

A Residual-Based Hybrid SVR–Ordinary Kriging–IDW with Residual Correction for Spatial Interpolation of Seismicity Parameters in Sumatra

Dwi Agustin Nuriani Sirodji^{1,2*}, Muhammad Nur Aidi¹, Bagus Sartono¹, Utami Dyah Syafitri¹, Bayu Pranata³

¹School of Data Science, Mathematics and Informatics, IPB University, Bogor, West Java, 16680, Indonesia

²Statistics Study Program, Universitas Islam Bandung, Bandung, West Java, 40116, Indonesia

³Earthquake and Tsunami Center, Agency for Meteorology, Climatology, and Geophysics (BMKG), Jakarta, 10610, Indonesia

*Corresponding author: dwi.agustinnuriani@unisba.ac.id

Abstract

Accurate spatial prediction of seismicity parameters is important for characterizing seismotectonic behavior in tectonically complex regions. This study developed a hybrid spatial interpolation framework integrating Support Vector Regression–Ordinary Kriging (SVR–OK) with Inverse Distance Weighting (IDW) residual correction for estimating seismic activity (a-value) and rock fragility (b-value) across Sumatra. A total of 238 seismicity data points derived from ISC and BMKG earthquake catalogs ($M_w > 5.0$) for 2000–2023 were analyzed. SVR–OK was used to capture global spatial variation through flexible empirical semivariogram modeling, while IDW-based residual correction captured local variation remaining unexplained by the primary model. Performance was evaluated using LOOCV, independent testing, RMSE, MAE, RMSEP, residual diagnostics, and Moran's I. The Hybrid SVR–OK–IDW Residual model achieved the lowest test RMSEP, reducing errors from 0.961 to 0.837 for the a-value and from 0.166 to 0.149 for the b-value, while reducing remaining positive spatial autocorrelation. Spatial contour maps identified distinct seismotectonic zonation, particularly in the Nias–Simeulue and Mentawai–Pagai segments. Simulation experiments showed that its advantage was most pronounced for clustered spatial patterns at small to moderate sample sizes ($n = 64$ and 225), whereas Gaussian–OK and SVR–OK performed better for unclustered patterns. These findings demonstrate the advantage of combining SVR–OK for global spatial variation with IDW residual correction for localized variation, particularly under clustered and spatially heterogeneous conditions.

Keywords

Absolute Residual Correction, Geostatistical Modeling, Hybrid Interpolation, Seismicity Parameters, SVR–Ordinary Kriging

Received: 2 July 2026, Accepted: 6 September 2026

<https://doi.org/10.26554/sti.2026.11.4.1688-1704>

1. INTRODUCTION

Understanding the spatial variability of seismicity parameters is fundamental to characterizing seismotectonic processes and improving earthquake hazard assessment in tectonically active regions. In geostatistics, kriging is one of the most widely used interpolation methods due to its ability to incorporate spatial dependence through semivariogram modeling. However, kriging relies on several key assumptions, including linearity and second-order stationarity, where the mean is constant across the domain and spatial dependence depends only on distance. In practice, these assumptions are often difficult to satisfy, as real-world spatial data frequently exhibit complex, non-linear, and non-stationary structures (Erdogan Erten et al., 2022; Van Beers and Kleijnen, 2003). To address these limitations, recent studies have explored the integration of multiple interpolation techniques. Hybrid spatial interpolation methods combine geostatistical approaches, such as Ordinary Kriging

(OK), with non-geostatistical methods, including Inverse Distance Weighting (IDW), as well as machine learning algorithms. These hybrid approaches aim to improve predictive accuracy by leveraging the strengths of each method. Previous studies have demonstrated that hybrid models can outperform individual interpolation techniques, particularly in complex spatial environments (Li and Heap, 2011; Ramos et al., 2015).

One important advancement in hybrid interpolation is the incorporation of residual-based modeling. In this framework, the primary model is first used to generate predictions, and the remaining spatial residuals are subsequently modeled to improve prediction accuracy. Regression kriging, for example, combines regression-based prediction with kriging of the residual component (Hengl et al., 2007). This principle has also been extended to machine-learning-based hybrid approaches, in which machine learning is generally used to model the response or deterministic component, followed by geostatistical modeling of the remaining spatial residuals (Li et al., 2011;

Seo et al., 2015). Other hybrid approaches have combined distance-based and geostatistical interpolation methods, including IDW and kriging, to exploit their complementary ability to represent local and spatially structured variation. These studies demonstrate the potential of hybrid interpolation; however, their methodological structures differ in how machine learning, kriging, and residual information are incorporated.

Many existing hybrid approaches also rely on parametric semivariogram models, such as spherical, exponential, and Gaussian functions, which may not adequately capture complex spatial patterns observed in empirical data (Li, 2022). To overcome this limitation, nonparametric and data-driven approaches for semivariogram modeling have been proposed. Among these, Support Vector Regression (SVR) has demonstrated the ability to model empirical semivariograms without requiring a predefined functional form, allowing nonlinear spatial relationships to be represented more flexibly (Han et al., 2016; Huang et al., 2012). Unlike conventional machine-learning-kriging approaches, the framework developed in this study uses SVR specifically for nonparametric semivariogram modeling rather than for direct prediction of the response variable. The resulting SVR-based semivariogram is incorporated into Ordinary Kriging to obtain the baseline spatial prediction, after which IDW is applied to the remaining residual variation as a local correction. Two formulations, based on signed and absolute residuals, are evaluated within this framework. Thus, the methodological contribution lies in the sequential integration of SVR-based nonparametric semivariogram modeling, Ordinary Kriging, and IDW-based residual correction within a unified spatial interpolation framework.

Sumatra Island is one of the most seismically active regions in Indonesia, characterized by a complex tectonic setting involving the subduction of the Indo-Australian Plate beneath the Eurasian Plate and the presence of the Sumatran Fault Zone (SFZ). This tectonic complexity leads to heterogeneous seismicity patterns across the region. Seismicity parameters, such as the seismic activity (a-value) and rock fragility (b-value), play a crucial role in characterizing earthquake behavior and assessing seismic hazard (Chasanah and Handoyo, 2021). Previous studies have highlighted significant spatial variability in these parameters, reflecting differences in tectonic conditions and stress accumulation across fault segments (Putra et al., 2023).

Motivated by these challenges, this study develops a residual-based hybrid spatial interpolation framework integrating Support Vector Regression–Ordinary Kriging (SVR–OK) with Inverse Distance Weighting (IDW). SVR provides flexible nonparametric semivariogram modeling, while signed and absolute residual corrections are incorporated to capture local spatial variation not explained by the primary model. The study contributes a unified SVR–OK–IDW framework and evaluates both residual correction strategies for seismic activity (a-value) and rock fragility (b-value) under the heterogeneous tectonic conditions of Sumatra. This framework provides an adaptive approach for improving spatial prediction while allowing the effectiveness of residual correction to be evaluated according

to the underlying spatial structure.

2. EXPERIMENTAL SECTION

2.1 Empirical Data

The dataset consisted of seismicity parameters (a-value and b-value) computed using ZMAP 6.0 from ISC Bulletin and BMKG earthquake catalogs covering 1 January 2000 to 31 December 2023. Earthquakes with $M_w > 5.0$ were selected to focus the analysis on moderate-to-large seismic events with greater potential impacts, rather than to represent the magnitude of completeness (M_c) of the earthquake catalog (Ma et al., 2021). The study area (95°E – 105°E , 6°N – 6°S) was divided into a $0.50^\circ \times 0.50^\circ$ grid, yielding 238 cells with valid seismicity parameter estimates. This resolution balanced spatial detail and sufficient earthquake events for reliable a-value and b-value estimation, avoiding data limitations at finer grids and loss of local variation at coarser grids. The 238 locations were randomly divided into 190 training (80%) and 48 testing (20%) locations. LOOCV was performed exclusively on the training data for model development and selection, while the testing data were kept independent and used only to evaluate the predictive performance of the selected model. The spatial distribution of the testing locations across the study area is presented in Sirodj et al. (2025b). The earthquake magnitude–frequency relationship follows the Gutenberg–Richter law (Gutenberg and Richter, 1944), as expressed in Equation 1:

$$\log N(M) = a - bM \quad (1)$$

where $N(M)$ is the cumulative number of earthquakes with magnitudes $\geq (M)$. The parameter a-value represents seismic activity, whereas b-value reflects the relative proportion of small and large earthquakes and provides insight into crustal stress conditions. Both parameters were estimated using the maximum likelihood method (Palacios et al., 2006), as expressed in Equation 2.

$$\hat{b} = \frac{\log e}{(\overline{M} - m_0)}, \quad (2)$$

$$\hat{a} = \log N(M \geq M_0) + \log(\hat{b} \ln 10) + m_0 \hat{b}.$$

where $\overline{M} = \frac{\sum_{i=1}^n M_i N_i}{\sum_{i=1}^n N_i}$ is the mean magnitude of the earthquake data; M_i is the magnitude of the i -th earthquake; N_i is the number of earthquake events with the i -th magnitude; and m_0 is the magnitude of completeness of the earthquake catalog. To ensure consistency, earthquake magnitudes reported as Surface Wave Magnitude (M_s), Body Wave Magnitude (m_b), and Local Magnitude (M_L) were homogenized to Moment Magnitude (M_w) following (Arimuko et al., 2023). The magnitude conversion relationships were defined as expressed in Equations 3–5.

$$M_w = 1.0107m_b + 0.0801, \quad 3.7 \leq m_b \leq 8.2 \quad (3)$$

$$M_w = \begin{cases} 0.6016M_s + 2.476, & 2.8 \leq M_s \leq 6.1, \\ 0.9239M_s + 0.8183, & 6.2 \leq M_s \leq 8.7. \end{cases} \quad (4)$$

$$M_w = M_L \quad (5)$$

The conversion equations were applied according to their respective magnitude ranges to obtain a homogeneous M_w -based earthquake catalog. The homogenized catalog was subsequently declustered using the Uhrhammer algorithm to remove dependent events and retain mainshock events for further analysis (Galina et al., 2019).

Table 1. Simulation Data Scenarios

Scenario	Sample Size	Spatial Response Pattern
1	64	Unclassified
2		Classified
3	225	Unclassified
4		Classified
5	625	Unclassified
6		Classified

2.2 Simulation Data

The simulation design combined different sample sizes and spatial response patterns under a Gaussian semivariogram structure. Each combination defined a simulation scenario for comparing Gaussian–Ordinary Kriging, SVR–Ordinary Kriging, Hybrid SVR–Ordinary Kriging–IDW Residual, and Hybrid SVR–Ordinary Kriging–IDW Absolute Residual. The simulation scenarios are summarized in Table 1.

