EKSAKTA
Journal of Sciences and Data Analysis

E-ISSN: 2720-9326
P-ISSN: 2716-0459
DOI : 10.20885/EKSAKTA.vol5.iss2.art9

Comparison of Inverse Distance Weighted and Thin Plate Spline Interpolation Methods in Projecting the Strength of the West Sumatra Earthquake

Nabila Azzahra Haris Putri[1] , **Achmad Fauzan**[1,*]

[1] Statistics Department, Faculty of Mathematics and Natural Sciences Universitas Islam Indonesia

**Corresponding author*: achmadfauzan@uii.ac.id

Received: 26 August 2024; Accepted: 29 October 2024; Published: 31 October 2024

Abstract: The Indonesian archipelago is situated in a highly active geological zone, making it prone to frequent earthquakes. West Sumatra, located on the west coast of central Sumatra, comprises lowland coastal areas and volcanic plateaus formed by the Barisan Mountains, covering a land area of 42,297.30 km² (2.17% of Indonesia's territory). This research aims to determine which interpolation method—Inverse Distance Weighted (IDW) and Thin Plate Spline (TPS)—provides more accurate predictions of earthquake strength in West Sumatra. The dataset consists of 229 earthquake events, divided into 90% for training (206 points) and 10% for testing (23 points). The training data was further subdivided into 80% training data 2 (164 points) and 20% validation data (42 points). The interpolation processes using the IDW and TPS methods were repeated 100 times, with the training 2 and validation data randomly shuffled in each iteration. Visualization of the interpolation results indicated that the earthquake magnitudes ranged from 2.0 to 4.5. Although the Mean Absolute Percentage Error (MAPE) values for the TPS method on the test and validation datasets were 16.42 and 14.29, respectively—slightly lower than the MAPE values for the IDW method—the t-test results showed no statistically significant difference between the two methods. Statistically, there is no significant difference between IDW and TPS in terms of predictive accuracy. However, researchers prefer the IDW method due to its computational efficiency and simplicity. Therefore, IDW is considered the most suitable method for analyzing earthquake strength in the West Sumatra region.

Keywords: Earthquake, Interpolation, Inverse Distance Weighted, Thin Plate Spline

Introduction

Indonesia frequently experiences earthquakes due to its location within a highly active geological region. West Sumatra, located on the western coast of Sumatra Island, is one of the most earthquake-prone areas [1]. Earthquakes in this region not only cause significant damage to buildings but can also trigger secondary disasters such as tsunamis and landslides. West Sumatra is located on the west coast of the Central Sumatra island, consisting of lowlands on the west coast and volcanic highlands formed by the Barisan mountains. The province covers an area of 42,297.30 km², equivalent to 2.17 percent of Indonesia's territory[2].

Earthquake strength measurements in West Sumatra are often uneven and face some limitations in certain locations [3]. To overcome this problem, interpolation techniques can be used to estimate earthquake strength in areas that are not well illustrated. By using interpolation, we can help identify spatial patterns of earthquake strength and provide a more complete picture of earthquake distribution in the region [4].

Spatial interpolation is the process of estimating values at locations between observation points, a method used to fill data gaps or create more detailed maps based on points in a geographic area [5]. Spatial interpolation is a mathematical technique used to estimate values at locations where there is no observation data, by referring to the values in the surrounding area. This spatial interpolation approach is based on the assumption that data attributes have continuous characteristics in space, and that these attributes have spatial relationships or relationships that can be used to estimate data at locations where there is no direct observation [6].

Consequently, it is essential to predict earthquake strength in areas where data is sparse. Interpolation methods are commonly employed to estimate earthquake intensity in unmeasured areas [7]. Among these methods, Inverse Distance Weighted (IDW) and Thin Plate Spline (TPS) are widely used. IDW estimates values between data points by assigning greater weight to those that are closer [8], whereas TPS employs a more complex mathematical approach to generate smoother and potentially more accurate maps of earthquake strength [9].

