Hindawi
Shock and Vibration
Volume 2022, Article ID 3109609, 13 pages
https://doi.org/10.1155/2022/3109609

Research Article

Fuzzy Clustering Analysis of HVSR Data for Seismic Microzonation at Lahore City

Sarfraz Khan,[1] Muhammad Waseem,[2] Shahzad Khalid,[1] Denise-Penelope N. Kontoni,[3,4] Mahmood Ahmad,[5] and Suraparb Keawsawasvong[6]

[1]*National Centre of Excellence in Geology, University of Peshawar, Peshawar 25130, Pakistan*
[2]*Department of Civil Engineering, University of Engineering and Technology, Peshawar 25000, Pakistan*
[3]*Department of Civil Engineering, School of Engineering, University of the Peloponnese, GR-26334, Patras, Greece*
[4]*School of Science and Technology, Hellenic Open University, GR-26335, Patras, Greece*
[5]*Department of Civil Engineering, University of Engineering and Technology Peshawar (Bannu Campus), Bannu 28100, Pakistan*
[6]*Department of Civil Engineering, Thammasat School of Engineering, Thammasat University, Pathumthani 12120, Thailand*

Correspondence should be addressed to Denise-Penelope N. Kontoni; kontoni@uop.gr

Received 6 August 2022; Revised 6 October 2022; Accepted 13 October 2022; Published 28 October 2022

Academic Editor: Roberto Nascimbene

A microzonation study deals with the classification of hazards in a town or city in terms of surface ground motions that result from amplification and resonance frequency in soils against seismic tremors. This paper presents the result of a microzonation study in terms of resonance frequency and peak amplitude for Lahore city, Pakistan. In order to recognize the local soil effects of the covered geology at 159 sites in Lahore city, the horizontal-to-vertical spectral ratio (HVSR) Nakamura method was implemented. A fuzzy C-mean (FCM) clustering algorithm was adopted to obtain the best cluster solution of the analyzed HVSR parameters. The results of the Silhouette Index suggest that the FCM clustering solution of observation data points is more consistent. The results of clustering reveal three solutions. Clusters 1 and 2 reveal that a major part of the research area possesses low to moderate frequencies (0.66–1.03 Hz) with a peak amplitude of 2.25–4.38 mm, indicating the presence of soft to hard rock and thick alluvial sedimentary cover. Cluster 3 reveals the presence of soft to compact rocks (with frequencies and amplitudes of 0.73–1.03 Hz and 3.02–4.11 mm, respectively) overlaying the bedrock. Lahore city has 60% of soil cover with an amplitude of 2–3 mm (for the central part) and about 40% of 3–4 mm in the northern, southern, and southwest portions. According to the NEHRP soil classification code of 1997, a major part of the city has stiff nature of the soil, while a few places reveal the presence of very dense soil. The maps produced in this study will provide expected ground motion-related useful information to reduce the seismic risk for infrastructure in Lahore city.

1. Introduction

An earthquake is a natural phenomenon that has caused huge loss of physical damage in different countries, e.g., Armenia, Iran, California, Japan, Greece, Turkey, India, Sumatra, Pakistan, China, and Haiti. Therefore, globally multi-dimensional earthquake hazard assessment studies by [1–4] have been performed. Several studies have shown that most Asian regions, including Pakistan, are vulnerable to seismic events [5] because they are located near active faults [6]. Historically, Pakistan has gone through many damaging earthquakes that resulted in the loss of lives and important infrastructures, e.g., the 1935 Quetta and 2005 Kashmir earthquakes [7]. Northern and northwestern parts of Pakistan are also considered as vulnerable due to moderate to high seismicity [8].

Seismic hazard analysis for a region can result in two forms, i.e., seismic macrozonation and microzonation. In seismic macrozonation, the probability of earthquake ground shaking corresponding to a return period is calculated by considering the shear wave velocity of bedrock. Structures such as roads, buildings, dams, are mostly built on soils

having different characteristics (shear wave velocity, frequency, amplification, etc.). These parameters cannot be estimated at a regional level, and hence a process called seismic microzonation is introduced to map and identify the response of an earthquake on local soil by estimating these parameters.

Various seismic microzonation techniques have been proposed, among which the site effect technique is generally followed to estimate the peak amplitude, resonance frequency, vulnerability to ground instabilities [9], top thirty-meter shear wave velocity (Vs30), etc. [10]. The recent experimental, numerical, and theoretical studies [11] of local soil in terms of damage distribution against earthquakes proved to be very effective. It is now accepted that local soil conditions play a significant role in earthquakes. Calculation of sediment thickness from resonance frequency is an appropriate way to characterize site conditions [12]. Mapping the average top 30-meter shear wave velocity (Vs30) of the sedimentary layers is a globally accepted parameter used in the recent development of buildings for forecasting the potential amplification against seismic shaking [13, 14].

