SEISMIC SITE QUALITY ASSESSMENT IN NORTH SUMATRA USING SPECTRAL DENSITY ANALYSIS AND MACHINE LEARNING-BASED CLUSTERING

Triya Fachriyeni[1,2], Katherin Indriawati[1*], Kevin W. Pakpahan[3], Irfan Rifani[4], Anne M. M. Sirait[5], Yusran Asnawi[6], Hendro Nugroho[2], Andrean V. H. Simanjuntak[2]

[1]Department of Engineering Physics, Faculty of Industrial Technology and System Engineering, Institute Technology of Sepuluh Nopember, Surabaya 44511, Indonesia
[2]Indonesian Agency for Meteorology Climatology and Geophysics (BMKG), Kemayoran 23111, Indonesia
[3]Indonesian State School of Meteorology Climatology and Geophysics (STMKG), Tangerang 15119, Indonesia
[4]Department of Geographic Education, Universitas Negeri Manado, Tondano 95618, Indonesia
[5]Geophysics Program, Faculty of Mathematics and Natural Sciences, University of Indonesia, Depok 16421, Indonesia
[6]Department of Physics Education, State Islamic University Ar-Raniry, Banda Aceh 23111, Indonesia
e-mail : katherin@its.ac.id

Abstract. *Seismic noise strongly influences the accuracy and reliability of earthquake monitoring, particularly in tectonically active regions such as North Sumatra. This study investigates the quality of seismic stations by analyzing noise characteristics using Power Spectral Density (PSD), Probability Density Functions (PDFs), and machine learning clustering. PSD was computed through the Fast Fourier Transform (FFT) and compared against the New High Noise Model (NHNM) and New Low Noise Model (NLNM) benchmarks. Noise variability was further quantified using PDFs, while fuzzy c-means (FCM) clustering was applied to classify temporal noise patterns. Results from the MUTSI seismic station demonstrate strong diurnal and weekly variability, with horizontal components (SHE and SHN) exhibiting significantly higher noise levels and fluctuations than the vertical component (SHZ). Noise amplitudes peaked during morning hours (06:00–09:00 UTC), correlating with anthropogenic activity, and decreased substantially at night, indicating that optimal recording conditions occur during late evening to early morning. FCM clustering identified five dominant noise regimes, separating stable low-noise baselines from sporadic high-noise anomalies likely associated with human activity or instrumental disturbances. These findings highlight the importance of integrating spectral analysis with clustering techniques to evaluate seismic station performance, improve real-time monitoring, and guide optimal site selection and operational scheduling for earthquake detection.*

Keywords: *fuzzy c-means; machine learning; North Sumatra; seismic noise; spectral density*

INTRODUCTION

Sumatra Island, located on the western side of Indonesia, lies within an active seismic zone. This region has a complex tectonic system consisting of the Sumatra Fault and the subduction zone, caused by the convergence of the Indo-Australian and Eurasian plates. Such tectonic activity results in very high seismicity. As of 2024, the Meteorology, Climatology, and Geophysics Agency (BMKG) has operated 512 seismic sensors across Indonesia, with around 20% located in Sumatra. Historically, several damaging earthquakes have occurred within the last 50 years. Therefore, the accuracy of earthquake information delivered to the public is highly dependent on the quality of seismic signals.

However, seismic signal recordings always contain noise (Stutzmann et al., 2012; Parolai et al., 2006; Anthony et al., 2020). Seismic noise refers to unwanted signal components recorded by seismometers (Johnson et al., 2020). Physically, seismic noise primarily originates from surface or near-surface sources and mostly consists of elastic surface waves (Stutzmann et al., 2012). Seismic noise sources can be classified into two categories. The first is natural noise, generated by tectonic earthquakes, volcanic earthquakes, and rockfalls (Johnson et al., 2020). The second is anthropogenic (man-made) noise, also known as environmental noise. This type of noise arises from human activities near or on the Earth's surface, such as underground explosions, drilling, mining, industrial operations, and traffic.

Environmental noise propagates as surface waves in the high-frequency range (5-10 Hz), decaying with distance (Zhong et al., 2021; Ke et al., 2023). Digital seismic data allows spectral analysis to be performed using relatively simple operations. This method involves examining the frequency domain within seismic

DOI : http://dx.doi.org/10.12962/j25023659.v11i3.8972

Received : 30-09-2025 Reviewed : 10-10-2025 Accepted : 19-12-2025 Available Online : 31-12-2025

noise. In addition, seismic noise has practical applications, including estimating seismic microzonation maps by characterizing local and regional site responses during earthquakes (Tchawe et al., 2020; Janiszewski et al., 2023). The quality of seismic stations reflects evaluations of the sites where seismometers are installed (Birnie et al., 2021). These evaluations are fixed and do not change over time.