To systematically evaluate the performance of the proposed interpolation methods under controlled spatial conditions, simulation data were generated across different sample sizes and spatial response patterns. A common underlying spatial dependence structure was maintained across scenarios to ensure comparability among methods. The simulation data were generated as follows:

a) Simulation parameters. The spatial dependence structure was generated using a Gaussian semivariogram model with a nugget $c_0 = 0.1$, a partial sill $c_1 = 1$, and a range parameter $a = 5$, as expressed in Equation 6.

$$\gamma(h) = \begin{cases} 0, & h = 0, \\ c_0 + c_1 \left[1 - \exp \left(- (h/a)^2 \right) \right], & h > 0. \end{cases} \quad (6)$$

b) Spatial domain and sample size. Simulations were conducted over the study area ($95^\circ\text{E} - 105^\circ\text{E}$, $6^\circ\text{N} - 6^\circ\text{S}$) using sample sizes of $n = 64$, 225 , and 625 , corresponding to 8×8 , 15×15 , and 25×25 grids, respectively.

c) Spatial process generation. The covariance matrix Σ was constructed from pairwise Euclidean distances using the Gaussian semivariogram model. The baseline spatial process was generated from a multivariate normal distribution (Equation 7) and subsequently rescaled to the specified response interval (Equation 8).

$$\mathbf{Z} \sim N_n(\mathbf{0}, \Sigma) \quad (7)$$

$$Z_i^* = 5 + \frac{Z_i - \min(\mathbf{Z})}{\max(\mathbf{Z}) - \min(\mathbf{Z})} (50 - 5) \quad (8)$$

d) Unclassified pattern. The rescaled Gaussian spatial process was directly used as the response without introducing an additional clustering component, as expressed in Equation 9.

$$Z_{i,\text{unclassified}} = Z_i^* \quad (9)$$

e) Classified pattern. Two cluster centers separated by at least 10 spatial units were randomly generated. Their local effects were modeled using Gaussian kernels (Equation 10), representing high- and low-response regions. The final response combined the baseline spatial process, cluster effects, and random error (Equations 11 and 12).

$$w_{ik} = \exp \left(- \frac{d_{ik}^2}{2\sigma_k^2} \right), \quad k \in \{1, 2\} \quad (10)$$

$$Z_{\text{classified}}(u_i) = (1 - w_{i2})(Z_i^* + A_1 w_{i1}) + w_{i2}(B_2 + A_2 w_{i2}) + \varepsilon_i \quad (11)$$

$$\varepsilon_i \sim N(0, \sigma_\varepsilon^2) \quad (12)$$

where d_{ik} denotes the Euclidean distance between the i -th location and the k -th cluster center, w_{ik} is the corresponding Gaussian kernel weight, and σ_k controls the spatial spread of each cluster. Z_i^* denotes the baseline spatial process, $A_1 = 100$ is the amplitude of the high-response cluster, while $B_2 = 2$ and $A_2 = 8$ determine the level and amplitude of the low-response cluster, respectively. The random error ε_i follows $N(0, \sigma_\varepsilon^2)$, with $\sigma_\varepsilon = 1$.

f) Simulation replications. Each combination of sample size and spatial pattern was independently generated over 100 simulation replications to evaluate the performance of the spatial interpolation methods under different sampling densities and spatial structures.

2.3 SVR–Ordinary Kriging Method

Traditional kriging is limited by the availability of theoretical semivariogram models and manual parameter selection, which may reduce fitting accuracy (Chen et al., 2025). Support Vector Regression (SVR) provides a data-driven alternative capable

Table 2. SVR-Ordinary Kriging

Input: Observed values z_i , coordinates $u_i = (x_i, y_i)$, and training set $D_{\text{train}} = \{x_i, y_i, z_i\}, i = 1, \dots, n$
Output: Predictions $\hat{z}_i^{\text{SVR-OK}}$

LOOCV Procedurefor each location $i = 1, \dots, n$ doRemove $\{x_i, y_i, z_i\}$ to form subset $D^{(-i)}$ **Step 1: Empirical Semivariogram**Compute pairwise distances $h_{jk} = \|u_j - u_k\|, j, k \neq i$ Estimate empirical semivariance $\gamma^{(-i)}(h)$ using binning**Step 2: SVR-Based Semivariogram Fitting**Train SVR with input h and output $\gamma^{(-i)}(h)$ Optimize hyperparameters (C, ϵ) Obtain fitted function $\hat{\gamma}^{(-i)}(h)$ **Step 3: Construct Kriging System**Build semivariance matrix: $\Gamma_{jk}^{(-i)} = \hat{\gamma}_{\text{SVR}}^{(-i)}(\|u_j - u_k\|), j, k \neq i$ Compute semivariance vector: $\gamma_{0j}^{(-i)} = \hat{\gamma}_{\text{SVR}}^{(-i)}(\|u_0 - u_j\|), j \neq i$

of capturing irregular spatial variability through kernel-based nonlinear modeling (Han et al., 2016; Huang et al., 2012; Montesinos López et al., 2022). Its performance is primarily controlled by the penalty parameter C , the width of the ϵ -insensitive tube, and the kernel function, such as Gaussian (RBF), polynomial, or linear kernels (Wu et al., 2009). The integration of SVR into Ordinary Kriging (SVR-OK), previously developed by (Sirodj et al., 2025a), replaces the conventional parametric semivariogram with an SVR-fitted empirical semivariogram. The resulting semivariogram is incorporated into the kriging system to determine interpolation weights, enabling SVR-OK to capture the overall spatial variation in heterogeneous seismicity parameters. In the present study, this SVR-OK framework serves as the primary interpolation model and is further extended through IDW-based residual correction to capture local spatial variation remaining unexplained by the primary model. The complete SVR-Ordinary Kriging procedure is presented in Table 2.

2.4 Hybrid SVR-Ordinary Kriging-IDW Method

The development of the proposed hybrid method was motivated by a key limitation of the existing IDW-OK hybrid framework introduced by Li (2022), in which the semivariogram fitting stage relies on traditional parametric models. Under heterogeneous spatial dependence structures, these models often fail to adequately represent the empirical semivariogram, resulting in fitting bias and reduced predictive accuracy. To address this limitation, a non-parametric semivariogram modeling approach based on Support Vector Regression (SVR) was incorporated into the empirical semivariogram fitting stage of the hybrid interpolation framework.

The proposed Hybrid SVR-Ordinary Kriging-IDW (SVR-OK-IDW) method was developed by first generating primary predictions using the SVR-OK model. Residuals were subsequently computed and interpolated using the Inverse Distance

Weighting (IDW) method to improve local prediction accuracy. Although IDW generally exhibits lower standalone predictive performance, it was retained as a baseline method because of its widespread use as a benchmark in spatial interpolation studies (Li and Heap, 2011). Residual-based hybrid modeling has been widely applied in time-series and forecasting applications due to several methodological advantages, including: (i) improved predictive accuracy; (ii) efficient model handling; and (iii) correction of biased predictions (de Mattos Neto et al., 2022). In this study, the residual-based hybrid framework was extended to the spatial interpolation domain.

Two types of residuals were considered: standard residuals and absolute residuals, as defined in Equations 13 and 14, respectively.

$$r_i = z_i - \hat{z}_i^{(\text{SVR-OK})} \quad (13)$$

$$|r_i| = |z_i - \hat{z}_i^{(\text{SVR-OK})}| \quad (14)$$

Standard residuals preserve directional error information but may introduce instability due to asymmetry. In contrast, absolute residuals remove sign bias and provide a more robust measure of prediction error (García-Díaz, Noel and Verduzo-Ramirez, Alberto and Garcia-Virgen, Juan and Muñoz, Lilia, 2016; Shepperd and MacDonell, 2012). The Hybrid SVR-Ordinary Kriging-IDW Residual model enhances the SVR-Ordinary Kriging (SVR-OK) prediction by incorporating an IDW-based interpolation of the residuals, allowing local spatial variability to be captured and improving prediction accuracy. The final prediction is expressed in Equation 15.

$$\hat{z}_i^{\text{Hyb-Res}} = \hat{z}_i^{\text{SVR-OK}} + \hat{r}_i^{\text{IDW}} \quad (15)$$

Table 3. Hybrid SVR–Ordinary Kriging–IDW Residual

Input: $z_i, u_i = (x_i, y_i)$
Output: $\hat{z}_i^{Hyb-Res}$
Step 1: SVR–OK Prediction Obtain $\hat{z}_{i,-i}^{SVR-OK}$
Step 2: Residual Construction for each location $i = 1, \dots, n$ do Exclude r_i from the residual dataset when estimating the residual correction at location i : $R^{(-i)} = \{(u_j, r_j) : j \neq i\}$
Step 3: IDW Interpolation $d_{ij} = \ u_i - u_j\ , \quad j \neq i$ $w_{ij} = \frac{d_{ij}^{-p}}{\sum_{k \neq i} d_{ik}^{-p}}, \quad j \neq i$ $\hat{r}_{i,-i}^{IDW} = \sum_{j \neq i} w_{ij} r_j$
Step 4: Hybrid Prediction $\hat{z}_i^{Hyb-Res} = \hat{z}_{i,-i}^{SVR-OK} + \hat{r}_{i,-i}^{IDW}$
Step 5: Residual $e_i^{Hyb-Res} = z_i - \hat{z}_i^{Hyb-Res}$ End for
Step 6: Evaluation Compute RMSE and MAE

where $\hat{z}_i^{SVR-OK}$ denotes the prediction obtained from the SVR–Ordinary Kriging model and $\hat{r}_i^{IDW}$ represents the residual value estimated through IDW interpolation.

To account for the magnitude of local prediction errors, an absolute residual correction is introduced by interpolating the absolute residuals using IDW. Unlike the signed residual approach, this formulation uses only the residual magnitude as an additive local correction to the SVR–Ordinary Kriging prediction. The final prediction is expressed in Equation 16.

$$\hat{z}_i^{Hyb-AbsRes} = \hat{z}_i^{SVR-OK} + |\hat{r}_i|^{IDW} \quad (16)$$

where $|\hat{r}_i|^{IDW}$ is the IDW-interpolated absolute residual. The pseudocode of the proposed hybrid approaches is presented in Table 3 and Table 4.