Each of these methods has its own advantages and disadvantages. The Inverse Distance Weighted (IDW) method is straight forward to use and effective for unevenly distributed data [10]. However, it may be less accurate when dealing with data that exhibits highly complex variations [11]. Conversely, the Thin Plate Spline (TPS) method is better suited to handling such complexities, though its application is more computationally intensive[10]. By utilizing these interpolation methods, a more accurate representation of earthquake strength in West Sumatra can be achieved [2]. This enhanced accuracy is crucial for aiding both the public and government authorities in implementing preventive measures and disaster mitigation strategies. This method uses a polynomial-based spline function to connect known data points in a smooth and continuous manner. Thin Plate Splines are used in a variety of applications, including image processing, spatial data interpolation, and topographic surface modeling [9].

Materials

The study focuses on earthquake pressure data from Padang, West Sumatra, within the coordinates -2.570° South Latitude to -0.198° North Latitude and 97.954° West Longitude to 102.502° East Longitude. The research sample consists of earthquake pressure data within these coordinates, collected from January 1, 2023, to May 10, 2024. This study was conducted using Microsoft Excel, RStudio, and QGIS software for data analysis and visualization. The resulting map visualization is shown in **Figure 1**.

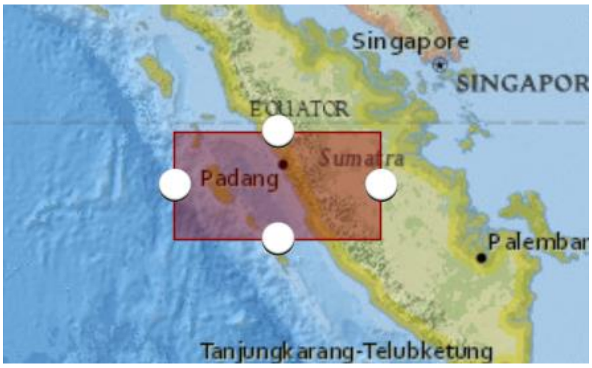


Figure 1. West Sumatra Map (source: BMKG)

Table 1. Earthquake data in West Sumatra

No	Date	Time	X (mE)	Y (mS)	UTM Zone	Magnitude (Ritchter)
1	07/01/2023	01.52.35	633506.3	9857716.1	47S	2.9
2	09/01/2023	17:17:01	648340.3	9962298.9	47S	2.6
⋮	⋮	⋮	⋮	⋮	⋮	⋮
228	12/05/2024	09:47:09	627414.9	9952794.2	47S	2.8
229	13/05/2024	11:01:38	600021.6	9876964.7	47S	3.2

The variables used in this study have been expanded to include earthquake magnitude, earthquake depth, x-coordinate (longitude), y-coordinate (latitude), and time of each event. Earthquake magnitude indicates the energy released during an earthquake, while earthquake depth indicates the distance below the Earth's surface where the earthquake originated. The x- and y-coordinates are expressed in the Universal Transverse Mercator (UTM) system, with the longitude coordinate measured in meters East (mE) and the latitude coordinate in meters South (mS)[13]. These coordinates are important for determining the geographic position of the observation points in a Geographic Information System (GIS). In addition, the timestamp provides the exact time of occurrence for each earthquake, which is valuable for analyzing temporal patterns[14]. By including the depth and timestamp of earthquakes, this study obtains a more comprehensive dataset that aids in the spatial and temporal analysis of earthquake characteristics, improving the accuracy and insight of the interpolation methods used[15].

The coordinate system is a very important geographic coordinate system that uses latitude and longitude to indicate points on the earth's surface [4]. Latitude is a measure of the distance north or south between the equator and the North Pole or South Pole. North Latitude is stated as positive, while South Latitude is stated as negative. The longest latitude is 0 degrees at the equator, the greatest latitude is 90 degrees at the North Pole and South Pole [16]. Longitude is a measure of the distance east or west of the prime reference meridian, known as the Greenwich Prime Meridian. The Greenwich Prime Meridian is symbolized by 0 degrees, the longitude in the East is called East Longitude and in the West is called West Longitude. The stars reach a maximum angle of 180 degrees in both directions and meet at the International Date Line [17].