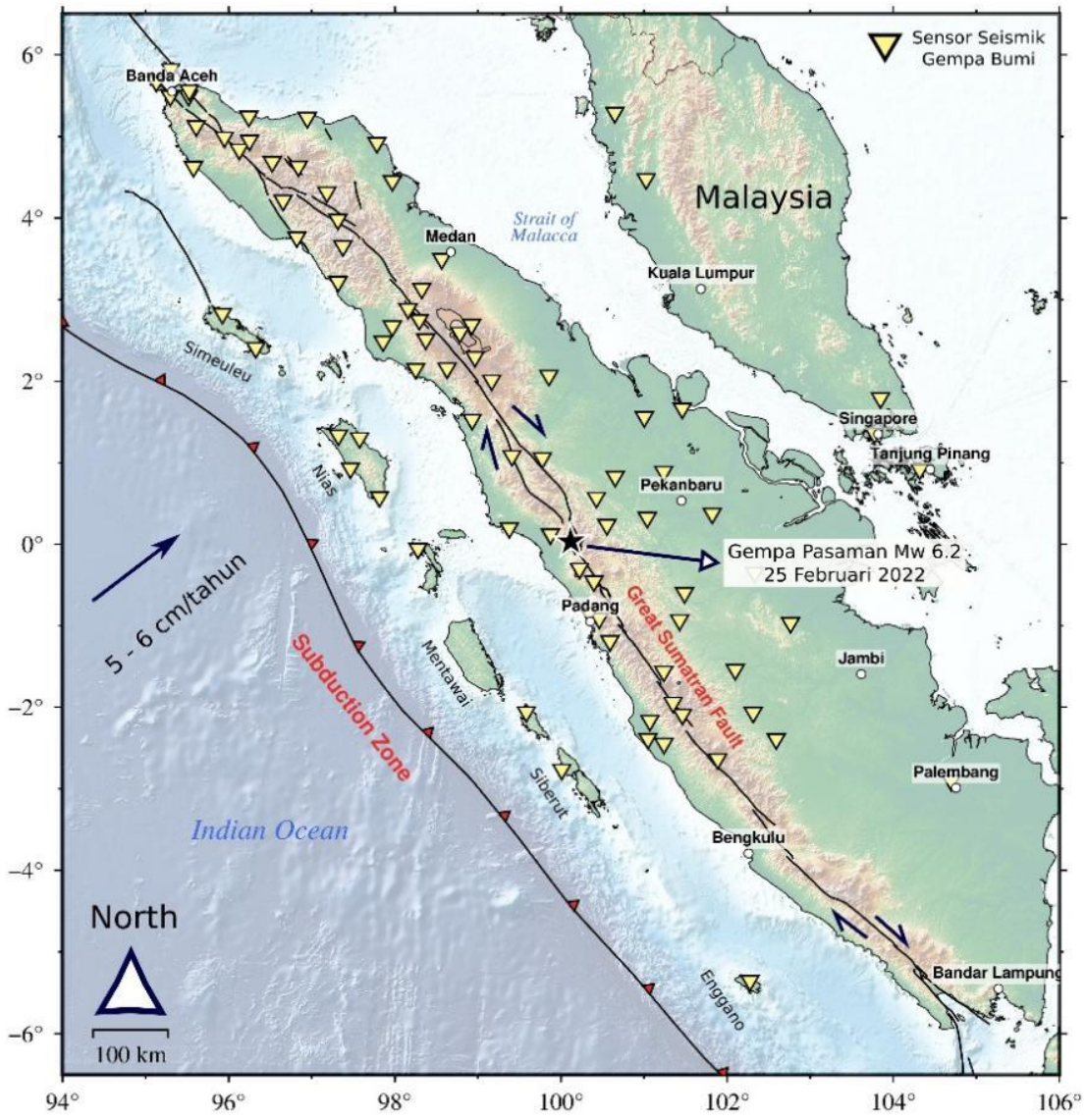

Figure 1. Distribution of seismic sensors across Sumatra Island and the epicenter location of the 2022 Pasaman earthquake. This map provides spatial context for station placement and highlights the regional tectonic setting relevant to noise analysis.

One key parameter is geology, assessed by classifying sensor site conditions according to rock age. This makes it possible to analyze geological conditions for each seismic sensor site in North Sumatra. Therefore, this study evaluates seismic signal quality by considering both natural noise levels and subsurface geological models of seismic sensor sites. The results of this research are expected to significantly improve earthquake monitoring, enabling earthquake information to be delivered more quickly, accurately, and reliably.

DOI : http://dx.doi.org/10.12962/j25023659.v11i3.8972

Received : 30-09-2025 Reviewed : 10-10-2025 Accepted : 19-12-2025 Available Online : 31-12-2025

METHODOLOGY

Data Acquisition

Seismic data were obtained from the MUTSI seismic station in North Sumatra, operated by the Meteorology, Climatology, and Geophysics Agency (BMKG). The dataset consisted of continuous digital waveform recordings across three components: east-west horizontal (SHE), north-south horizontal (SHN), and vertical (SHZ). Data were collected in counts and converted to physical units as $m^2/s^4/Hz$, representing digitized ground motion amplitudes, and analyzed primarily within the frequency range of 1-10 Hz, which is highly sensitive to both local seismic events and anthropogenic disturbances.

The dataset used in this study consists of 30 days of continuous broadband seismic with 40 Hz sampling rate for all three components (SHZ, SHN, and SHE). Prior to processing, the raw waveforms were subjected to a comprehensive quality-control workflow. First, the continuity of the time series was evaluated to ensure zero data gaps, and any segments containing gaps longer than 0.5 seconds were removed. Second, amplitude stability and instrument performance were assessed to eliminate intervals affected by clipping, timing drift, or sensor saturation (Simanjuntak & Jihad, 2017; Maryana et al., 2021). Third, signal quality was measured using a signal-to-noise ratio (SNR) threshold of 3 or higher within the 1-10 Hz frequency band to retain only reliable segments for spectral analysis (e.g Simanjuntak et al., 2021; Sirait et al., 2025). Segments with SNR values below 3 or with spectral anomalies exceeding three standard deviations from the median Power Spectral Density (PSD) curve were excluded. These procedures ensured that the final dataset consisted of high-quality broadband recordings suitable for accurate PSD estimation and machine-learning-based clustering.

In addition, geological information was integrated by classifying station site conditions according to rock age, allowing correlation between subsurface structure and noise characteristics.Power Spectral Density (PSD) analysis is used to evaluate the noise characteristics of seismic stations. The standardization of seismic noise levels, whether high or low, refers to Peterson's (1983) model, namely the New High Noise Model (NHNM) and the New Low Noise Model (NLNM). The PSD was computed via the Fast Fourier Transform (FFT), which decomposes time-domain seismic signals into their frequency components (Cooley & Tukey, 1965).

This approach enables efficient identification of frequency-dependent noise characteristics. Noise evaluation was further supported by geological characterization of the seismic station sites. Geological conditions were classified based on rock age, enabling correlation between subsurface structure and noise levels. This integration allows site-specific interpretation of seismic signal quality. The PSD results were interpreted relative to three main noise bands (McNamara et al., 2009):

1. Short-period noise: 0.1- 1 s (high-frequency range)
2. Microseismic noise: 2 - 20 s (intermediate range)
3. Long-period noise: 20 - 900 s (low-frequency range)

Additional anomalies such as instrumental disturbances (e.g., sensor malfunctions, digitizer errors, or power supply instability) were also considered. Noise originating from short-period characteristics (0.1 - 1 s), microseismic characteristics (2 - 20 s), and long-period characteristics (period 20 - 900 s) (McNamara et al., 2009). Noise originating from other causes such as sensor disturbances, digitizer issues, power supply problems, and so forth. The method used to analyze the level of seismic noise is PSD. Referring to the studies of Peterson (1983) and McNamara & Buland (2004), PSD is calculated using the direct Fourier transform method (Cooley & Tukey, 1965). This method employs the Fast Fourier Transform (FFT) to identify frequency components contained within seismic noise signals in the time domain over a limited range, thus providing efficiency in computation.

In addition to analyzing PSD values, statistical analysis is also carried out by calculating the Probability Density Function (PDF) of seismic signals for all PSD results. The probability density function is obtained from a random variable as a function that represents each member of the sample space S in real numbers. The PSD solution is considered as a function $f(x)$, which is the probability distribution function of X for continuous random variables. The PDF of a continuous random variable X must satisfy the following properties:

DOI : http://dx.doi.org/10.12962/j25023659.v11i3.8972

Received : 30-09-2025 Reviewed : 10-10-2025 Accepted : 19-12-2025 Available Online : 31-12-2025

1. $0 \leq fX(x) \leq 1$, meaning the PDF ranges from zero to one.
2. $\int x\, fX(x)\, dx = 1$, ameaning the PDF of random variable X over the sample space equals one.
3. $P(a \leq X \leq b) = \int xb = a\, fX(x)\, dx$, where a and b are arbitrary values of X such that $a < b$.
4. $P(X \leq x) = P(X < x)$, meaning that the inequality symbols (< and ≤, or > and ≥) are considered equivalent.