2.5 Interpolation Model Validation and Accuracy Assessment

The model validation process was conducted using the Leave-One-Out Cross-Validation (LOOCV) approach to evaluate the predictive performance of each interpolation model. The performance measures employed in this study were the Root

Mean Square Error (RMSE) and Mean Absolute Error (MAE), as defined in Equations 17 and 18, respectively.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (z_i - \hat{z}_i)^2} \quad (17)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |z_i - \hat{z}_i| \quad (18)$$

where z_i represents the observed value, $\hat{z}_i$ denotes the predicted value obtained from the SVR–Ordinary Kriging ($\hat{z}_i^{SVR-OK}$), Hybrid SVR–Ordinary Kriging–IDW Residual ($\hat{z}_i^{Hyb-Res}$), or Hybrid SVR–Ordinary Kriging–IDW Absolute Residual ($\hat{z}_i^{Hyb-AbsRes}$) models, and n is the total number of observations.

2.6 Residual Diagnostic Analysis

Model adequacy was evaluated through residual normality and spatial autocorrelation analyses. Residual normality was assessed using the Shapiro–Wilk statistic, with the test statistic defined in Equation 19.

Table 4. Hybrid SVR–Ordinary Kriging–IDW Absolute Residual

Input: $z_i, u_i = (x_i, y_i)$
Output: $\hat{z}_i^{Hyb-Abs-Res}$
Step 1: SVR–OK Prediction Obtain $\hat{z}_{i,-i}^{SVR-OK}$
Step 2: Absolute Residual Construction for each location $i = 1, \dots, n$ do Exclude r_i from the residual dataset when estimating the absolute residual correction at location i : $R^{(-i)} = \{(u_j, r_j) : j \neq i\}$
Step 3: IDW Interpolation $d_{ij} = \ u_i - u_j\ , \quad j \neq i$ $w_{ij} = \frac{d_{ij}^{-p}}{\sum_{k \neq i} d_{ik}^{-p}}, \quad j \neq i$ $\hat{r}_{i,-i}^{IDW} = \sum_{j \neq i} w_{ij} r_j$
Step 4: Hybrid Prediction $\hat{z}_i^{Hyb-Abs-Res} = \hat{z}_{i,-i}^{SVR-OK} + \hat{r}_{i,-i}^{IDW}$
Step 5: Residual $e_i^{Hyb-Abs-Res} = z_i - \hat{z}_i^{Hyb-Abs-Res}$ End for
Step 6: Evaluation Compute RMSE and MAE

$$W = \frac{(\sum_{i=1}^n a_i e(i))^2}{\sum_{i=1}^n (e_i - \bar{e})^2} \quad (19)$$

where $e(i)$ denotes the ordered residual, $\bar{e}$ is the residual mean, and a_i represents constants derived from the expected values and covariance matrix of order statistics from a standard normal distribution. Residuals were considered normally distributed when ($p > 0.05$).

Spatial dependence was evaluated using Moran's I, as defined in Equation 20.

$$I = \frac{n}{S_0} \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} (e_i - \bar{e})(e_j - \bar{e})}{\sum_{i=1}^n (e_i - \bar{e})^2} \quad (20)$$

where e_i and e_j are residuals at locations i and j , w_{ij} denotes the spatial weight, and S_0 is the sum of all spatial weights. Residuals with non-significant Moran's I values ($p > 0.05$) were

considered spatially independent, indicating that the interpolation model adequately captured the underlying spatial dependence structure. Both diagnostics were applied to the residuals of the SVR–Ordinary Kriging, Hybrid SVR–Ordinary Kriging–IDW Residual, and Hybrid SVR–Ordinary Kriging–IDW Absolute Residual models. The overall workflow of the proposed methodology is presented in Figure 1.

3. RESULTS AND DISCUSSION

3.1 Empirical Results

A critical component of the kriging interpolation procedure is the characterization of the empirical semivariogram, which governs the spatial dependence structure of spatial data. The selection of an appropriate semivariogram model is essential for ensuring accurate spatial prediction because it directly influences the kriging weights and, consequently, the resulting interpolation surface. An inadequate representation of the empirical semivariogram may lead to biased spatial dependence estimates and reduced interpolation accuracy. The empirical semivariograms derived from the seismicity parameters representing seismic activity (a-value) and rock fragility (b-value) across Sumatra are presented in Figure 2.

Figure 2 illustrates the theoretical semivariogram fitting

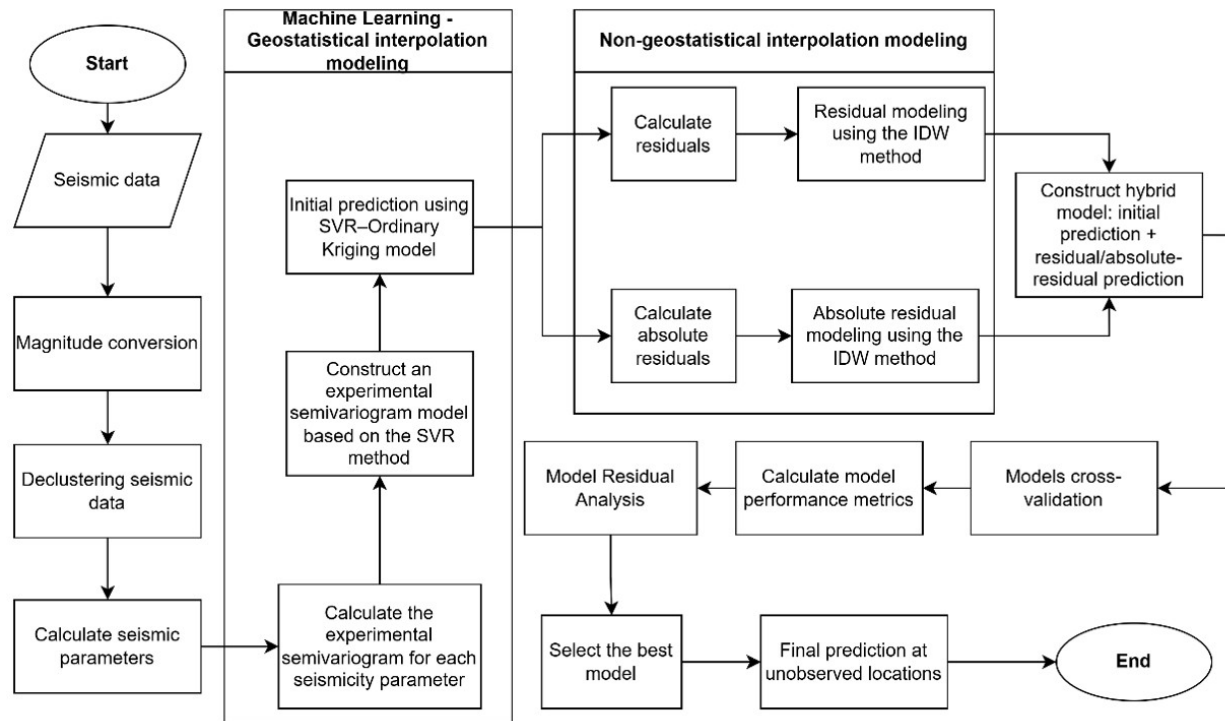


Figure 1. Research Workflow of the Proposed Hybrid SVR–Ordinary Kriging–IDW Framework

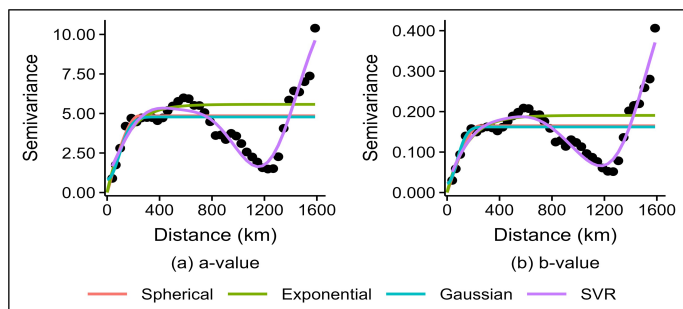


Figure 2. Empirical Semivariograms of Seismicity Parameters Associated With Seismic Activity (a-Value) and Rock Fragility (b-Value) Across Sumatra

results for the seismicity parameters representing seismic activity (a-value) and rock fragility (b-value) across Sumatra. The empirical semivariograms exhibit pronounced non-monotonic patterns, indicating that spatial dependence does not change uniformly with increasing separation distance. These fluctuations reflect the complex and heterogeneous spatial structure of seismicity across Sumatra, which is influenced by variations in earthquake occurrence associated with tectonic segmentation. This interpretation is consistent with the spatial distribution of seismicity identified in our previous study, where earthquake activity was predominantly clustered in northern and southern Sumatra, while a more scattered distribution was observed in the central region (Sirodj et al., 2025b).

The conventional Exponential, Gaussian, and Spherical semivariogram models generally capture the overall short-range spatial continuity but are less capable of reproducing fluctuations in empirical semivariance at medium and longer separation distances because their predefined functional forms impose relatively smooth distance-dependent structures. Consequently, these parametric models may not fully represent the complex spatial dependence exhibited by the seismicity parameters. In contrast, Support Vector Regression (SVR) provides greater flexibility by learning the relationship between separation distance and empirical semivariance directly from the data without imposing a predefined semivariogram functional form. This data-driven formulation enables SVR to accommodate local fluctuations and non-monotonic changes in empirical semivariance more effectively. The advantage of this approach was quantitatively demonstrated in our previous study (Sirodj et al., 2025a), in which SVR produced lower semivariogram fitting errors than the Gaussian model for both the a-value and b-value. These findings support the use of an SVR-based nonparametric semivariogram as a more flexible representation of the complex spatial dependence underlying seismic activity and rock fragility across Sumatra.

