequal observational capacity introduces spatial bias, potentially producing artificially high-density values in better-monitored regions.

Third, magnitude thresholding and event filtering used during data preprocessing, such as deleting duplicates or excluding aftershocks, can affect KDE outcomes. For example, although methodologically reasonable, removing aftershock clusters may result in underestimating local seismicity in high-risk zones.

Finally, uncertainties in earthquake epicenter locations may lead to spatial noise, especially in offshore regions where precise localization is inherently more complex. While these data limitations do not compromise the validity of the results, they underscore the need for a cautious interpretation of the resulting KDE-based density maps. Future research may benefit from combining catalog completeness analysis, station density modifications, or uncertainty

modeling to improve the robustness of geographic seismic hazard estimations.

Comparison with other methods

Previous studies on earthquake-prone area classification have predominantly utilized clustering algorithms such as K-Means, DBSCAN, and ST-DBSCAN to group seismic events based on magnitude, depth, and spatiotemporal characteristics. These clustering methods effectively identify discrete seismic patterns and are commonly applied in exploratory analyses and regional zoning. However, they inherently produce categorical outputs (i.e., clusters) rather than continuous hazard surfaces, limiting their capacity to represent spatial gradients of seismic risk.

In contrast, Kernel Density Estimation (KDE) offers a probabilistic and spatially continuous approach to representing earthquake occurrence. By generating smooth density surfaces or heatmaps, KDE effectively visualizes seismic hazard intensity, making it particularly useful for risk communication, spatial planning, and delineation of high-risk zones. This study compares KDE and 3 widely adopted clustering techniques, K-Means, DBSCAN, and ST-DBSCAN, to highlight their strengths and limitations in seismic

hazard mapping. While KDE presents a continuous depiction of event density, clustering methods partition seismic events into discrete groups based on spatial and/or temporal proximity. Each approach demonstrates unique capabilities, summarized in **Table 12**.

The comparative analysis in **Table 12** enhances our understanding of how these techniques function within the context of seismic hazard mapping. KDE excels at producing spatially continuous, interpretable heat maps that highlight zones of high seismic intensity. Conversely, K-Means and DBSCAN are better suited for detecting discrete seismic regions, with ST-DBSCAN extending this functionality by integrating temporal dynamics. Ultimately, selecting an appropriate method depends on the objectives of the analysis, the dataset’s characteristics, and the desired output format. Clustering methods are advantageous for pattern recognition and classification, whereas KDE provides a gradient-based visualization of seismic risk. This study suggests that integrating KDE with clustering approaches may offer a more robust and nuanced representation of seismic hazards, capturing density patterns and discrete event groupings.

Table 12 Comparison of kernel density estimation and clustering methods in seismic hazard mapping.

Feature	KDE (Kernel Density Estimation)	K-Means Clustering	DBSCAN	ST-DBSCAN
Spatial Continuity	It provides a continuous density surface, effectively visualizing event concentration as smooth heatmaps [32].	It produces discrete cluster centers without inherent spatial continuity [43].	Identifies clusters based on density, allowing for contiguous regions depending on parameter settings [44].	ST-DBSCAN to incorporate temporal data for identifying spatiotemporal clusters [45].
Sparse Data Handling	May oversmooth areas with sparse data, potentially underrepresenting isolated events [32].	Tends to form clusters of similar size, which may not accurately reflect areas with sparse seismic activity [43].	Effectively detects clusters in dense regions and identifies noise in sparse areas, suitable for varied densities [46].	Capable of detecting clusters in both spatial and temporal dimensions, even in datasets with varying densities [47].