Procedurally, The Power Spectral Density (PSD) estimates in this study were computed using a windowed Fast Fourier Transform (FFT) following standard procedures for ambient seismic noise analysis. A Hanning window with a length of 1024 samples (approximately 25.6 seconds at a 40 Hz sampling rate) was applied, together with a 50 percent overlap between adjacent windows to improve spectral stability. Each PSD estimate was obtained by averaging a minimum of 10 overlapped segments to reduce variance in the spectral estimates, consistent with recommendations for noise characterization in continuous broadband data.

Machine Learning Clustering

To further classify temporal patterns of seismic noise, we employed Fuzzy C-Means (FCM) clustering (Bezdek et al., 1984). Unlike hard clustering methods such as k-means, FCM allows each data point to have partial membership across multiple clusters, providing greater flexibility in representing seismic noise, which often exhibits ambiguous and overlapping characteristics (Nuranna et al., 2021). Five clusters were identified, corresponding to baseline low-noise conditions, moderate anthropogenic noise, and sporadic high-noise anomalies. Temporal distributions of these clusters were analyzed on both daily and weekly scales to evaluate the influence of human activity and environmental factors.

For temporal pattern classification, the fuzzy c-means (FCM) algorithm was implemented with five clusters (k = 5), selected empirically based on the temporal variability observed in the PSD and RMS distributions. The fuzziness coefficient was set to m = 2.0, a standard value widely used in seismological clustering applications to balance cluster separation and membership ambiguity. The algorithm was run with a convergence threshold of 10^{-5} and a maximum of 300 iterations. These parameter choices provide a stable, reproducible framework for capturing the dominant noise regimes at the MUTSI station while avoiding overfitting.

RESULTS AND DISCUSSION

We successfully analyze all responsible station that influenced by the noise by hourly analysis as show in Figure 2. The hourly noise illustrates the noise pattern over a 24-hour period at the MUTSI seismic station. The plot includes all three component channels: east-west horizontal (SHE), north-south horizontal (SHN), and vertical (SHZ), recorded on January 5, 2025 (Sunday). This visualization provides insights into the distribution of seismic noise amplitude, where the radial axis indicates amplitude and the angular axis represents time (00-23 hours). From this analysis, we observed that the horizontal components (SHE and SHN) experienced greater noise fluctuations than the vertical component (SHZ), as indicated by the longer radii of the blue and green curves compared to the red curve.

The highest noise amplitude occurred between 06:00-07:00 on the SHE component (blue), most likely associated with anthropogenic activities such as morning traffic or increased human activity before working hours. Meanwhile, the lowest noise amplitude generally occurred at night until early morning (around 00:00-05:00 and 18:00-23:00), indicating that the environment during these hours was relatively quiet. The SHZ (vertical) component showed higher stability throughout the day, consistent with studies stating that vertical components are more resistant to surface disturbances, as explained by McNamara & Buland (2004) in the context of background noise in global broadband seismic networks.

DOI : http://dx.doi.org/10.12962/j25023659.v11i3.89

Received : 30-09-2025 Reviewed : 10-10-2025 Accepted : 19-12-2025 Available Online : 31-12-2025

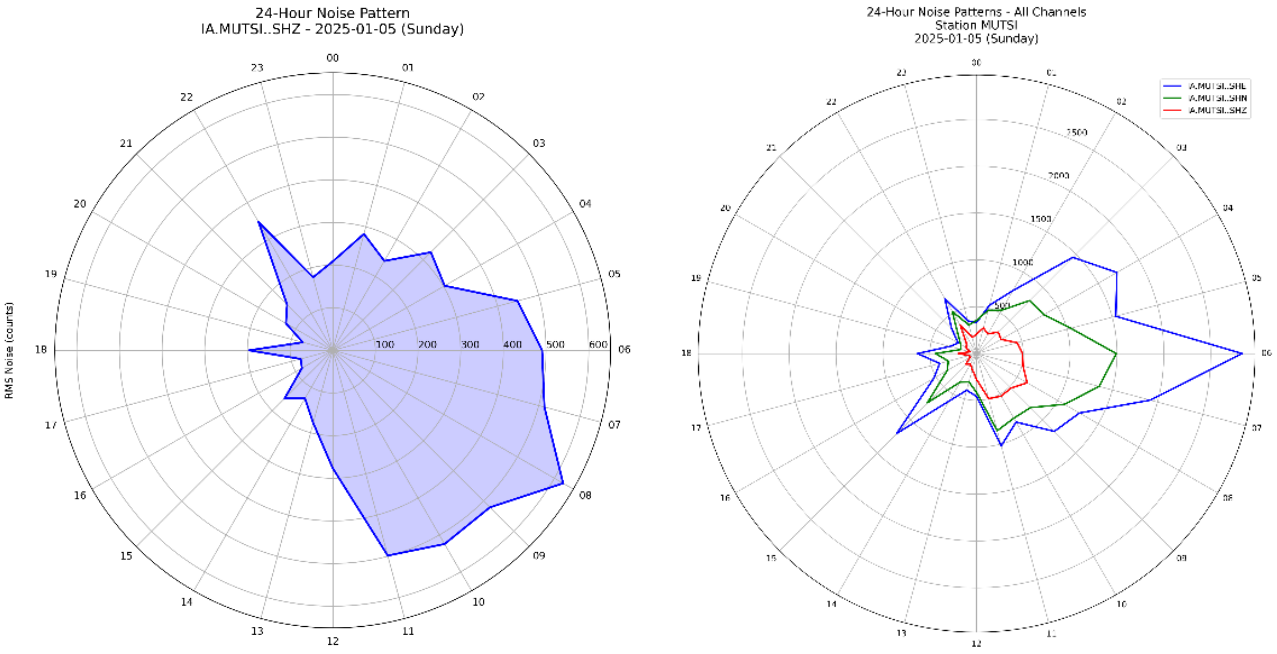

Figure 2. Example of daily polar plots of seismic noise amplitude at the MUTSI station (a) Vertical component (SHZ) over a 24-hour period. (b) Combined east-west horizontal (SHE), north-south horizontal (SHN), and vertical (SHZ) components. The radial axis represents amplitude, while the angular axis corresponds to time (00-23 hours).

High and fluctuating noise patterns, especially in the horizontal components, can reduce the signal-to-noise ratio (SNR), making it difficult to detect small waves from micro-earthquakes or distant earthquakes. By understanding daily noise patterns like this, station operators and researchers can select optimal times for high-quality waveform analysis and design more effective frequency filtering and signal preprocessing strategies. Overall, this graph shows that noise activity is strongly influenced by local environmental factors and time, reinforcing the importance of noise monitoring in managing seismic networks.