To obtain an optimal representation of the empirical semivariogram, the SVR model employs a Radial Basis Function (RBF) kernel, with its hyperparameters optimized through a grid-search procedure. The grid search systematically evaluates predefined combinations of hyperparameters and selects the configuration that provides the best semivariogram fitting

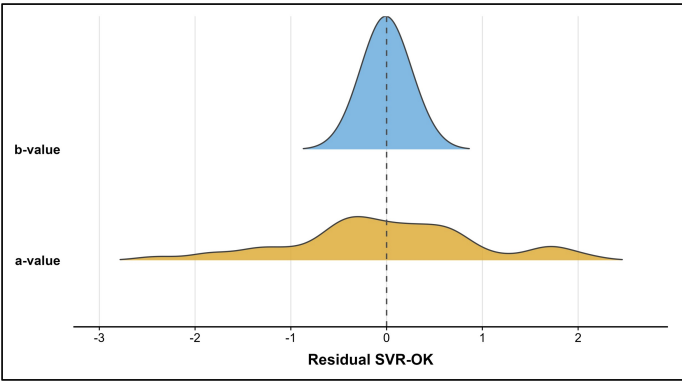


Figure 3. Residual Distributions of the SVR–Ordinary Kriging Model for Seismic Activity (a-Value) and Rock Fragility (b-Value)

performance. The detailed optimization procedure, including the parameter search space and selection criteria, is provided in Sirodj et al. (2025a). The selected configuration includes an ϵ -insensitive loss value of 0.2 and a penalty parameter of $C=100$. The RBF kernel is well suited to this application because of its flexibility in capturing nonlinear and irregular patterns (Ding et al., 2021; Elen et al., 2022; Han et al., 2016).

The predictive applicability of the fitted SVR semivariogram was further demonstrated through its integration into the Ordinary Kriging framework (Sirodj et al., 2025a). Instead of using a conventional parametric semivariogram, the semivariance values required in the kriging system were derived from the optimized SVR semivariogram. Within this framework, SVR models the spatial dependence structure represented by the empirical semivariogram, while Ordinary Kriging uses the resulting semivariance estimates to determine spatial weights and predict values at unobserved locations. This integration resulted in lower RMSEP values than those obtained using IDW and Gaussian-based Ordinary Kriging for both seismicity parameters, demonstrating that the improved representation of the semivariogram was also reflected in improved spatial prediction performance.

Table 5. Descriptive Statistics of SVR–OK Residuals for Seismicity Parameters

Parameters	Mean	Standard Deviation	Skewness	Kurtosis
Seismic Activity	-0.012	0.972	-0.069	3.170
Rock Fragility	-0.004	0.168	0.054	3.030

Building on this previously established SVR–Ordinary Kriging framework, the present study introduces an IDW-based residual correction to account for local spatial variation that remains unexplained by the baseline SVR–Ordinary Kriging predictions. This extension forms the basis of the proposed

Hybrid SVR–Ordinary Kriging–IDW methods evaluated in this study. The performance of the SVR–Ordinary Kriging (SVR–OK) model was evaluated using residual distribution analysis (Figure 3), descriptive statistics (Table 5), and Moran’s I residual test (Table 6) for the seismic activity (a-value) and rock fragility (b-value) parameters. These analyses were conducted to assess residual characteristics and determine whether spatial autocorrelation remained in the SVR–Ordinary Kriging residuals.

Table 6. Moran’s I Statistics for SVR–OK Residuals

Parameters	Moran I Statistics	p-value
Seismic Activity	0.174	0.032
Rock Fragility	0.158	0.045

Figure 3 and Table 5 show that the residuals of the SVR–Ordinary Kriging (SVR–OK) model were centered close to zero for both seismic activity (a-value) and rock fragility (b-value), indicating relatively small average prediction errors. However, the magnitude of residual variability differed between the two parameters. The a-value residuals showed substantially greater dispersion, with a standard deviation of 0.972, compared with 0.168 for the b-value residuals. For both parameters, skewness values were close to zero and kurtosis values were near 3, suggesting approximately symmetric residual distributions without substantial departures from normality. Although the residuals were centered near zero, their spatial distribution revealed an important limitation of the baseline model. As shown in Table 6, significant positive spatial autocorrelation remained in the residuals, with Moran’s I values of 0.174 ($p = 0.032$) for the a-value and 0.158 ($p = 0.045$) for the b-value. Thus, while SVR–OK produced residuals with small average errors, it did not completely account for the local spatial dependence in either seismicity parameter. The remaining positive spatial autocorrelation indicates that neighboring locations tended to exhibit residuals with similar values, suggesting the presence of local spatial variation that was not fully captured by the baseline model. This residual spatial structure therefore provides the methodological basis for introducing an IDW-based residual correction to account for the remaining local spatial variation and further improve spatial predictions.

The predictive performance of the hybrid interpolation models was evaluated using the Leave-One-Out Cross-Validation (LOOCV) method, with RMSE and MAE employed as the evaluation metrics for the seismic activity (a-value) and rock fragility (b-value) parameters. The comparative performance results of the SVR–OK, Hybrid SVR–OK–IDW Residual, and Hybrid SVR–OK–IDW Absolute Residual models are presented in Table 7.

The LOOCV and testing results in Table 7 and Table 8 show that the Hybrid SVR–Ordinary Kriging–IDW Residual model provided the best overall predictive performance. In LOOCV, this model achieved the lowest RMSE (0.984) and

Table 7. LOOCV Performance Metrics (RMSE and MAE) for Seismicity Parameters Across Interpolation Models

Model	Seismic Activity		Rock Fragility	
	RMSE	MAE	RMSE	MAE
SVR–Ordinary Kriging	1.005	0.752	0.179	0.131
Hybrid SVR–Ordinary Kriging–IDW Residual	0.984	0.723	0.181	0.131
Hybrid SVR–Ordinary Kriging–IDW Absolute Residual	1.258	1.001	0.221	0.178

Table 8. RMSEP Values on Testing Data for Seismicity Parameters Across Interpolation Models

Model	Seismic Activity		Rock Fragility	
	RMSEP	MAE	RMSEP	MAE
SVR–Ordinary Kriging	0.961	0.166		
Hybrid SVR–Ordinary Kriging–IDW Residual	0.837	0.149		
Hybrid SVR–Ordinary Kriging–IDW Absolute Residual	1.234	0.219		

MAE (0.723) for seismic activity (a-value). For rock fragility (b-value), SVR–Ordinary Kriging produced a slightly lower RMSE (0.179) than the Hybrid Residual model (0.181), while both models yielded the same MAE (0.131). On the testing data, the Hybrid Residual model consistently achieved the lowest RMSEP for both seismic activity (0.837) and rock fragility (0.149), compared with SVR–Ordinary Kriging (0.961 and 0.166) and Hybrid Absolute Residual (1.234 and 0.219), respectively. In contrast, the Hybrid Absolute Residual model showed higher prediction errors across both evaluation stages. Overall, these results demonstrate that local signed-residual correction improves the predictive performance of the SVR–Ordinary Kriging framework, particularly in terms of generalization to unseen data.

Table 9 presents the Shapiro–Wilk normality test results for the residuals of the SVR–Ordinary Kriging, Hybrid SVR–OK–IDW Residual, and Hybrid SVR–OK–IDW Absolute Residual models for the seismic activity parameter (a-value) and the rock fragility parameter (b-value). All models produced p-values greater than 0.05, indicating that the residuals satisfied the normality assumption and showed no significant deviation from a normal distribution.

Figure 4 shows that the residual distributions differ across the three interpolation models for both seismic activity (a-value) and rock fragility (b-value). The SVR–Ordinary Kriging and Hybrid SVR–OK–IDW Residual models produced residuals centered close to zero, indicating relatively balanced

Table 9. Shapiro–Wilk Normality Test Results for Residuals of Seismic Activity (a-value) and Rock Fragility (b-value) Across Interpolation Models

Model	Seismic Activity		Rock Fragility	
	W	p-value	W	p-value
SVR–Ordinary Kriging	0.985	0.785	0.986	0.820
Hybrid SVR–Ordinary Kriging–IDW Residual	0.987	0.861	0.990	0.950
Hybrid SVR–Ordinary Kriging–IDW Absolute Residual	0.979	0.539	0.979	0.532

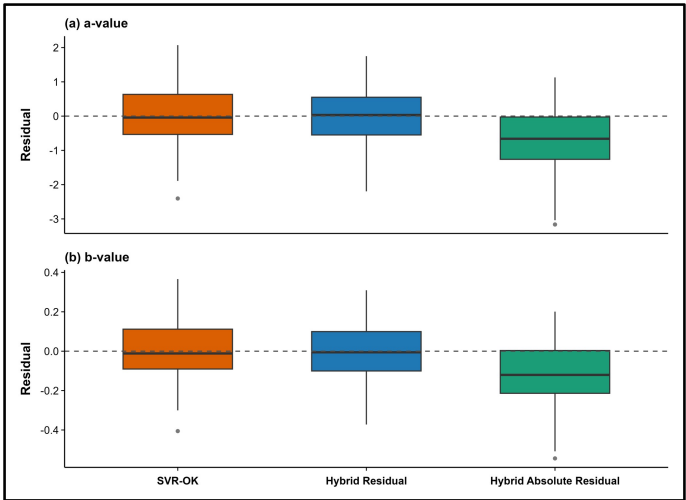


Figure 4. Boxplots of Residuals for the SVR–Ordinary Kriging and Hybrid SVR–OK–IDW Models for Seismic Activity (a-Value) and Rock Fragility (b-Value)

predictions. In contrast, the Hybrid SVR–OK–IDW Absolute Residual model showed a clear shift toward negative residuals for both parameters, indicating a systematic tendency to overestimate the observed values. Although its residual distribution appears relatively concentrated, this reduced dispersion is accompanied by substantial negative bias. Overall, the Hybrid Residual model provides a more balanced residual distribution, maintaining residuals near zero without the systematic overprediction observed in the Hybrid Absolute Residual model.

To further evaluate model performance, Moran’s I analysis was conducted across different numbers of nearest neighbors ($k = 3–10$) to assess the spatial independence of the residuals. As shown in Table 10, the SVR–Ordinary Kriging model exhibited significant positive residual spatial autocorrelation at $k = 3$ for both seismicity parameters ($p < 0.05$), indicating that local spatial dependence remained after the baseline interpolation. The Hybrid Residual model, however, showed no significant residual spatial autocorrelation across $k = 3–10$

Table 10. Moran’s I Residuals Across K (3–10) for SVR-OK, Hybrid-Res, and Hybrid-Abs Models of a-Value and b-Value.