Feature	KDE (Kernel Density Estimation)	K-Means Clustering	DBSCAN	ST-DBSCAN
Output Type	It generates continuous heatmaps representing event density [32].	It produces discrete, non-overlapping clusters [43].	It produces discrete clusters based on density connectivity. [44].	It produces discrete clusters considering spatial proximity and temporal occurrence [45].
Parameter Sensitivity	Highly sensitive to bandwidth selection; inappropriate bandwidth can lead to over- or under-smoothing [25].	Requires prior specification of the number of clusters (k), which can be challenging without domain knowledge [43].	Sensitive to parameters like epsilon (ϵ) and minimum points (minPts); improper settings can lead to poor results [44].	The method involves multiple parameters (spatial radius, temporal window, and minPts), and each must be carefully tuned to ensure optimal clustering performance [45].
Suitability for Seismic Mapping	It is well-suited for identifying and visualizing high seismic risk zones through continuous density estimation [25].	Limited applicability in seismic hazard mapping due to the assumption of spherical clusters and equal sizes [43].	It is effective for detecting irregularly shaped clusters of seismic events and is helpful for seismic patterns [46].	It is beneficial for analyzing seismic sequences where both space and time are critical [47].

Conclusions

This study examines the role of kernel function, bandwidth, and grid resolution 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 highest concentration of risk in the medium category, whereas the Gaussian kernel 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 hot spots.

Furthermore, bandwidth significantly impacts risk mapping resolution and data interpretation quality. A small bandwidth (0.1) enables high spatial resolution but results in a granular and fragmented distribution that is difficult to discern at large scales. A broader 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 the best since it provided a balance of local detail and global coverage, allowing for more consistent detection of different degrees of earthquake risk.

Grid resolution is key in determining spatial resolution and map accuracy. Small grids ($0.25^\circ \times 0.25^\circ$) provide the best resolution, but the resulting data is sometimes too detailed for regional study. A large grid ($5^\circ \times 5^\circ$) produces excessively coarse data aggregation, which is less meaningful. The $1^\circ \times 1^\circ$ grid strikes the ideal mix between resolution and regional coverage, allowing for high-accuracy Identification of danger hot spots while preserving the larger spatial context. The main contribution of this study lies in systematically evaluating the interplay between kernel functions, bandwidth sizes, and grid resolutions, resulting in the recommendation of a Gaussian kernel, 0.5 bandwidth size, and $1^\circ \times 1^\circ$ grid as

the optimal combination for comprehensively mapping seismic risk in Indonesia.

However, this study has certain limitations. It relies solely on historical earthquake data without incorporating additional geophysical or geological variables, such as fault slip rates, soil types, or tectonic stress fields, which could improve the robustness of the risk assessment. Moreover, using a static bandwidth in Kernel Density Estimation (KDE) may not fully capture the spatial heterogeneity of seismic patterns, especially in highly complex or evolving tectonic regions. Additionally, the smoothing inherent in KDE may underrepresent low-frequency seismic events or isolated earthquakes in sparse-data regions, potentially leading to underestimated risk in these areas. Future research could explore adaptive or data-driven KDE approaches to address these limitations more effectively.

Future research aiming to address these limitations may explore advanced methodologies to improve the accuracy of earthquake event density analysis. One such approach is the Local Outlier Factor (LOF) method, which offers significant advantages in detecting extreme and rare seismic events, particularly in risk assessment and hazard mapping. Unlike KDE, which relies on a fixed bandwidth parameter to smooth spatial distributions, LOF dynamically evaluates the local density of seismic occurrences relative to their neighboring events. This characteristic makes LOF particularly effective in identifying localized anomalies that deviate significantly from the expected spatial pattern, thus providing a more robust mechanism for detecting potential seismic precursors or hidden fault activity.

Furthermore, the assumption of a continuous spatial distribution inherent in Kernel Density Estimation (KDE) may obscure critical variations in seismicity, especially in tectonically active regions characterized by substantial spatial heterogeneity in earthquake occurrences. In contrast, the Local Outlier Factor (LOF) method assesses the degree of isolation of each event relative to its local neighborhood, offering a more adaptive and context-sensitive approach to identifying risk-prone areas. This localized analytical framework enhances LOF's utility in discriminating genuine seismic clusters from statistical noise, ultimately improving the robustness and

reliability of seismic risk prediction models. Integrating LOF into earthquake hazard assessment allows researchers to refine risk mapping, capturing spatial variations in seismicity with greater accuracy.