The highest noise amplitude occurred between 06:00 and 09:00 UTC, with RMS values reaching around 600 $m^2/s^4/Hz$. This indicates that the station experienced increased seismic disturbances in the morning, generally associated with anthropogenic activities such as traffic, industrial machinery, or local community activities. Noise amplitude decreased significantly after 10:00, particularly in the afternoon to evening, between 17:00 and 20:00 UTC, with RMS dropping to around 100-150 $m^2/s^4/Hz$. The vertical component (SHZ) was more stable against noise compared to the horizontal components, as it is more resistant to lateral disturbances. Surface waves (Rayleigh waves) can increase noise, especially if sensors are located in open or soft-soil areas. From a data quality perspective, this pattern shows that the best time to obtain clean seismic signals with a high signal-to-noise ratio (SNR) is at night when noise amplitudes are low.

The two graphs presented technically show the statistical daily distribution and trends of seismic noise levels (RMS noise) based on the three recording components of the seismometer at the MUTSI station: east-west horizontal (SHE), north-south horizontal (SHN), and vertical (SHZ). The analyzed noise values are represented in digital count units, converted from ground vibration amplitudes by the digitizer system. It is assumed that RMS noise values are measured in the frequency range of about 1-10 Hz, a band sensitive to both local seismic waves and anthropogenic noise (McNamara & Buland, 2004; Bensen et al., 2007). Thus, this analysis is highly relevant for evaluating signal quality for detecting small earthquakes and early warning systems. The first graph (boxplot) shows that statistically, the SHE component has a very wide RMS noise distribution and is prone to extreme outliers, with maximum values exceeding 40,000 $m^2/s^4/Hz$ at certain hours (especially between 02:00-06:00 UTC).

DOI : http://dx.doi.org/10.12962/j25023659.v11i3.8972

Received : 30-09-2025 Reviewed : 10-10-2025 Accepted : 19-12-2025 Available Online : 31-12-2025

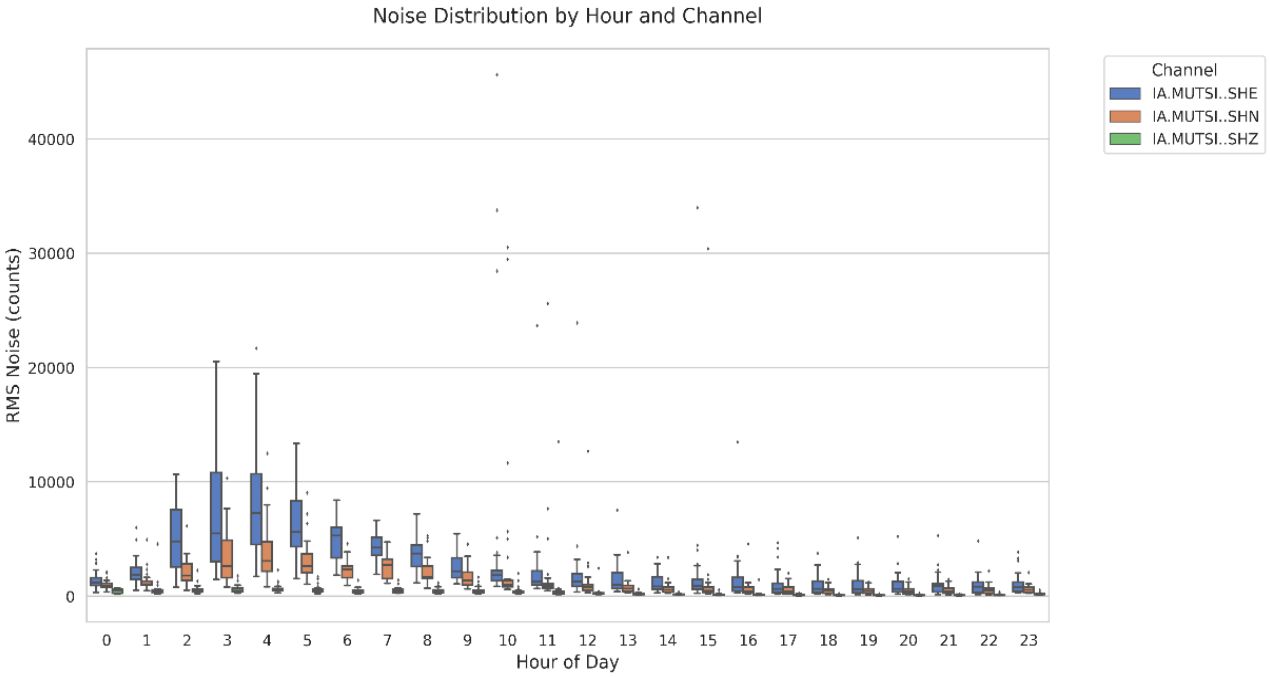

Figure 3. Boxplot of hourly RMS noise values for the MUTSI station across three components: SHE, SHN, and SHZ. Horizontal components exhibit higher variability and extreme outliers compared to the vertical channel, indicating stronger sensitivity to anthropogenic disturbances.

The median RMS noise value for this component reached around 5,000-7,000 m^2/s^4/Hz in the early morning hours, with a wide range and many extreme values, indicating the presence of non-Gaussian noise spikes most likely caused by anthropogenic disturbances such as heavy vehicle vibrations, industrial machinery, or local soil–structure interactions. The SHN component also showed a similar distribution pattern but with lower medians, around 2,000-3,000 m^2/s^4/Hz. Meanwhile, the SHZ (vertical) component showed a more concentrated and stable distribution, with a median of about 300-500 m^2/s^4/Hz, a narrow interquartile range (IQR), and relatively few outliers, indicating that the vertical component is more resistant to lateral disturbances. The summaries result for the RMS noise level in the three components is shown in the Table 1.

Table 1. Numerical comparison of RMS noise levels between daytime and nighttime at MUTSI station.

Component	Daytime RMS	Nighttime RMS	Reduction (%)	Horizontal-to-Vertical Ratio
SHE (E–W)	5,500-7,000	800-1,200	~80-85%	8-10x higher than SHZ
SHN (N–S)	3,000-4,000	700-1,000	~70-80%	5-7x higher than SHZ
SHZ (Vertical)	500-600	200-300	~45-60%	Reference

The table above shows that the horizontal components (SHE and SHN) experience substantially higher noise amplitudes compared to the vertical component, with daytime values reaching up to 7,000 counts, while nighttime noise is reduced by as much as 80–85 percent. In contrast, the SHZ component remains more stable, with only modest diurnal variations. These numerical differences reinforce the interpretation that horizontal channels are more sensitive to anthropogenic activity, local surface conditions, and site environment.

This pattern confirms that early morning hours (03:00-06:00 UTC) are the worst periods for noise, while 16:00-22:00 UTC are the best times for recording high-quality seismic signals with optimal SNR. Physically, noise in the frequency range 1-10 Hz is highly sensitive to exogenous sources such as traffic, human activity, structural resonance, and wind (Withers et al., 1996; Groos & Ritter, 2009). The statistical analysis shown by the boxplot and median patterns together demonstrate that noise characteristics have strong temporal dependence. Therefore, noise mitigation strategies such as placing stations far from settlements, using

DOI : http://dx.doi.org/10.12962/j25023659.v11i3.8972
Received : 30-09-2025 Reviewed : 10-10-2025 Accepted : 19-12-2025 Available Online : 31-12-2025

underground vaults, and applying bandpass filters of 1-10 Hz in pre-processing are highly recommended. Understanding these noise dynamics can also be used to schedule automatic earthquake detection and adjust trigger thresholds (STA/LTA) in real-time systems.