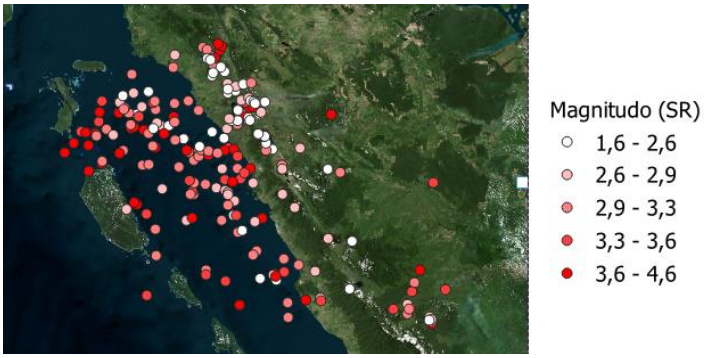


Figure 2. Visualization Distribution Occurrence Points West Sumatra (source: data processed).

Methods

After collecting the data on earthquakes in West Sumatra, the next step is data processing. Figure 3 illustrates the research flow diagram used in this study.

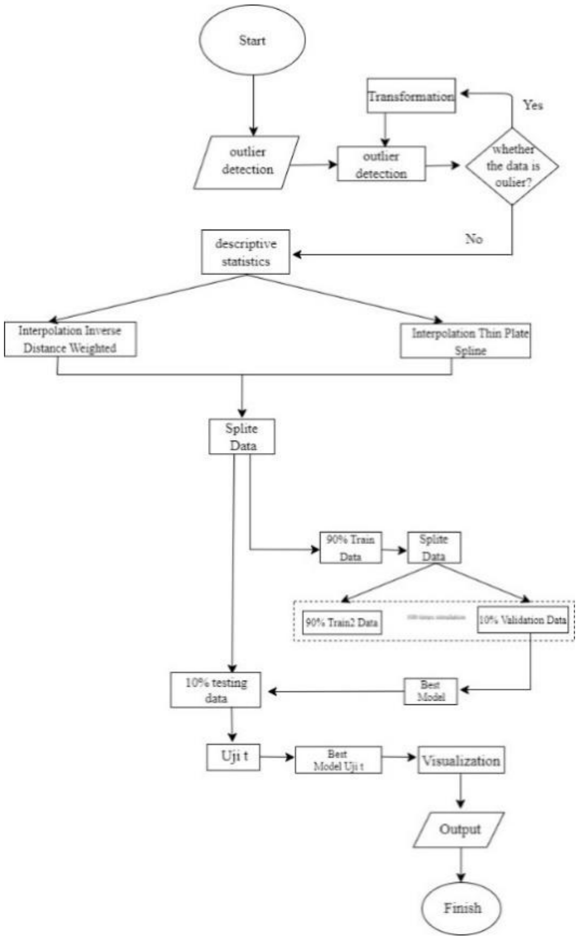


Figure 3. Flow Diagram

The research began with a literature review to identify relevant topics. Once the data was collected, outliers were checked using a predefined equation. If no outliers were detected, the research proceeded to the next stage. A descriptive analysis was then performed to provide a general overview of the data. The dataset, consisting of 229 observations, was split into two proportions: 90% for training data (206 observations) and 10% for testing data (23 observations). The training data was further divided, with 80% used as training data 2 (164 observations) and 20% as validation data (42 observations).

Spatial interpolation was performed using the IDW method to predict earthquake strength, with 100 simulations conducted to calculate the Mean Absolute Percentage Error (MAPE) by varying the training 2 and validation data. This approach allowed for variations in MAPE results across simulations. The second interpolation method applied was the TPS, which also involved 100 MAPE simulations for result variation. A t-test was then conducted to determine the superior method between IDW and TPS. Finally, the interpolation results were visualized as a simple map using QGIS software, based on the most accurate method identified. Inverse Distance Weighted (IDW) interpolation is a local interpolation technique where the estimated value of a point is influenced by its neighboring points [18]. In this method, points closer to the target location have a greater influence, or weight, on the estimated value, while this weight decreases as the distance from the target point increases. The weight, typically denoted as w_i, represents the magnitude of the impact that neighboring points exert on the estimated point.The equations used in the IDW interpolation method are as follows [19].

$$w_i = \frac{\frac{1}{{d_i}^p}}{\sum_i^n \frac{1}{{d_i}^p}} \tag{1}$$

Calculation of the estimated point value uses the following equation [10].