A hybrid approach combining LOF and KDE is also promising, where KDE provides a global density estimation. At the same time, LOF enhances local anomaly detection, ultimately improving the spatial resolution of earthquake forecasting models. Furthermore, future studies may integrate KDE with neural network-based approaches, particularly those grounded in stability analysis frameworks, to build more adaptive and data-driven models for detecting earthquake-prone areas.

Moreover, the k-nearest neighbors (k-NN) density estimation method provides another promising alternative. Unlike KDE and LOF, k-NN dynamically adjusts density estimation based on the number of nearest neighbors, making it highly effective in analyzing non-uniformly distributed seismic data. This method effectively handles complex spatial variations and enhances real-time seismic risk prediction and mitigation when integrated into machine learning or AI-based systems.

The findings of this study offer practical implications for seismic risk mitigation in Indonesia, providing a methodological reference for national disaster agencies in designing early warning systems and spatial planning strategies. Future implementations may utilize the proposed KDE-LOF hybrid model in real-time monitoring systems to improve responsiveness to emerging seismic threats, especially in tectonically complex regions.

Acknowledgments

This work is supported by the Indonesian Education Scholarship (BPI), under the Center for Higher Education Funding and Assessment, Ministry of Higher Education, Science, and Technology of the Republic of Indonesia (PPAPT), and the Indonesia Endowment Fund for Education (LPDP), as acknowledged in Decree No. 02462/BPPT/BPI.06/9/2024

Declaration of Generative AI in Scientific Writing

The authors employed ChatGPT to assist in grammar correction and language refinement during the manuscript preparation. The authors thoroughly reviewed and edited the final content; we take full responsibility for its accuracy.

CRedit author statement

Ari Fadli: Writing – original draft, Conceptualization, Methodology, Software, Data curation, Visualization. Tri Kuntoro Priyambodo: Supervision, Writing – review & editing, Methodology. Agfianto Eko Putra: Supervision, Writing – review & editing, Methodology. Wiwit Suryanto: Supervision, Writing – review & editing, Methodology, Validation, Formal analysis.