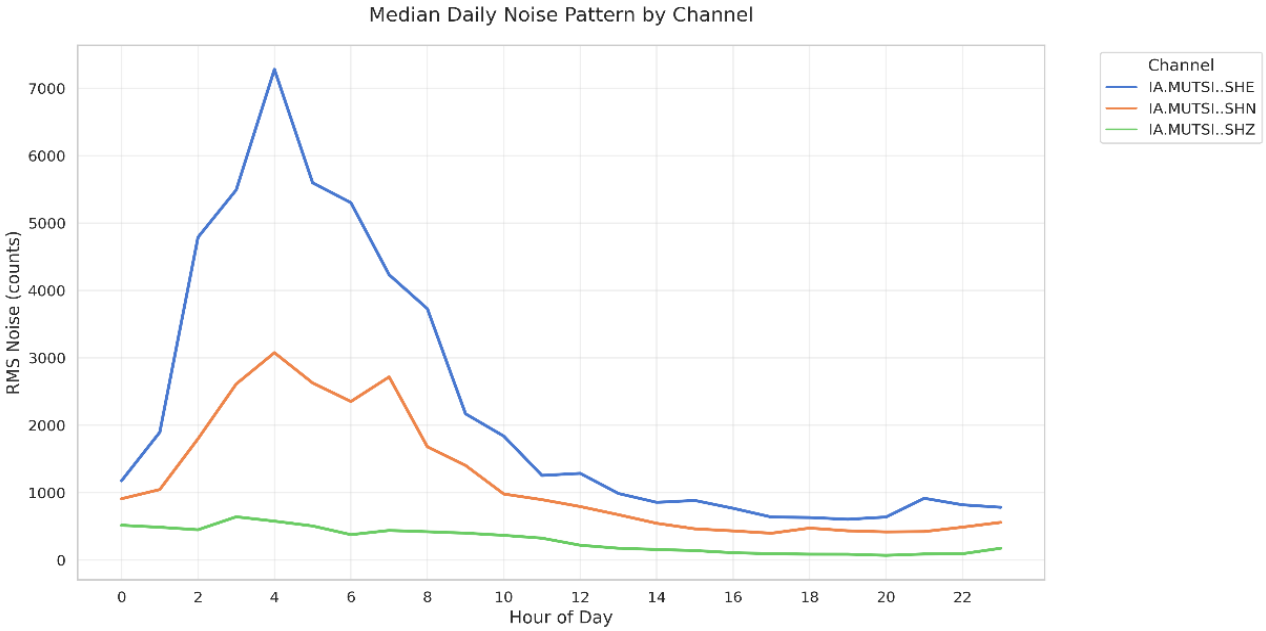

Figure 4. Diurnal variation of median RMS noise at the MUTSI station. Horizontal components (SHE, SHN) show strong fluctuations with a peak around 04:00 UTC, while the vertical component (SHZ) remains more stable throughout the day.

The second graph (line chart) presents daily median RMS noise values as a function of time. It can be seen that the median value for the SHE component peaked around 04:00 UTC, at about 7,300 $m^2/s^4/Hz$, before decreasing sharply after 06:00 UTC. This decline was exponential, approaching 800-1,000 $m^2/s^4/Hz$ in the afternoon to evening. The SHN component showed a similar pattern, although at lower amplitude levels. The SHZ component, as a reference for vertical signal stability, showed small fluctuations, with the highest median value only around 600 $m^2/s^4/Hz$, decreasing to 200-300 $m^2/s^4/Hz$ at night.

The boxplot above presents the weekly distribution pattern of seismic noise based on the three recording components at the MUTSI station, namely east-west horizontal (SHE), north-south horizontal (SHN), and vertical (SHZ). Noise levels were measured in RMS (Root Mean Square), directly correlated with ground vibration amplitudes in the dominant frequency range of 1-10 Hz, which is relevant for detecting local and regional waves (McNamara & Buland, 2004; Bensen et al., 2007). Statistically, the horizontal components showed the highest and most fluctuating RMS noise distributions throughout the week.

Median RMS noise on weekdays (Monday to Friday) ranged from 2,500 to 3,000 $m^2/s^4/Hz$, with upper quartiles reaching 5,000-6,000 $m^2/s^4/Hz$ and many outliers exceeding 20,000 $m^2/s^4/Hz$, even approaching 40,000 $m^2/s^4/Hz$ on some days. This indicates the dominance of anthropogenic noise, likely originating from human activities such as traffic, industrial operations, and other environmental disturbances (Groos & Ritter, 2009). The north-south horizontal component showed a similar trend but at lower median levels, generally below 2,000 $m^2/s^4/Hz$ and with more controlled distribution. Conversely, the vertical component showed the most stable and lowest noise distribution throughout the week. Median RMS noise generally remained below 600 $m^2/s^4/Hz$, with a narrow interquartile range and without extreme outliers.

DOI : http://dx.doi.org/10.12962/j25023659.v11i3.8972

Received : 30-09-2025 Reviewed : 10-10-2025 Accepted : 19-12-2025 Available Online : 31-12-2025