$$\hat{z}_0 = \sum_{i=1}^{n} w_i . z_i \tag{2}$$

$\hat{z}_0$ is the value of the point to be estimated, w_i is weight factor at point i, z_i is The value of the estimator point, d_i is the distance between point i and the estimated point, p power/exponent factor [8].

$$d_i = \sqrt{(x_0 - x_i)^2 + (y_0 - y_i)^2} \tag{3}$$

d_i is The distance between point i and the estimated point, x_0, y_0 is estimated point coordinates, x_i, y_i is coordinates of the measured point [7].

In TPS method, the relative influence of sample data points is determined, with adjacent data points exerting a more significant influence. This results in a surface with higher detail where data points are closer. Conversely, as the distance increases, the influence diminishes, producing a smoother surface with less detailed characteristics [20]. The TPS method aims to estimate values by minimizing the total surface curvature. The equation used in this method is provided in the image below [21].

$$\hat{z}_i = f(x_i y_i) + \sum_{i=1}^{p} \beta_j \, \psi_{ij} + \varepsilon_i \quad (i = 1,2,....,n) \tag{4}$$

$\hat{z}_i$ *is* independent variable, $x_i y_i$ is independent variables observation data i, ψ_{ij} is the number of covariates (j = 1,2,...,p), β_j is weighting parameters, ε_i is errors that are random errors with an average of 0 [22].

$$\iint_R = ((\frac{\partial^2 f}{\partial x^2})^2 + 2\,(\frac{\partial^2 f}{\partial x \partial x})^2 + (\frac{\partial^2 f}{\partial y^2})^2)\, dx\, dy \tag{5}$$

Basis for solving the algorithm will be solved using the U carnel function as follow [1].

$$U = d_i^{\,2} . \text{Log}(d_i^{\,2}) \tag{6}$$

$$d_i = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)}^2 \tag{7}$$

d_i is The distance between point i and the estimated point, x_0, y_0 is estimated point coordinates, x_i, y_i is coordinates of the measured point. By defining the point source as P, which is a (n × 3) matrix and n is the number of data points [7].

$$P = \begin{bmatrix} 1 & x_1 & y_1 \\ 1 & x_2 & y_2 \\ \dots & \dots & \dots \\ 1 & x_n & y_n \end{bmatrix}, n\,x\,3 \tag{8}$$

Using the carnel function, another matrix can be defined, namely the K matrix.

$$K = \begin{bmatrix} 0 & U_{(r12)} & \dots & U_{(r12)} \\ 1 & 0 & \dots & U_{(r12)} \\ \dots & \dots & \dots & \dots \\ 1 & U_{(r12)} & \dots & 0 \end{bmatrix}, nxn \tag{9}$$

$$L = \begin{bmatrix} K & P \\ P^T & 0 \end{bmatrix}, (n + 3) \times (n+3) \tag{10}$$

$$Y = (V \mid 0\ 0\ 0)^T \tag{11}$$

Matrix P This matrix stores the coordinates of the longitude and latitude measurement points and a constant 1 for each observation point. Matrix K This matrix is a kernel matrix defined using the kernel function whose function is to calculate the distance or spatial relationship between points. Matrix L is a combination of K and P, with additional structures to expand the size of the matrix so that it allows the calculation of coefficients in the interpolation method. Matrix Y stores the observation values in vector form[7]. In correlation, P determines the position of the point, K measures the relationship between the points, L is the system used for the interpolation calculation, and Y is the data to be interpolated[12].

Where V is any n-vector of (v1, v2, ..., vn) so that we obtain a vector W = (w1, w2, ..., wn) and the coefficients are α1, αx, αy [22].

$$L^{-1}Y = (w \mid \alpha_1, \alpha_x, \alpha_y)^T \tag{12}$$

By using the L^{-1}Y element, you can define a function f (x, y) with all these formulas.

$$f(x_i, y_i) = \alpha_1, \alpha_x x, \alpha_y y + \textstyle\sum_{i=1}^{n} w_i U(|x_i, y_i) - x_i, y)|] \tag{13}$$

So that $f(x_i, y_i)$ has the second derivative of the integral function, then the value[12].

$$\textstyle\sum_{i=1}^{n} w_i = 0 \text{ and } \sum_{i=1}^{n} w_i x_i = \sum_{i=1}^{n} w_i y_i \tag{14}$$

The results of the spatial interpolation process in this study will be compiled using QGIS software. QGIS is a Geographic Information System (GIS) application used for mapping, spatial analysis, and other functions. The specialty of this software is its open source nature, which allows anyone to use it freely and flexibly, without being limited by time or place [5].