Model	k	Seismic Activity		Rock Fragility	
		Moran I	p -value	Moran I	p -value
SVR–Ordinary Kriging	3	0.174	0.032	0.158	0.045
	4	0.119	0.060	0.111	0.072
	5	0.111	0.050	0.111	0.050
	6	0.085	0.071	0.084	0.074
	7	0.078	0.069	0.076	0.072
	8	0.040	0.163	0.038	0.169
	9	0.040	0.149	0.041	0.144
	10	0.040	0.131	0.042	0.125
Hybrid SVR–Ordinary Kriging–DW Residual	3	0.103	0.119	0.099	0.128
	4	0.056	0.195	0.059	0.189
	5	0.054	0.175	0.061	0.153
	6	0.037	0.212	0.042	0.193
	7	0.036	0.194	0.040	0.177
	8	-0.001	0.371	0.003	0.349
	9	0.008	0.310	0.013	0.281
	10	0.009	0.290	0.014	0.262
Hybrid SVR–Ordinary Kriging–DW Absolute Residual	3	0.248	0.005	0.263	0.004
	4	0.188	0.011	0.213	0.005
	5	0.170	0.008	0.205	0.002
	6	0.145	0.011	0.180	0.003
	7	0.120	0.017	0.154	0.004
	8	0.075	0.061	0.105	0.021
	9	0.076	0.049	0.108	0.014
	10	0.080	0.033	0.112	0.008

Table 11. Multifactor ANOVA Results for RMSE by Interpolation Method, Sample Size, and Response Value Pattern

Source of Variation	df	Sum of Squares	Mean Square	F-value	p-value
Method	3	400.520	133.510	117.670	0.000*
n	2	5586.030	2793.020	2461.780	0.000*
Pattern	1	421.880	421.880	371.850	0.000*
Method $\times n$	6	122.230	20.370	17.960	0.000*
Method $\times$ Pattern	3	253.730	84.580	74.550	0.000*
$n \times$ Pattern	2	658.990	329.500	290.420	0.000*
Method $\times n \times$ Pattern	6	22.550	3.760	3.310	0.000*
Residuals	2376	2695.690	1.130		

Note: Significant at $\alpha = 5\%$

for either the a-value or b-value ($p > 0.05$), suggesting that the signed residual correction effectively reduced the remaining local spatial structure. The Hybrid Absolute Residual model generally retained positive residual spatial autocorrelation, with statistical significance varying across neighborhood sizes and occurring in most cases. This remaining spatial dependence was accompanied by higher prediction errors and predominantly negative residuals, indicating a tendency toward overprediction. Taken together, these results demonstrate that the Hybrid

Residual model provided a more effective residual correction than the Hybrid Absolute Residual model. By preserving both the magnitude and direction of the prediction errors, the signed residual correction was better able to account for local spatial variation not captured by SVR–Ordinary Kriging, resulting in improved predictive performance and residuals with no statistically significant spatial autocorrelation across the evaluated neighborhood sizes.

Figure 5 presents the spatial distributions of the seismic

Table 12. Multifactor ANOVA Results for MAE by Interpolation Method, Sample Size, and Response Value Pattern

Source of Variation	df	Sum of Squares	Mean Square	F-value	p-value
Method	3	267.400	89.130	180.440	0.000*
n	2	2263.850	1131.930	2291.390	0.000*
Pattern	1	987.500	987.500	1999.030	0.000*
Method × n	6	88.830	14.810	29.970	0.000*
Method × Pattern	3	109.490	36.500	73.880	0.000*
n × Pattern	2	61.530	30.770	62.280	0.000*
Method × n × Pattern	6	9.560	1.590	3.220	0.004*
Residuals	2376	1173.720	0.490		

Note: Significant at $\alpha = 5\%$

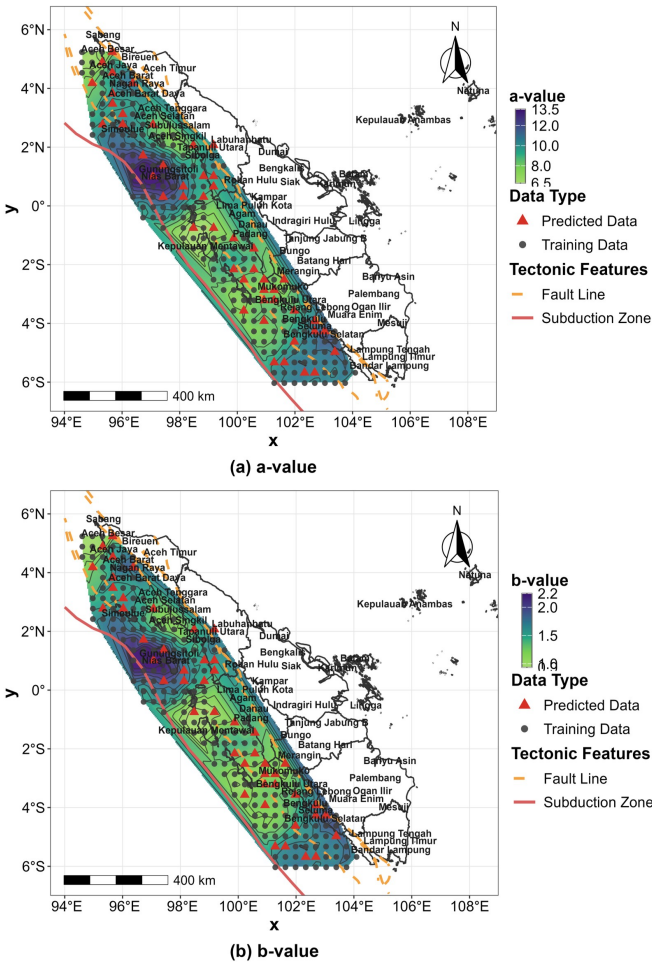


Figure 5. Spatial Distributions of (a) Seismic Activity (a-Value) and (b) Rock Fragility (b-Value) Estimated Using the Hybrid SVR–OK–IDW Residual Model Across Sumatra

activity parameter (a-value) and rock fragility parameter (b-value) estimated using the Hybrid SVR–OK–IDW Residual model across Sumatra. The resulting spatial patterns correspond to the major tectonic features of the region, including the Barisan Mountains, Sumatran Fault, forearc system,

Sunda Trench, and subducting Indo-Australian Plate (McCaffrey, 2009). High a-values were predominantly concentrated in the Nias–Simeulue segment and extended into parts of northern and central Sumatra, whereas lower values were observed in the Mentawai–Pagai–Enggano segment. The relatively high a-values in Nias–Simeulue indicate greater earthquake occurrence during the study period and are consistent with the active megathrust setting of this region, which has experienced major seismic events, including the 2005 Mw 8.7 Nias–Simeulue earthquake southeast of the 2004 Mw 9.1 Aceh–Andaman rupture (Hsu et al., 2006). In contrast, the lower a-values in the Mentawai–Pagai–Enggano segment indicate comparatively lower earthquake occurrence during the analyzed period. Nevertheless, this region has generated major earthquakes, including the 2007 Mw 8.4 and 2010 Mw 7.8 Mentawai events, as well as the historical great earthquakes of 1797 and 1833 (Qin et al., 2024). Therefore, low a-values should not be interpreted directly as evidence of low seismic hazard; rather, their tectonic implications should be considered together with recurrence behavior, fault coupling, geodetic deformation, and other independent evidence.

A comparable spatial pattern was observed for the b-value, with relatively high values in the Nias–Simeulue segment and lower values in the Mentawai–Pagai segment, demonstrating substantial spatial variability in rock fragility along the Sumatra megathrust. When considered jointly, the a-value and b-value distributions reveal distinct spatial variations among the major seismotectonic segments. Nias–Simeulue is characterized by relatively high a-values and b-values, indicating more frequent seismic activity with a relatively greater proportion of smaller-magnitude earthquakes, whereas Mentawai–Pagai exhibits comparatively low values of both parameters. Although low a-values and b-values in the Mentawai region have previously been discussed in relation to seismic gaps (Megawati and Pan, 2009; Yuliatmoko, 2020), the interpolated values alone cannot establish the existence of a seismic gap or indicate the potential occurrence of a future large earthquake. Such interpretations require independent supporting evidence from earthquake recurrence statistics, interseismic coupling, geodetic deformation, fault geometry, and prediction uncertainty. Accordingly, the interpolated surfaces in Figure 5 are interpreted

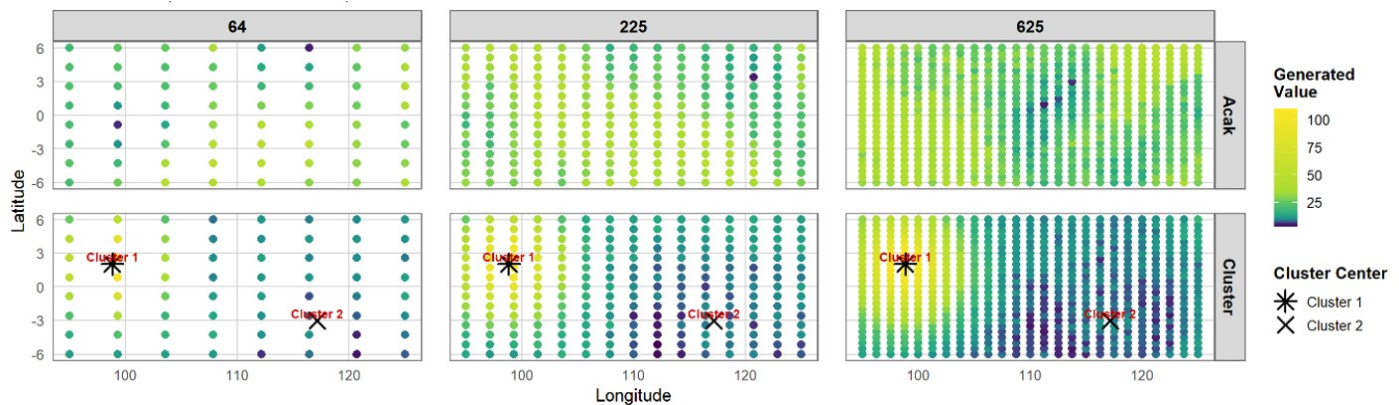


Figure 6. Simulated Data Based on the Theoretical Gaussian Semivariogram Model

primarily as representations of the spatial characteristics of seismicity parameters rather than as direct indicators of seismic hazard. The observed spatial variations are broadly consistent with the tectonic segmentation of the Sumatra megathrust (Rizal et al., 2023).