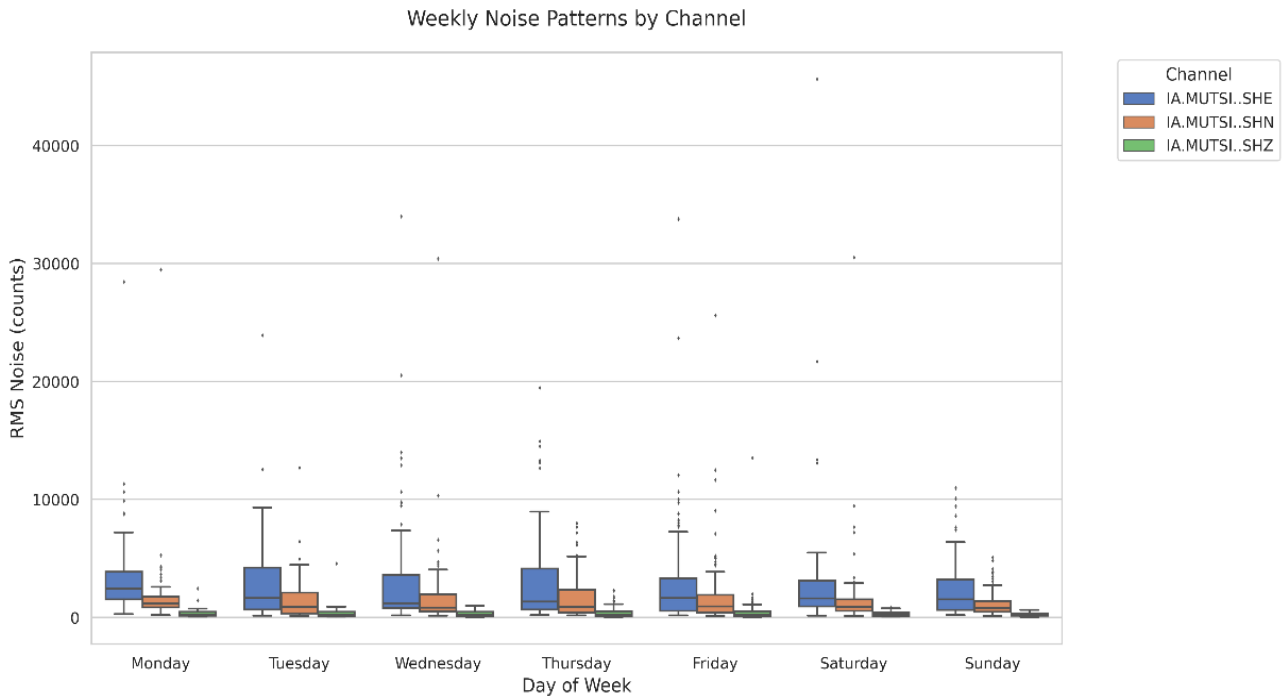

Figure 5. Weekly distribution of RMS noise at the MUTSI station across horizontal and vertical components. Horizontal channels display higher noise during weekdays, with reduced levels during weekends. The vertical channel shows consistently lower and more stable values.

This reflects the relative resistance of the vertical component to lateral disturbances, making it more suitable for capturing primary seismic waves (P-waves), which are crucial parameters in early detection systems and earthquake source analysis (Bormann, 2012). Weekly trends also showed that noise levels were relatively lower on weekends (Saturday and Sunday), reflecting reduced anthropogenic activity. This indicates that seismic signal quality can improve significantly when human activity decreases. The technical implications of these findings are the need to consider timing and specific components in optimizing automatic earthquake detection algorithms and determining ideal times for earthquake parameter extraction.

In addition, this information is very useful for selecting new seismic station locations to minimize the effect of surface disturbances on recordings. The hourly spectrogram pattern in Figure 5 shows the distribution of seismic noise levels in the east-west horizontal component (SHE) at the MUTSI station based on time on a daily and hourly scale for one month. The colours represent RMS noise values, reflecting the strength of seismic disturbances in the dominant frequency range of about 1-10 Hz.

This frequency range is important because it is within the sensitive zone for detecting surface and body waves from local and regional seismic events (McNamara & Buland, 2004; Bensen et al., 2007). In general, most RMS noise values were in the low-to-moderate range (dark purple to blue), but there were significant anomalies recorded as green, yellow, and bright orange. Extreme noise peaks were identified on several days and times, such as on the 18th day around 10:00-11:00, with RMS values exceeding 45,000 m^2/s^4/Hz. Other significant spikes occurred on the 5th, 8th, 14th, 23rd, and 31st days, usually around 10:00-12:00 noon, which are the busiest hours of human activity.

This is consistent with anthropogenic noise patterns that usually increase during the day, particularly due to vibrations from heavy vehicles, construction, or industrial activity (Groos & Ritter, 2009). Statistically, daily noise fluctuation patterns showed uneven distribution, with high variance at certain hours. This analysis reinforces the results of the previous boxplots and median graphs, showing that the SHE component is most vulnerable to surface disturbances. High and unpredictable noise can reduce the signal-to-noise ratio (SNR), thereby disrupting the sensitivity of automatic detection of weak seismic waves (Withers et al., 1996).

DOI : http://dx.doi.org/10.12962/j25023659.v11i3.8972

Received : 30-09-2025 Reviewed : 10-10-2025 Accepted : 19-12-2025 Available Online : 31-12-2025

Therefore, information from this heatmap is very useful for disturbance mitigation strategies, such as frequency filters and scheduling analysis based on optimal times.

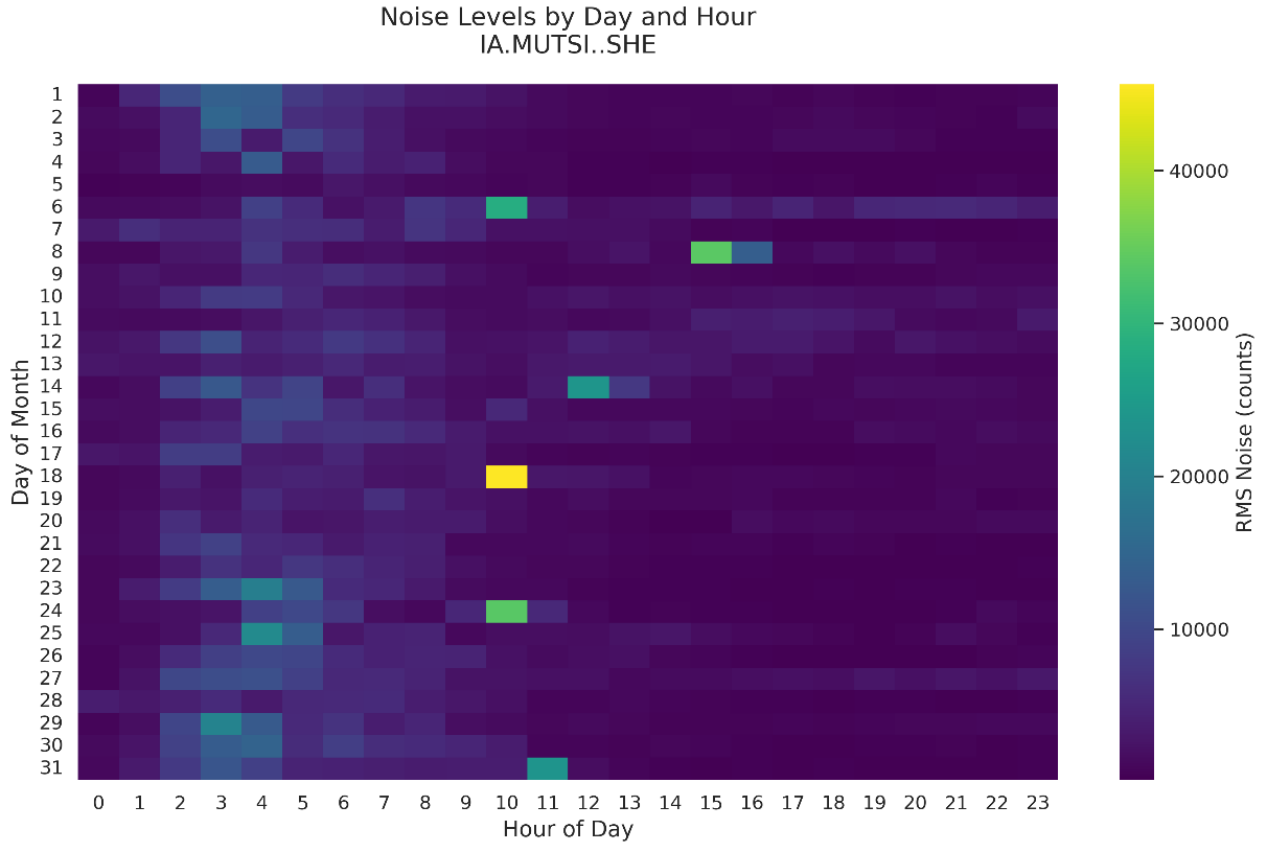

Figure 6. Hourly spectrogram of RMS noise in the east–west horizontal component (SHE) at the MUTSI station over a one-month period. Colors indicate noise intensity (counts in physical units $m^2/s^4/Hz$) within the dominant frequency range of 1-10 Hz, with extreme anomalies observed during daytime hours.