Result and Discussions

Before performing 100 simulations, it is important to interpret the values along the X and Y axes, which are coordinates in the Universal Transverse Mercator (UTM) system. The X (Longitude) and Y (Latitude) coordinates are very important in mapping earthquake magnitudes spatially in the West Sumatra study area. These coordinates reflect the actual values for each earthquake event in the current data period from January 2023 to May 2024. At this stage, the data is divided into an 80:20 proportion. Not only one simulation was carried out but 100 simulations were carried out. In each simulation, different train data and validation data were used, resulting in varying MAPE (Mean Absolute Percentage Error) values across simulations. Each iteration produces a new 90:10 division of train data and validation data, resulting in the collection of 100 different MAPE values.

Table 2. 100 Simulations IDW and TPS Prediction Model.

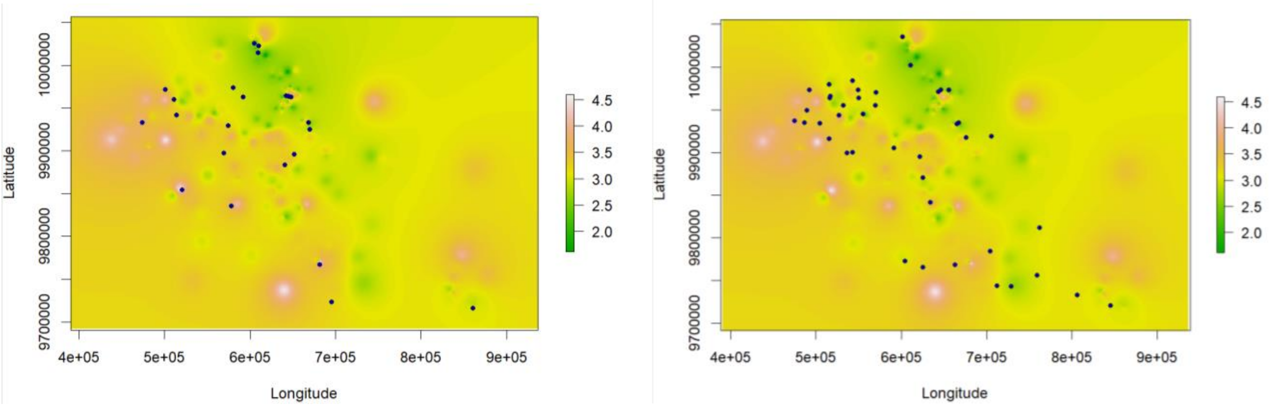
Simulations	MAPE (IDW)	MAPE (TPS)
1	12.80518	16.43780
2	13.66082	13.42932
⋮	⋮	⋮
99	15.13186	13.20750
100	11.71679	14.29466

After completing the 100 repetitions, the next step is to analyze and evaluate the model's estimates by calculating the mean, standard deviation, and confidence interval of the obtained MAPE values. The confidence interval is used to provide a range within which we can be 95% certain that the actual average MAPE of the population lies. This analysis helps to determine whether the model is stable, even when the data is simulated differently.