Slow slip events (SSEs) provide additional tectonic context for interpreting these spatial variations because aseismic slip may influence strain accumulation and stress transfer along different megathrust segments. Coral microatoll and geodetic observations have provided evidence of long-term or episodic aseismic deformation in several parts of the Sumatra subduction zone, including the Enggano and Banyak Islands (Dragert et al., 2001; Mallick et al., 2021; Obara, 2002). Long-duration SSEs may contribute to the redistribution of strain and stress between adjacent megathrust segments (Rousset et al., 2019). However, observations are not spatially or temporally uniform. SuGAR measurements showed no significant short-term SSE activity following the 2004, 2005, and 2007 megathrust earthquakes (Prawirodirdjo et al., 2010) whereas observations from the Banyak Islands suggest prolonged SSE activity (Tsang et al., 2015). Because geodetic measurements and temporal observations of SSEs were not incorporated as explanatory variables in the proposed interpolation framework, these studies are considered only as complementary tectonic context and should not be interpreted as direct validation of the interpolated a-value and b-value surfaces.

3.2 Simulation Results

To examine how sample size and spatial response patterns influence the underlying spatial structure, the simulated datasets were visually assessed through their spatial distributions and empirical semivariograms. Figure 6 presents representative realizations across the simulation scenarios, providing an initial characterization of the spatial patterns before evaluating the performance of the interpolation methods.

Figure 6 illustrates the simulated spatial response distributions across different sample sizes and spatial response patterns. As the sample size increased from $n=64$ to $n=625$, the denser spatial coverage provided a clearer representation of the un-

derlying spatial structure. The unclustered pattern exhibited relatively homogeneous spatial variation without distinct concentration zones, whereas the clustered pattern showed pronounced high- and low-response regions, reflecting stronger local heterogeneity. These structures became increasingly evident at larger sample sizes. The prediction errors of the spatial interpolation methods were further evaluated using the distributions of RMSE and MAE across simulation replications, as presented in Figure 7.

Figure 7 presents the RMSE and MAE distributions across 100 simulation replications. Increasing the sample size from $n=64$ to $n=625$ generally reduced prediction errors and variability. For the unclustered pattern, Gaussian-OK and SVR-OK consistently produced lower errors than the hybrid methods. Under the clustered pattern, Hybrid SVR-OK-IDW Residual performed best at $n=64$ and $n=225$, while Gaussian-OK and SVR-OK achieved lower errors at $n=625$. Hybrid SVR-OK-IDW Absolute Residual generally produced higher errors than the signed residual approach. Overall, interpolation performance depended on the method, sample size, and spatial response pattern, with residual correction being most beneficial for clustered patterns at small to moderate sample sizes.

Figure 8 shows that interpolation performance varied with sample size and spatial response pattern. At $n=64$ and $n=225$, Hybrid Residual achieved the lowest errors for clustered patterns, indicating its effectiveness under local heterogeneity and limited sampling density, while Gaussian-OK and SVR-OK performed better for unclustered patterns. At $n=625$, SVR-OK and Gaussian-OK achieved lower errors, suggesting reduced benefit from residual correction at high sampling density. Overall, Hybrid Residual was most effective for clustered patterns at small to moderate sample sizes, whereas Hybrid Absolute Residual showed no consistent advantage.

Residual spatial dependence was further assessed using Moran's I. Figure 9 compares the residual Moran's I distributions across methods, sample sizes, and spatial response patterns. Values close to zero indicate that little spatial structure remains in the prediction errors, suggesting that the underlying spatial dependence has been more effectively captured by the

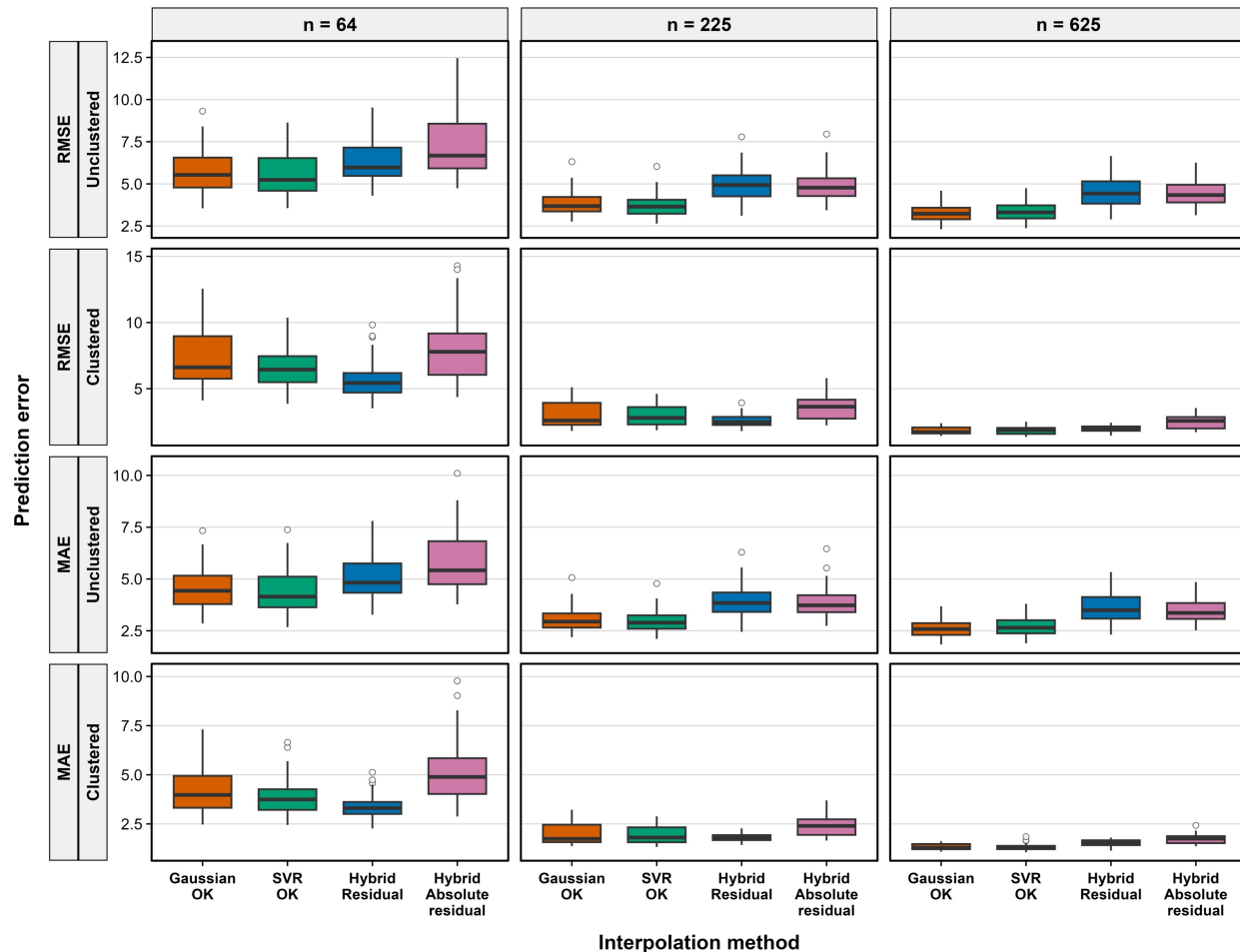


Figure 7. Boxplots of RMSE and MAE For Spatial Interpolation Methods Based on Response Value Patterns and Sample Sizes

interpolation method.

Figure 9 shows that residual spatial autocorrelation varied by spatial response pattern, sample size, and interpolation method. For unclustered patterns, Moran's I was generally negative, particularly at $n = 225$ and $n = 625$, with Gaussian-OK and SVR-OK closer to zero, suggesting that residual correction may induce spatial over-dispersion. In contrast, clustered patterns produced positive Moran's I for Gaussian-OK and SVR-OK, indicating remaining spatial dependence. Hybrid Residual effectively reduced this autocorrelation, although it occasionally shifted Moran's I toward negative values. Overall, residual correction was most beneficial when baseline residuals retained positive spatial autocorrelation, particularly under clustered conditions.

In addition to RMSE, MAE, and Moran's I, mean prediction error was evaluated to assess systematic tendencies toward over- or underprediction. Values close to zero indicate approximately unbiased predictions, whereas positive or negative values reflect systematic deviations from the observed responses. The distributions of mean prediction error across methods, sample sizes, and spatial response patterns are presented in

Figure 10.

Figure 10 shows that mean prediction error varied by spatial response pattern, sample size, and interpolation method. Gaussian-OK, SVR-OK, and Hybrid Residual generally produced median errors close to zero, with decreasing dispersion as sample size increased, indicating improved prediction stability. In contrast, Hybrid Absolute Residual consistently showed negative mean prediction errors, indicating systematic over-prediction, although the bias decreased with increasing sample size. Overall, larger samples improved prediction stability, while Hybrid Absolute Residual exhibited persistent negative bias.

To further assess whether these performance differences were statistically significant, a multifactor ANOVA was conducted using RMSE and MAE as response variables. The analysis evaluated the main effects of interpolation method, sample size, and spatial response pattern, as well as their interactions, to determine whether method performance depended on sampling density and spatial structure. The ANOVA results are presented in Table 11 and Table 12.