The first figure shows the noise distribution pattern based on clustering results using the fuzzy c-means (FCM) algorithm, while the second figure shows the distribution of each cluster by day of the week. From the FCM results, five main clusters (from cluster 0 to 4) successfully grouped noise characteristics based on their temporal patterns. Cluster 0 dominated with more than 380 observations, representing low-noise or stable daily baseline conditions. Cluster 1 was also significant with more than 200 observations, possibly indicating consistent moderate noise. Clusters 2, 3, and 4 had fewer observations and tended to represent anomalies or sporadic high-noise conditions.

In terms of temporal distribution, cluster distribution per day showed that cluster 0 dominated on weekdays (Tuesday to Friday), indicating relatively stable and low noise during those days. Meanwhile, cluster 1 tended to appear more on Mondays and Saturdays, possibly associated with anthropogenic fluctuations at the start and end of the week. Cluster 4, which visually increased on Wednesdays and Thursdays, may be related to specific local activity increases. Clusters 3 and 2, although rare, are important to note because they may represent significant disturbances in seismic signals, such as local large earthquakes, mining activity, or instrumental disturbances.

DOI : http://dx.doi.org/10.12962/j25023659.v11i3.8972

Received : 30-09-2025 Reviewed : 10-10-2025 Accepted : 19-12-2025 Available Online : 31-12-2025

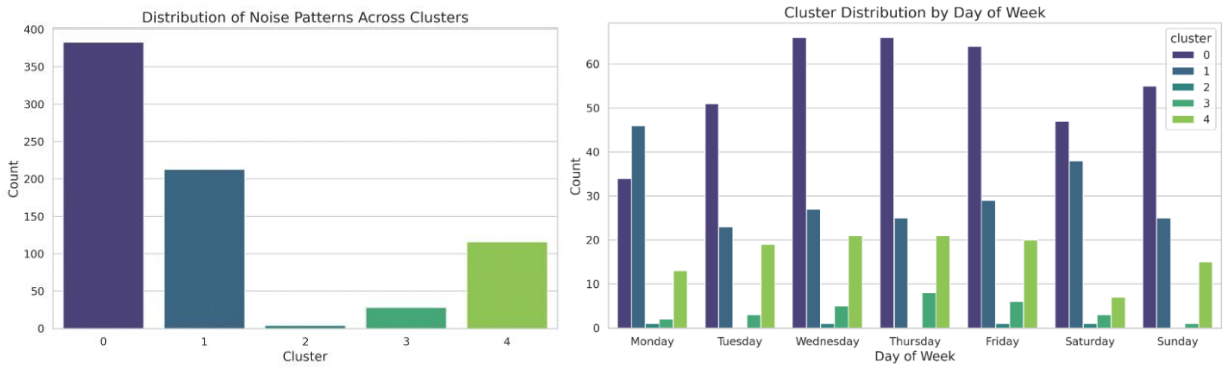

Figure 7. Fuzzy c-means (FCM) clustering of seismic noise at the MUTSI station. (top panel) Distribution of five noise clusters (0-4) based on temporal characteristics. (bottom panel) Cluster occurrence by day of the week, highlighting stable baseline noise during weekdays and increased variability at the beginning and end of the week.

Physically, the fuzzy clustering method provides flexibility in representing noise data that contain spatial–temporal ambiguities (Bezdek et al., 1984). Each observation has a degree of membership in each cluster, allowing for richer interpretation compared to hard clustering methods such as k-means. This is highly relevant in seismic noise analysis, where variability is strongly influenced by environmental, anthropogenic, as well as weather and soil conditions (McNamara & Buland, 2004). Therefore, these cluster distribution patterns can serve as a basis for developing mitigation strategies and seismic data acquisition planning, including scheduling optimal times for recording earthquake signals with the best signal-to-noise ratio (SNR).

CONSLUSION

This study provides a comprehensive evaluation of seismic site quality in North Sumatra by integrating spectral and machine learning approaches. Power Spectral Density (PSD) analysis successfully characterized noise levels across short-period, microseismic, and long-period frequency bands, revealing strong temporal variability driven by both environmental and anthropogenic factors. Probability Density Function (PDF) statistics further quantified the distribution of noise, highlighting the greater stability of vertical components (SHZ) compared to horizontal channels (SHE and SHN), which are more susceptible to lateral disturbances. In addition, fuzzy c-means clustering effectively classified noise characteristics into five temporal regimes, distinguishing stable baseline conditions from sporadic anomalies linked to local activity or instrumental effects. Collectively, these findings demonstrate that seismic site quality is highly time-dependent and emphasize the importance of combining spectral density analysis with machine learning to improve earthquake monitoring, optimize site selection, and enhance the accuracy of real-time seismic detection systems in North Sumatra and similar tectonically active regions.

ACKNOWLEDGEMENT

The authors would like to thank the editorial team and the anonymous reviewers for their constructive and insightful comments, which have significantly improved the quality of this manuscript. We also gratefully acknowledge the Indonesian Agency for Meteorology, Climatology, and Geophysics (BMKG) for providing the seismic data used in this study. This research was supported by funding from the Human Resources Development Center of BMKG (PPSDM BMKG).

DAFTAR PUSTAKA

Anthony, R.E., Ringler, A.T., Wilson, D.C. & Wolin, E. (2020), "How much do microbaroms matter for seismic noise?", *Seismological Research Letters*, Vol.91, No.1, hal. 611–624. https://doi.org/10.1785/0220190176.

Received : 30-09-2025 Reviewed : 10-10-2025 Accepted : 19-12-2025 Available Online : 31-12-2025

Bensen, G.D., Ritzwoller, M.H., Barmin, M.P., Levshin, A.L., Lin, F.C., Moschetti, M.P., Shapiro, N.M. & Yang, Y. (2007), "Processing seismic ambient noise data to obtain reliable broad-band surface wave dispersion measurements", *Geophysical Journal International*, Vol.169, No.3, hal. 1239–1260. https://doi.org/10.1111/j.1365-246X.2007.03374.x.