Table 3. Statistical Values Description of IDW and TPS Prediction Model

	Standard Deviation	Standard Error	T_{Val}	Lower Bound	Upper Bound	MAPE Average 100 times
IDW	1.624	0.162	1.984	13.869	14.514	14.191
TPS	0.159	1.984	13.921	14.555	14.238	0.159

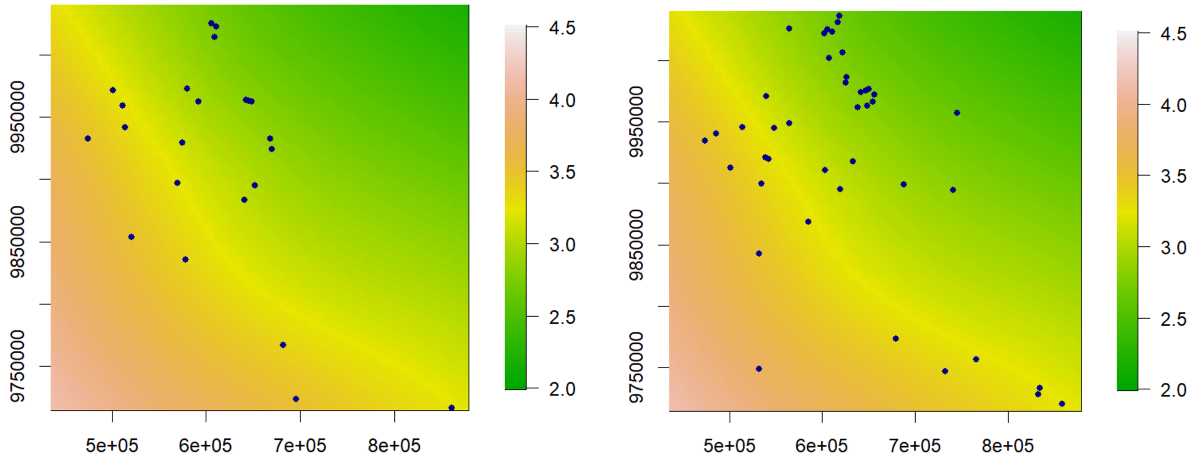
The Hold-Out Method for the Inverse Distance Weighting (IDW) model involves a repetition process, where each simulation begins by resetting the simulation parameters. This approach is designed to achieve consistent results, with a focus on obtaining the median or average MAPE value closest to the true value. Data distribution is performed randomly, with 80% allocated for training data and 20% for validation data in each simulation.



(a) (b)

Figure 4. (a) IDW Estimation Results with Testing Data, (b) IDW Estimation Results with Validation Data

The magnitude interpolation map for the validation data displays earthquake points within the same area. Similar to the testing plot, the colored dots represent the earthquake locations, with various colors indicating the magnitude of the earthquakes on a scale from 2.0 to 4.5. The dominant green areas indicate regions with lower magnitudes, while red to orange areas signify regions with higher magnitudes. This map serves as a valuable tool for future earthquake risk mitigation. The Hold-Out Method for the Thin Plate Spline (TPS) model is conducted using a repetition process, where each simulation begins by resetting the simulation parameters. This approach is designed to achieve consistent results, focusing on obtaining the median or average MAPE value closest to the true value. Data distribution is performed randomly, with 80% allocated for training data and 20% for validation data in each simulation.



(a) (b)

Figure 5. (a) TPS Estimation Results with Testing Data, (b) TPS Estimation Results with Validation Data

The visualization in Figure 5 depicts the results of the Thin Plate Spline (TPS) interpolation for the validation data. In this plot, the horizontal axis represents longitude, while the vertical axis represents latitude. The color scale on the right ranges from 2.0 to 4.5, indicating the magnitude of earthquake strength across different locations. Warmer colors, ranging from yellow to red, signify areas with higher potential earthquake strength, whereas green hues indicate regions with lower potential earthquake strength. The scattered dots correspond to the specific locations where earthquakes have occurred. To determine the optimal method for estimating earthquake strength in West Sumatra, MAPE values obtained from both IDW and TPS methods. The MAPE results for both interpolation techniques are summarized in the table 4.

Table 4. MAPE IDW and TPS values

	IDW	**TPS**
Testing	18.56726	16.41764
Validation	14.28197	14.2946

The visualization of earthquake strength interpolation results is represented in Figure 6. In the West Sumatra region, it is evident that the northern area, particularly around Bukittinggi and Lubuk Basung, exhibits the highest earthquake activity, as indicated by the numerous orange dots representing moderate to high magnitude earthquakes. The central region, encompassing Padang Panjang, Padang, and Solok, also shows a concentration of green and yellow dots, reflecting moderate magnitude earthquake activity.