The multifactor ANOVA results presented in Table 11 and

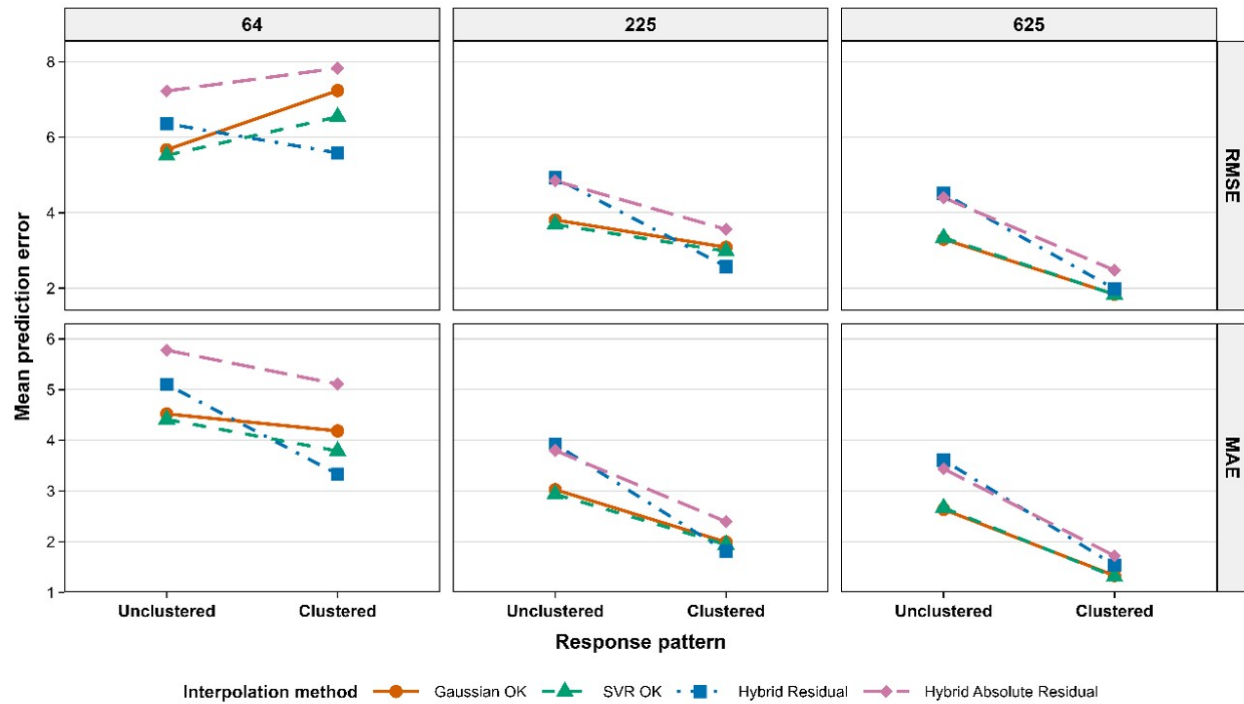


Figure 8. Comparison of Mean RMSE and MAE Across Interpolation Methods by Spatial Pattern and Sample Size

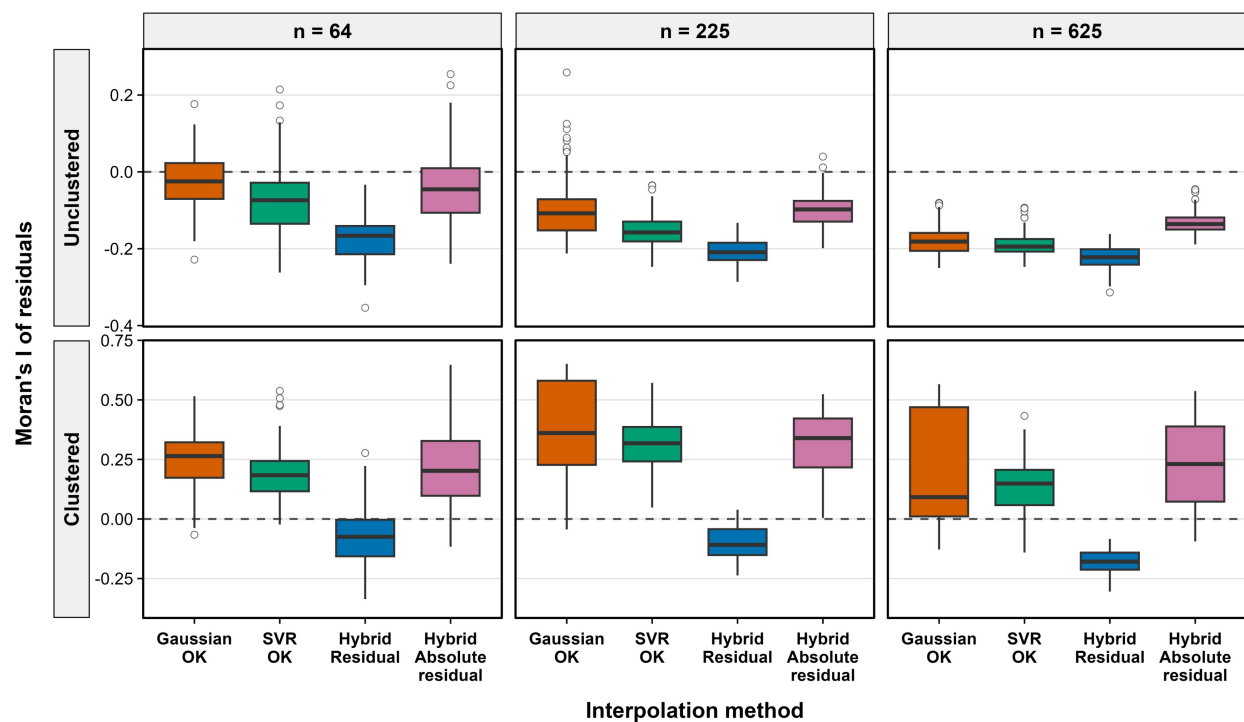


Figure 9. Boxplots of Residual Moran's I for Spatial Interpolation Methods by Response Value Pattern and Sample Size

Table 12 show that interpolation method, sample size, and spatial response pattern significantly affected both RMSE and

MAE ($p < 0.05$). All two-way interactions were also significant, indicating that the effect of one factor varied depending on the

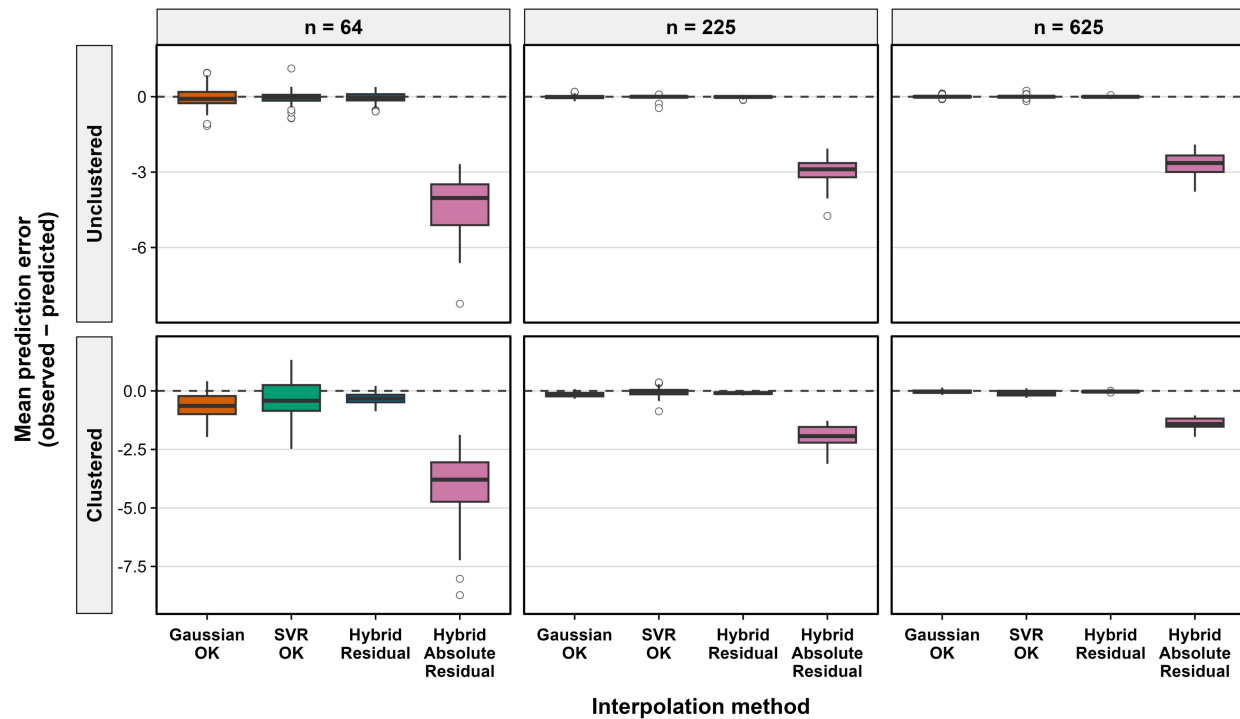


Figure 10. Distribution of Prediction Errors Across Interpolation Methods by Response Value Pattern and Sample Size

level of another factor. More importantly, the significant three-way interaction among interpolation method, sample size, and spatial response pattern indicates that differences in interpolation performance cannot be attributed to any single factor independently; rather, the relative performance of the methods depends jointly on sampling density and the underlying spatial structure. This interaction is reflected in the observed performance across the simulation scenarios. Increasing sample size generally reduced prediction errors and improved prediction stability, consistent with the expected benefit of greater spatial sampling density (Cressie, 1993; Li and Heap, 2011). For unclustered spatial patterns, Gaussian-OK and SVR-OK generally performed better than the hybrid methods, suggesting that additional residual correction was less beneficial when the spatial structure could already be adequately represented by the baseline interpolation models. For clustered patterns, however, Hybrid SVR-OK-IDW Residual was more effective at $n = 64$ and $n = 225$ and also reduced the remaining spatial dependence in the residuals, indicating that signed residual correction was particularly useful under heterogeneous spatial configurations with limited to moderate sampling density. At $n = 625$, Gaussian-OK and SVR-OK performed better, suggesting that the additional benefit of residual correction diminished as sampling density increased and more spatial information became available to the baseline models. In contrast, Hybrid SVR-OK-IDW Absolute Residual showed no consistent improvement and tended toward overprediction. Overall, these findings demonstrate that no single interpolation

method performs uniformly best across all simulation conditions. SVR-OK is particularly relevant when the empirical semivariogram exhibits complex or non-monotonic behavior, whereas Hybrid SVR-OK-IDW Residual provides greater benefit when residuals from the baseline model retain positive local spatial dependence, particularly for clustered spatial patterns with small to moderate sample sizes.

4. CONCLUSIONS

This study developed a hybrid spatial interpolation framework integrating Support Vector Regression–Ordinary Kriging (SVR-OK) with IDW-based residual correction using signed and absolute residual formulations for predicting seismicity parameters across Sumatra. The empirical results showed that the Hybrid SVR-OK-IDW Residual model provided the best predictive performance for both seismic activity (a-value) and rock fragility (b-value), while also reducing the remaining residual spatial autocorrelation. In contrast, the Hybrid SVR-OK-IDW Absolute Residual model tended to overpredict and produced higher prediction errors. The simulation results further demonstrated that interpolation performance depended on spatial response pattern and sample size. Gaussian-OK and SVR-OK performed better for unclustered patterns, whereas the Hybrid Residual model was more effective for clustered patterns at small to moderate sample sizes; however, its advantage diminished as sampling density increased. Overall, these findings indicate that residual correction is particularly beneficial when the baseline model retains positive local spatial dependence

and should therefore be applied according to the characteristics of the spatial data rather than universally across all conditions. Future studies should explore alternative machine-learning approaches for semivariogram modeling, incorporate auxiliary geophysical information and spatially blocked validation, and develop location-specific prediction intervals or spatial uncertainty measures for the complete hybrid framework. Extending the framework to spatio-temporal modeling may further improve prediction reliability and the characterization of seismicity dynamics.