Bezdek, J.C., Ehrlich, R. & Full, W. (1984), "FCM: The fuzzy c-means clustering algorithm", *Computers & Geosciences*, Vol.10, No.2–3, hal. 191–203. https://doi.org/10.1016/0098-3004(84)90020-7.

Birnie, C., Chambers, K., Horleston, A., Kendall, J.M., Nowacki, A. & Wookey, J. (2021), "Quantifying the performance of seismic stations across the UK", *Geophysical Journal International*, Vol.226, No.3, hal. 1889–1909. https://doi.org/10.1093/gji/ggab159.

Bormann, P. (2012), *New manual of seismological observatory practice 2 (NMSOP-2)*, Deutsches GeoForschungsZentrum GFZ. https://doi.org/10.2312/GFZ.NMSOP-2.

Cooley, J.W. & Tukey, J.W. (1965), "An algorithm for the machine calculation of complex Fourier series", *Mathematics of Computation*, Vol.19, No.90, hal. 297–301. https://doi.org/10.1090/S0025-5718-1965-0178586-1.

Groos, J.C. & Ritter, J.R.R. (2009), "Time domain classification and quantification of seismic noise in an urban environment", *Geophysical Journal International*, Vol.179, No.2, hal. 1213–1231. https://doi.org/10.1111/j.1365-246X.2009.04343.x.

Janiszewski, H.A., Delph, J.R., Levander, A. & Niu, F. (2023), "Ambient seismic noise tomography for crustal structure in tectonically active regions", *Geophysical Research Letters*, Vol.50, No.2, e2022GL101234. https://doi.org/10.1029/2022GL101234.

Johnson, C.W., Meng, H., Vernon, F. & Ben-Zion, Y. (2020), "Characteristics of seismic noise recorded across the San Jacinto fault zone", *Seismological Research Letters*, Vol.91, No.1, hal. 595–610. https://doi.org/10.1785/0220190147.

Ke, H., Li, X., Wang, W. & Chen, Q. (2023), "High-frequency seismic noise attenuation with distance in different geological settings", *Journal of Seismology*, Vol.27, No.3, hal. 487–504. https://doi.org/10.1007/s10950-023-10125-6.

Maryana, M., Simanjuntak, A. V., Sinambela, M., Sihotang, N., Saragih, N. F., Aritonang, M., ... & Darnila, E. (2021). Waveform Evaluation of Seismic Network in the Aceh Province by using Power Spectral Density and Probability Density Function. *Internetworking Indonesia Journal*, 13(2), 49-52.

McNamara, D.E. & Buland, R.P. (2004), "Ambient noise levels in the continental United States", *Bulletin of the Seismological Society of America*, Vol.94, No.4, hal. 1517–1527. https://doi.org/10.1785/012003001.

McNamara, D.E., Buland, R.P., Ringler, A.T. & Hutt, C.R. (2009), "Seismic noise power spectral density estimates for global seismic stations", *Seismological Research Letters*, Vol.80, No.4, hal. 665–671. https://doi.org/10.1785/gssrl.80.4.665.

Nurana, I., Susanto, E., Simanjuntak, A. V. H., Rusdin, A. A., Syamsidik, & Umar, M. (2023, January). A possibility scenario of tsunami run-up model in Aceh from potential major earthquake in the Northern Part of Sumatra. In AIP Conference Proceedings (Vol. 2613, No. 1, p. 020007). *AIP Publishing LLC*. https://doi.org/10.1063/5.0119559

Parolai, S., Bormann, P. & Milkereit, C. (2006), "New relationships between Vs, thickness of sediments, and resonance frequency calculated by the H/V ratio of seismic noise", *Bulletin of the Seismological Society of America*, Vol.96, No.2, hal. 459–471. https://doi.org/10.1785/0120050102.

Peterson, J. (1983), *Observations and modeling of seismic background noise (Open-File Report 93-322)*, U.S. Geological Survey. https://doi.org/10.3133/ofr93322.

Simanjuntak, A. V. H., & Jihad, A. Perbandingan Setiap Kontur Transfomasi Pada Anomali Magnetik Untuk Identifikasi Sesar (Studi Kasus: Sesar Toru Pada Great Sumatera Fault). *Jurnal Fisika Flux: Jurnal Ilmiah Fisika FMIPA Universitas Lambung Mangkurat*, 14(2), 78-84. http://dx.doi.org/10.20527/flux.v14i2.3835

Simanjuntak, A. V., Sihotang, N., Simangunsong, A. V., Simamora, B. M., Kuncoro, D. C., Asnawi, Y., ... & Irwandi, I. (2021, October). Application of P-Wave Moment Magnitude (Mwp) and Rupture Time Duration (Tdur) to Analyze A Potential Tsunami Earthquake in Sumatra. In IOP Conference Series: Earth and Environmental Science (Vol. 873, No. 1, p. 012093). *IOP Publishing*. 10.1088/1755-1315/873/1/012093

Sirait, A. M., Suryanto, W., Simanjuntak, A. V., Magdalena, C., Maulidira, N., Rizqi, M. N., & Fitria, I. C. (2025, October). Preliminary analysis of identified earthquake swarms in eastern Indonesia. In IOP Conference Series: Earth and Environmental Science (Vol. 1547, No. 1, p. 012008). *IOP Publishing*. 10.1088/1755-1315/1547/1/012008

DOI : http://dx.doi.org/10.12962/j25023659.v11i3.8972

Received : 30-09-2025 Reviewed : 10-10-2025 Accepted : 19-12-2025 Available Online : 31-12-2025

Stutzmann, E., Schimmel, M., Patau, G. & Maggi, A. (2012), "Global climate imprint on seismic noise", *Geochemistry, Geophysics, Geosystems*, Vol.13, No.11. https://doi.org/10.1029/2012GC004333.

Tchawe, B.N., Kervyn, F. & Delvaux, D. (2020), "Microzonation of seismic hazards using ambient noise recordings: A case study from East Africa", *Journal of African Earth Sciences*, Vol.164, 103789. https://doi.org/10.1016/j.jafrearsci.2020.103789.

Withers, M.M., Aster, R.C., Young, C.J., Chael, E.P. & Randall, G.E. (1996), "A comparison of select trigger algorithms for automated global seismic phase and event detection", *Bulletin of the Seismological Society of America*, Vol.86, No.3, hal. 953–961. https://doi.org/10.1785/BSSA0860030953.

Zhong, X., Li, Y. & Li, C. (2021), "Propagation characteristics of high-frequency seismic noise in surface environments", *Earthquake Science*, Vol.34, No.3, hal. 265–278. https://doi.org/10.1016/j.eqsce.2021.06.004.

DOI : http://dx.doi.org/10.12962/j25023659.v11i3.8972

Received : 30-09-2025 Reviewed : 10-10-2025 Accepted : 19-12-2025 Available Online : 31-12-2025