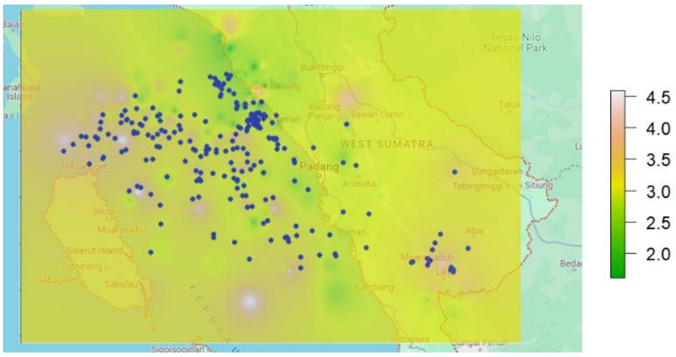


Figure 6. Visualization of Earthquake Strength Interpolation Results

Although earthquake activity in the central region is not as intense as in the northern region, it remains significant. In the southern region, including Painan, Muara Labuh, and Kambang, most earthquake points are green and yellow, indicating moderate to low magnitude activity. However, a few orange dots suggest higher magnitude earthquakes. Along the west coast of West Sumatra, several orange and red points are visible, indicating moderate to high magnitude earthquake activity[4].

Overall, the northern region of West Sumatra is the most earthquake-prone, with the highest frequency and strongest earthquake activity, followed by the central and southern regions. The visualization of interpolation results, conducted to estimate earthquake strength in West Sumatra, provides valuable information for the West Sumatra government in enhancing mitigation efforts and optimizing disaster risk reduction, particularly for earthquakes[19]. This information is crucial for enabling the government to respond more effectively and proactively to earthquake threats. The public is also urged to remain vigilant about the potential for earthquake disasters, given West Sumatra's high seismic activity. Therefore, enhancing both structural and non-structural disaster mitigation efforts is essential. The application of geographic information systems (GIS) in this context can significantly contribute to risk reduction and community preparedness for disaster threats[22].

Conclusion

The frequency of earthquake magnitudes in West Sumatra from January 2023 to May 2024 indicates that most earthquake events in the region fell into the moderate category. The majority of significant earthquakes had magnitudes ranging from 2.6 to 3.6, suggesting that earthquakes of this strength occur more frequently than both stronger and weaker ones. Although there were some earthquakes with higher magnitudes, with the highest recorded at 4.6, their frequency was considerably lower. Based on the analysis, the Inverse Distance Weighted (IDW) method was identified as the most effective approach for predicting earthquake strength in West Sumatra.

Acknowledgment

The author thanks to Statistics Department, Faculty of Mathematics and Natural Science, Universitas Islam Indonesia which has assisted in the research process.