References

- [1] WL Hakim and LC Wook. A review on remote sensing and GIS applications to monitor natural disasters in Indonesia. *Korean Journal of Remote Sensing* 2020; **36(6)**, 1303-1322.
- [2] TH Siagian, P Purnadi, S Suhartono and H Ritonga. Social vulnerability to natural hazards in Indonesia: Driving factors and policy implications. *Natural Hazards* 2014; **70(2)**, 1603-1617.
- [3] MN Malawani, F Lavigne, C Gomez, BW Mutaqin and DS Hadmoko. Review of local and global impacts of volcanic eruptions and disaster management practices: The Indonesian example. *Geosciences* 2021; **11(3)**, 109.
- [4] D Kurniati, MZ Fauzi, R Ripangi, A Falegas and I Indria. Klasterisasi daerah rawan gempa bumi di Indonesia menggunakan algoritma k-medoids. *MALCOM: Indonesian Journal of Machine Learning and Computer Science* 2021; **1(1)**, 47-57.
- [5] W Achmad. The effectiveness of earthquake disaster management policy in Indonesia. *Ganaya: Jurnal Ilmu Sosial dan Humaniora* 2023; **6(2)**, 367-377.
- [6] L Srikanth and I Srikanth. *A case study on kernel density estimation and hotspot analysis methods in traffic safety management*. IEEE, New York, 2020.
- [7] S Chainey, L Tompson and S Uhlig. The utility of hotspot mapping for predicting spatial patterns of crime. *Security Journal* 2008; **21(1-2)**, 4-28.
- [8] D Baranyai and T Sipos. Black-spot analysis in Hungary based on kernel density estimation. *Sustainability* 2022; **14(14)**, 8335.
- [9] N Boonsatit, G Rajchakit, R Sriraman, CP Lim and P Agarwal. Finite-/fixed-time synchronization of delayed Clifford-valued recurrent neural networks. *Advances in Difference Equations* 2021; **2021**, 276.
- [10] G Rajchakit, R Sriraman, N Boonsatit, P Hammachukiattikul, CP Lim and P Agarwal. Exponential stability in the Lagrange sense for Clifford-valued recurrent neural networks with time delays. *Advances in Difference Equations* 2021; **2021**, 256.
- [11] G Rajchakit, R Sriraman, N Boonsatit, P Hammachukiattikul, CP Lim and P Agarwal. Global exponential stability of Clifford-valued neural networks with time-varying delays and impulsive effects. *Advances in Difference Equations* 2021; **2021**, 208.
- [12] G Rajchakit, P Agarwal and S Ramalingam. *Stability Analysis of Neural Networks*. Springer Singapore, Singapore, 2021.
- [13] F Hisyam, A Susilo, M Anshori and MFR Hasan. Spatio-temporal variation seismicity pattern in east Java between 2002 and 2022 based on the b-value and seismic quiescence z-value. *Trends in Sciences* 2024; **21(4)**, 7608.
- [14] M Bil, R Andrasik, T Svoboda and J Sedonik. The KDE+ software: A tool for effective identification and ranking of animal-vehicle collision hotspots along networks. *Landscape Ecology* 2016; **31(2)**, 231-237.
- [15] CA Blazquez and MS Celis. A spatial and temporal analysis of child pedestrian crashes in Santiago, Chile. *Accident Analysis & Prevention* 2013; **50**, 304-311.
- [16] SS Pulgurtha, VK Krishnakumar and SS Nambisan. New methods to identify and rank high pedestrian crash zones: An illustration. *Accident Analysis & Prevention* 2007; **39(4)**, 800-811.
- [17] AL Achu, CD Aju, V Suresh, TP Manoharan and R Reghunath. 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* 2019; **69(4)**, 255-265.
- [18] S Hashimoto, S Yoshiki, R Saeki, Y Mimura, R Ando and S Nanba Development and application of traffic accident density estimation models using kernel density estimation. *Journal of Traffic and Transportation Engineering (English Edition)* 2016; **3(3)**, 262-270.
- [19] TK Anderson. Kernel density estimation and K-means clustering to profile road accident hotspots. *Accident Analysis & Prevention* 2009; **41(3)**, 359-364.
- [20] IM Gelfand, SI Guberman, ML Izvekova, VI Keilis-Borok and EJ Ranzman. Criteria of high seismicity, determined by pattern recognition. *Tectonophysics* 1972; **13(1-4)**, 415-422.
- [21] AD Gvishiani, AA Soloviev and BA Dzeboev. Problem of recognition of strong-earthquake-prone areas: A State-of-the-art review. *Izvestiya, Physics of the Solid Earth* 2020; **56(1)**, 1-23.
- [22] AA Soloviev, AD Gvishiani, AI Gorshkov, MN Dobrovolsky and OV Novikova. Recognition of earthquake-prone areas: Methodology and analysis of the results. *Izvestiya, Physics of the Solid Earth* 2014; **50(2)**, 151-168.
- [23] A Sharma, A Ahuja, S Devi and S Pasari. *Use of spatio-temporal features for earthquake forecasting of imbalanced data*. IEEE, New York, 2022.
- [24] M Yousefzadeh, SA Hosseini and M Farnaghi. Spatiotemporally explicit earthquake prediction using deep neural network. *Soil Dynamics and Earthquake Engineering* 2021; **144**, 106663.
- [25] M Danese, M Lazzari and B Murgante. *Kernel density estimation methods for a geostatistical approach in seismic risk analysis: The case study of potenza hilltop town (southern Italy)*. Springer Berlin Heidelberg, Berlin, Germany, 2008.
- [26] DA Luna, FM Alonso-Chaves and C Fernandez. Kernel density estimation for the interpretation of seismic big data in tectonics using QGIS: The Turkiye–Syria earthquakes (2023). *Remote Sensing* 2024; **16(20)**, 3849.
- [27] IH Rifa, H Pratiwi and R Respatiwan. clustering of earthquake risk in indonesia using k-medoids and k-means algorithms. *Media Statistika* 2020; **13(2)**, 194-205.
- [28] FR Senduk, I Indwiarti and F Nhita. Clustering of earthquake prone areas in indonesia using k-medoids algorithm. *Indonesia Journal on Computing (Indo-JC)* 2020; **4(3)**, 65-76.
- [29] A Jufriansah, Y Pramudya, A Khusnani and S Saputra. Analysis of earthquake activity in Indonesia by clustering method. *Journal of Physics: Theories and Applications*. 2021; **5(2)**, 92.
- [30] A Prasetyo, MM Effendi and MN Dwi. Analisis gempa bumi di Indonesia dengan metode clustering. *Bulletin of Information Technology (BIT)* 2023; **4(3)**, 338-343.
- [31] E Barlian, J Tarigan, Z Nasution and A Purwoko. Identification of earthquake-prone areas as a regional planning study for deli serdang regency. *Journal of Physics: Conference Series* 2022; **2193**, 012015.
- [32] AR Samsudin, DH Fudholi and L Iswari. Temporal spatial property profiling and identification of earthquake prone areas using st-dbscan and k-means clustering. *Jurnal Teknik Informatika* 2024; **5(3)**, 917-929.
- [33] BW Silverman. *Density estimation for statistics and data analysis*. Springer Netherlands, Dordrecht, Netherlands, 1998.
- [34] X Jin and J Kawczak. Birnbaum-saunders and lognormal kernel estimators for modelling durations in high frequency financial data. *Annals of Economics and Finance* 2003; **4(1)**, 103-124.
- [35] S Weglarczyk. Kernel density estimation and its application. *ITM Web of Conferences* 2018; **23**, 00037.
- [36] T Hart and P Zandbergen. Kernel density estimation and hotspot mapping. *Policing: An International Journal of Police Strategies & Management* 2014; **37(2)**, 305-323.
- [37] S Chainey and J Ratcliffe. *GIS and Crime Mapping*. Wiley, New Jersey, 2005.
- [38] N Levine. *CrimeStat: A spatial statistical program for the analysis of crime incidents*. Springer Natrue, New York, 2017.
- [39] TC Bailey, AC Gatrell and IK Crain. Interactive spatial data analysis. *Cartographica*, 1997; **34(1)**, 69.