5. ACKNOWLEDGEMENT

The authors acknowledge the Indonesian Agency for Meteorology, Climatology and Geophysics (BMKG) for providing the earthquake dataset used in this study. We also thank the School of Data Science, Mathematics and Informatics, IPB University, Indonesia, for its academic and research support. This research was financially supported by the Center for Higher Education Funding and Assessment (PPAPT) and the Indonesian Endowment Fund for Education (LPDP), Ministry of Finance of the Republic of Indonesia, under the Indonesian Education Scholarship (BPI ID: 202231103733).

REFERENCES

- Arimuko, A., S. Rohadi, and A. S. Rahman (2023). Seismotectonic Studies to Determine the Recurrence of Earthquakes $M_w > 7$ Using a Statistical Approach and Plate Motion in the Megathrust Western Part of Java. *Geotechnical and Geological Engineering*, **41**(2); 1397–1406
- Chasanah, U. and E. Handoyo (2021). Determination the Magnitude of Completeness, b-Value and a-Value for Seismicity Analysis in East Java, Indonesia. *Journal of Physics: Conference Series*, **1805**(1); 012009
- Chen, L., M. Ma, M. Wu, W. Pedrycz, and K. Hirota (2025). Extended Multi-Kernel Relevance Vector Machine Optimized Kriging Interpolation for Coal Seam Thickness Prediction in Coal-Bearing Strata. *Engineering Applications of Artificial Intelligence*, **145**; 110093
- Cressie, N. A. C. (1993). *Statistics for Spatial Data*. John Wiley & Sons, revised edition edition
- de Mattos Neto, P. S. G., G. D. C. Cavalcanti, D. S. Domingos, and E. G. Silva (2022). Hybrid Systems Using Residual Modeling for Sea Surface Temperature Forecasting. *Scientific Reports*, **12**(1)
- Ding, X., J. Liu, F. Yang, and J. Cao (2021). Random Radial Basis Function Kernel-Based Support Vector Machine. *Journal of the Franklin Institute*, **358**(18); 10121–10140
- Dragert, H., K. Wang, and T. S. James (2001). A Silent Slip Event on the Deeper Cascadia Subduction Interface. *Science*, **292**(5521); 1525–1528
- Elen, A., S. Baş, and C. Közkurt (2022). An Adaptive Gaussian Kernel for Support Vector Machine. *Arabian Journal for Science and Engineering*, **47**(8); 10579–10588
- Erdogan Erten, G., M. Yavuz, and C. V. Deutsch (2022). Combination of Machine Learning and Kriging for Spatial Estimation of Geological Attributes. *Natural Resources Research*, **31**(1); 191–213
- Galina, N. A., V. V. Bykova, R. N. Vakarchuk, and R. E. Tatevosian (2019). Effect of Earthquake Catalog Declustering on Seismic Hazard Assessment. *Seismic Instruments*, **55**(1); 59–69
- García-Díaz, Noel and Verduzo-Ramirez, Alberto and Garcia-Virgen, Juan and Muñoz, Lilia (2016). Applying Absolute Residuals as Evaluation Criterion for Estimating the Development Time of Software Projects by Means of a Neuro-Fuzzy Approach. *Journal of Information Systems Engineering & Management*, **1**(4)
- Gutenberg, B. and C. F. Richter (1944). Frequency of Earthquakes in California. *Bulletin of the Seismological Society of America*, **34**(4); 185–188
- Han, C., J. Wang, M. Zheng, E. Wang, J. Xia, G. S. Li, and S. Choe (2016). New Variogram Modeling Method Using MGPP and SVR. *Earth Science Informatics*, **9**(2); 197–213
- Hengl, T., G. B. M. Heuvelink, and D. G. Rossiter (2007). About Regression-Kriging: From Equations to Case Studies. *Computers & Geosciences*, **33**(10); 1301–1315
- Hsu, Y. J., M. Simons, J. P. Avouac, J. Galetka, K. Sieh, M. Chlieh, D. Natawidjaja, L. Prawirodirdjo, and Y. Bock (2006). Frictional Afterslip Following the 2005 Nias-Simeulue Earthquake, Sumatra. *Science*, **312**(5782); 1921–1926
- Huang, Z., H. Wang, and R. Zhang (2012). An Improved Kriging Interpolation Technique Based on SVM and Its Recovery Experiment in Oceanic Missing Data. *American Journal of Computational Mathematics*, **2012**(1); 56–60
- Li, J. (2022). *Spatial Predictive Modelling with R*. CRC Press
- Li, J. and A. D. Heap (2011). A Review of Comparative Studies of Spatial Interpolation Methods in Environmental Sciences: Performance and Impact Factors. *Ecological Informatics*, **6**(3–4); 228–241
- Li, J., A. Potter, Z. Huang, and A. D. Heap (2011). Predicting Seabed Mud Content across the Australian Margin II: The Performance of Machine Learning Methods and Their Combinations with Ordinary Kriging or Inverse Distance Squared. ResearchGate
- Ma, N., Y. Bai, S. Meng, A. De Santis, V. Kossobokov, D. Ouzounov, and S. Pulinet (2021). Return Period Evaluation of the Largest Possible Earthquake Magnitudes in Mainland China Based on Extreme Value Theory. *Sensors*, **21**(10); 3519
- Mallick, R., A. J. Meltzner, L. L. H. Tsang, E. O. Lindsey, L. Feng, and E. M. Hill (2021). Long-Lived Shallow Slow-Slip Events on the Sunda Megathrust. *Nature Geoscience*, **14**(5); 327–333
- McCaffrey, R. (2009). The Tectonic Framework of the Sumatran Subduction Zone. *Annual Review of Earth and Planetary Sciences*, **37**; 345–366
- Megawati, K. and T. C. Pan (2009). Regional Seismic Hazard Posed by the Mentawai Segment of the Sumatran Megathrust

- rust. *Bulletin of the Seismological Society of America*, **99**(2A); 566–584
- Montesinos López, O. A., A. Montesinos López, and J. Crossa (2022). Support Vector Machines and Support Vector Regression. In *Multivariate Statistical Machine Learning Methods for Genomic Prediction*. Springer, Cham
- Obara, K. (2002). Nonvolcanic Deep Tremor Associated with Subduction in Southwest Japan. *Science*, **296**(5573); 1679–1681
- Palacios, P., I. Molina, and M. Segovia (2006). The Gutenberg–Richter Law: Assumptions, Limitations and Interpretations. In *Statistics in Volcanology*. pages 115–127
- Prawirodirdjo, L., R. McCaffrey, C. D. Chadwell, Y. Bock, and C. Subarya (2010). Geodetic Observations of an Earthquake Cycle at the Sumatra Subduction Zone: Role of Interseismic Strain Segmentation. *Journal of Geophysical Research: Solid Earth*, **115**(B3)
- Putra, A. S., D. P. Sahara, A. D. Nugraha, N. T. Puspito, P. Supendi, and N. Heryandoko (2023). Segmentation of the Sumatran Fault Zone Based on Spatial Variation of b-Values. *Journal of Seismology*, **27**(5); 919–932
- Qin, Y., J. Chen, S. C. Singh, N. Hananto, H. Carton, and P. Tapponnier (2024). Assessing the Risk of Potential Tsunamigenic Earthquakes in the Mentawai Region by Seismic Imaging, Central Sumatra. *Geochemistry, Geophysics, Geosystems*, **25**(5); e2023GC011149
- Ramos, Y., B. St-Onge, J. P. Blanchet, and A. Smargiassi (2015). Spatio-Temporal Models to Estimate Daily Concentrations of Fine Particulate Matter in Montreal: Kriging with External Drift and Inverse Distance-Weighted Approaches. *Journal of Exposure Science & Environmental Epidemiology*, **26**(4); 405–414
- Rizal, J., A. Y. Gunawan, S. W. Indratno, and I. Meilano (2023). Seismic Activity Analysis of Five Major Earthquake Source Segments in the Sumatra Megathrust Zone: Each Segment and Two Adjacent Segments Points of View. *Bulletin of the New Zealand Society for Earthquake Engineering*, **56**(2); 55–70
- Rousset, B., R. Bürgmann, and M. Campillo (2019). Slow Slip Events in the Roots of the San Andreas Fault. *Science Advances*, **5**(2)
- Seo, Y., S. Kim, and V. P. Singh (2015). Estimating Spatial Precipitation Using Regression Kriging and Artificial Neural Network Residual Kriging (RKNNRK) Hybrid Approach. *Water Resources Management*, **29**(7); 2189–2204
- Shepperd, M. and S. MacDonell (2012). Evaluating Prediction Systems in Software Project Estimation. *Information and Software Technology*, **54**(8); 820–827
- Sirodj, D. A. N., M. N. Aidi, B. Sartono, U. D. Syafitri, and B. Pranata (2025a). Identification of Earthquake Prone Zones in Sumatra Using Density Based Spatial Clustering of Applications with Noise. *CAUCHY: Jurnal Matematika Murni dan Aplikasi*, **10**(2); 916–928
- Sirodj, D. A. N., M. N. Aidi, B. Sartono, U. D. Syafitri, and B. Pranata (2025b). Integrating Support Vector Regression and Kriging in Spatial Interpolation of Statistical Seismicity Parameters. *Indonesian Journal of Geography*, **57**(3); 463–470
- Tsang, L. L. H., A. J. Meltzner, B. Philipposian, E. M. Hill, J. T. Freymueller, and K. Sieh (2015). A 15 Year Slow-Slip Event on the Sunda Megathrust Offshore Sumatra. *Geophysical Research Letters*, **42**(16); 6630–6638
- Van Beers, W. C. M. and J. P. C. Kleijnen (2003). Kriging for Interpolation in Random Simulation. *Journal of the Operational Research Society*, **54**(3); 255–262
- Wu, C. H., G. H. Tzeng, and R. H. Lin (2009). A Novel Hybrid Genetic Algorithm for Kernel Function and Parameter Optimization in Support Vector Regression. *Expert Systems with Applications*, **36**(3); 4725–4735
- Yuliatmoko, R. S. (2020). Analisis Seismic GAP Sumatera Menggunakan Anomali Gravitasi. *Megasains*, **11**(1); 24–30. (in Indonesia)