References

[1] W. Erlangga, M. Teguh, M. Mushthofa, I. Trianggoro Saputro, and G. Setiadi, Sebaran gempa utama berdasarkan magnitudo dan kedalaman di wilayah Mamuju dan sekitarnya, Teknisia 27(2) (2022) 122–131.
[2] R. Metrikasari and A. Choiruddin, Pemodelan risiko gempa bumi di pulau Sumatera menggunakan model inhomogeneous neyman-scott cox process, Jurnal Sains dan Seni ITS, 9(2) (2021).
[3] D. M. Guwandi, H. Andreas, and M. A. Kusuma, Analisis permasalahan sistem koordinat pada kegiatan well tie data seismik, ITB Indonesian Journal of Geospatial, 6(1) (2017) 11–30.
[4] A. D. W. Nasrullah, S. Hanifarianty, and I. Fauzi, Pola sebaran kerusakan bangunan akibat gempa bumi berbasis metode Inverse Distance Weighting (studi kasus kota Palu, provinsi Sulawesi Tengah), Jurnal pendidikan ilmu sosial, 28(2) (2019) 94–104.
[5] I. W. S. Adnyana and Abd. R. As-syakur, Application of Geographic Information System (GIS) based raster data to classify land capability in Bali Province by using differentiator pixel value method, J.Manusia dan Lingkungan, 19(1) (2012) 21–29.
[6] S. P. Sejati, Perbandingan akurasi metode Inverse Distance Weighting dan Kriging dalam pemetaan kedalaman muka air tanah, Majalah Geografi Indonesia, 33(2) (2020).
[7] H. kurniadi, E. Aprilia, J. B. Utomo, A. Kurniawan, and A. Safril, Perbandingan metode IDW dan spline dalam interpolasi data curah hujan (studi kasus curah hujan bulanan di Jawa Timur periode 2012-2016), in Prosiding Seminar Nasional Geotik 2018, (2020) 213–220.
[8] F. Amalia, Interpolasi spasial dengan pendekatan Inverse Distance Weighted dan Ordinary Kriging dalam estimasi kekuatan gempa bumi, Final Task, Universitas Islam Indonesia, Yogyakarta, 2023.
[9] R. A. Yudanegara, D. Astutik, A. Hernandi, T. P. Soedarmodjo, and E. Alexander, Penggunaan metode inverse distance weighted (IDW) untuk pemetaan zona nilai tanah (studi kasus: kelurahan gedong meneng, Bandar Lampung), Elipsoida : Jurnal Geodesi dan Geomatika, 4(2) (2021) 85–90.
[10] J. P. Musashi, H. Pramoedyo, and R. Fitriani, Comparison of Inverse Distance Weighted and Natural Neighbor Interpolation method at air temperature data in Malang region, CAUCHY: Jurnal Matematika Murni dan Aplikasi, 5(2) (2018) 48–54.
[11] R. L. Islami and P. R. Sihombing, Interpolasi curah hujan ekstrim menggunakan model spatial di provinsi Jawa Barat, Jurnal Aplikasi Statistika & Komputasi Statistik, 13(2) (2021) 15–22.
[12] W. Keller and A. Borkowski, Thin plate spline interpolation, J Geod, 93(9) (2019) 1251–1269.
[13] A. Nur Halim, A. Karisma Aji Restu, and U. Muhammadiyah Kalimantan Timur, Geospatial data optimization for mapping updates in Sungai Pinang district using the QGIS approach, Pengabdian Kepada Masyarakat, 1(6) (2023).
[14] D. Hartono et al., Analisis sejarah, dampak, dan penanggulangan bencana gempa bumi pada saat pandemi covid-19 di Sulawesi Barat, PENDIPA Journal of Science Education, 5(2) (2021) 218–224.
[15] G. M. S. Febriyan, Pemanfaatan data geospasial dalam proses pembelajaran geografi pada kondisi bencana covid-19, Jurnal Geografika (Geografi Lingkungan Lahan Basah) , 1(1) (2020) 20–29.
[16] L. Singal and Y. D. Y. Rindengan, Comparative analysis of google maps coordinates points and professional GPS tools in Manado city, Jurnal Teknik Informatika, 16(2) (2021) 157–164.
[17] N. R. Hasna, A. Setiawan, and H. A. Parhusip, Penentuan lokasi lumbung pangan berdasarkan gravity location models dengan koordinat UTM di provinsi Maluku Utara, Jurnal Sains dan Edukasi Sains, 1(2) (2018) 7–16.
[18] N. Arif, Studi komparasi kriging dan IDW untuk estimasi spasial bahan organik tanah, Geomedia: Majalah Ilmiah dan Informasi Kegeografian, 17(2) (2019).
[19] E. Respatti, R. Goejantoro, and S. Wahyuningsih, Perbandingan metode Ordinary Kriging dan Inverse Distance Weighted untuk estimasi elevasi pada data topografi (studi kasus: topografi wilayah FMIPA Universitas Mulawarman), Jurnal EKSPONENSIAL, 5(2) (2014).
[20] W. Keller and A. Borkowski, Thin plate spline interpolation, J Geod, 93(9) (2019) 1251–1269.
[21] M. Dunitz, Thin-plate splines on the sphere for interpolation, computing spherical averages, and solving inverse problems, Image Processing On Line, 13 (2023) 199.
[22] R. E. Andarasmono, Penggunaan thin plate spline untuk memetakan penyebaran penyakit diare di kabupaten dan kota propinsi Jawa Timur, Thesis, Universitas Airlangga, (2003).