- [40] M Mathur. Spatial autocorrelation analysis in plant population: An overview. *Journal of Applied and Natural Science* 2015; **7(1)**, 501-513.
- [41] A Getis and JK Ord. The analysis of spatial association by use of distance statistics. *Geographical Analysis* 1992; **24(3)**, 189-206.
- [42] L Anselin. Local indicators of spatial association—LISA. *Geographical Analysis* 1995; **27(2)**, 93-115.
- [43] P Songchitruksa and X Zeng. Getis–ord spatial statistics to identify hot spots by using incident management data. *Transportation Research Record: Journal of the Transportation Research Board* 2010; **2165(1)**, 42-51.
- [44] AK Jain. Data clustering: 50 years beyond K-means. *Pattern Recognition Letters* 2010; **31(8)**, 651-666.
- [45] M Ester, H Kriegel, J Sander and X Xu. *A density-based algorithm for discovering clusters in large spatial databases with noise*. AAAI Press, Woshington DC, 1996.
- [46] D Birant and A Kut. ST-DBSCAN: An algorithm for clustering spatial–temporal data. *Data & Knowledge Engineering* 2007; **60(1)**, 208-221.
- [47] I Zaliapin and Y Ben-Zion. Earthquake clusters in southern California I: Identification and stability. *Journal of Geophysical Research: Solid Earth* 2013; **118(6)**, 2847-2864.
- [48] Y Wen, C Chen, S Wen and W Lu. Spatiotemporal seismicity pattern of the Taiwan orogen. *Natural Hazards and Earth System Sciences* 2023; **23(5)**, 1835-1846.