The existing seismic hazard assessment studies of [15–18] have classified Lahore and its surroundings into microzones without incorporating the soil/subsoil condition. Limited research has been carried out to highlight the site response parameter (amplification, frequency, Vs30, etc.) in Pakistan. [19] estimated site effects of local soils in Fateh Jang region; [20] performed site response analysis to mitigate the impact on natural hazards; [10] performed seismic microzonation of Islamabad-Rawalpindi metropolitan area; [21] produced microzonation map for Abbottabad basin and its immediate surrounding; [22] highlighted the application of machine learning techniques on ambient noise data for Potwar region; and [23] performed site response studies in Peshawar using HVSR Nakamura technique.

Since Lahore city majorly lies over alluvial cover (Figure 1), no such studies have been opted for Lahore city, Pakistan. Therefore, this study is focused on the microzonation of Lahore city that takes into account the soil/subsoil conditions while designing and constructing multistory buildings and other engineering structures.

Clustering techniques have been found very promising in developing solutions for several geoscience problems [24–28]. In this study, a fuzzy C-mean clustering algorithm is used to find the cluster solution for HVSR data in order to investigate the site's effect parameters [29–31]. In Figure 2, the flowchart represents the important steps that are performed in the clustering analysis of observation data. Based on clustering analysis, these site response parameters are further classified in Lahore city into microzones that incorporate local soil/subsoil conditions and characterize Lahore city into different risk classifications against future earthquakes in the design of engineering structures.

2. Materials and Methods

In this study, the Nakamura method [32] of HVSR (Horizontal-to-Vertical Spectral Ratio) for the calculation of site effects parameters was adopted.

In this method, a single station measurement can easily be done to obtain subsurface information. This method was initially used to investigate the risk of seismicity in Japan [32, 33]. The HVSR method of Nakamura has been adopted for the seismic site characterization studies of large cities all over the world and obtained fruitful outputs from HVSR surveys by different researchers, i.e., [34] characterized the site of Iran based on HVSR; [35] adopted this method for seismic microzonation study of Catania Italy; [11] in Lubaljana area Slovenia; in Almerìa City [36, 37] in Beijing; [38] in Urban Las Vegas Nevada basin and [10] in Islamabad and Rawalpindi, Pakistan, etc.

Data acquisition is the initial stage in performing research work in any area or field. Horizontal-to-vertical spectral ratio data were acquired from 159 test locations using a Tromino Engy system for high-resolution wireless seismic and geophysical surveys (housed at NCE in Geology, University of Peshawar). For each data collecting point, the sensor was set to record data for 10 minutes duration at a sampling frequency of 128 Hz. This instrument sends electromagnetic pulse in the form of Rayleigh waves to the ground and detects soft/compact sedimentary layers.

The acquired HVSR data set was processed in the Grilla software by setting the engineering interest frequency range of 0.5–20 Hz. The H/V curves of HVSR data were generated using 10% smooth triangular windows (Figure 3). The features of H/V curves were analyzed in accordance with the worldwide recommendation introduced by [39]. First, the reliability of the curves was checked (i.e., is significant cycles for a definite *f*o, low scattering, and enough windows number acceptable for a set frequency range). Secondly, HVSR curves were checked to determine whether the criteria for a clear and reliable peak are fulfilled or not [39], as shown in Table 1. Based on these two criteria, 8 points were skipped due to so much noise and poor data.

2.1. Clustering Analysis of HVSR Data

2.1.1. Data Preprocessing. After analyzing HVSR data in the Grilla software, site response parameters, i.e., fundamental frequency (*f*o) of soil and H/V spectral amplitude (Ao) of corresponding H/V spectral ratios, were selected for clustering analysis.

A preprocessing tool was applied prior to clustering to remove any inconsistent (missing, unknown, or redundant) values. The data set was found to be consistent, and no missing or unknown values were found. However, different ranges are found in the data which need to be scaled. This task was done using a z-standardization test on two variables of amplitude and frequency. For elements X*i* = A*oi* and *foi*, the z-standardization can be calculated by

$$X_{iz} = \left(\frac{X_i - \mu}{\sigma}\right), \tag{1}$$

where σ = standard deviation and μ = mean value of variable Ao and *f*o. Some upscale points were found (Figure 4) using the z-standardization of the data. These points were scaled

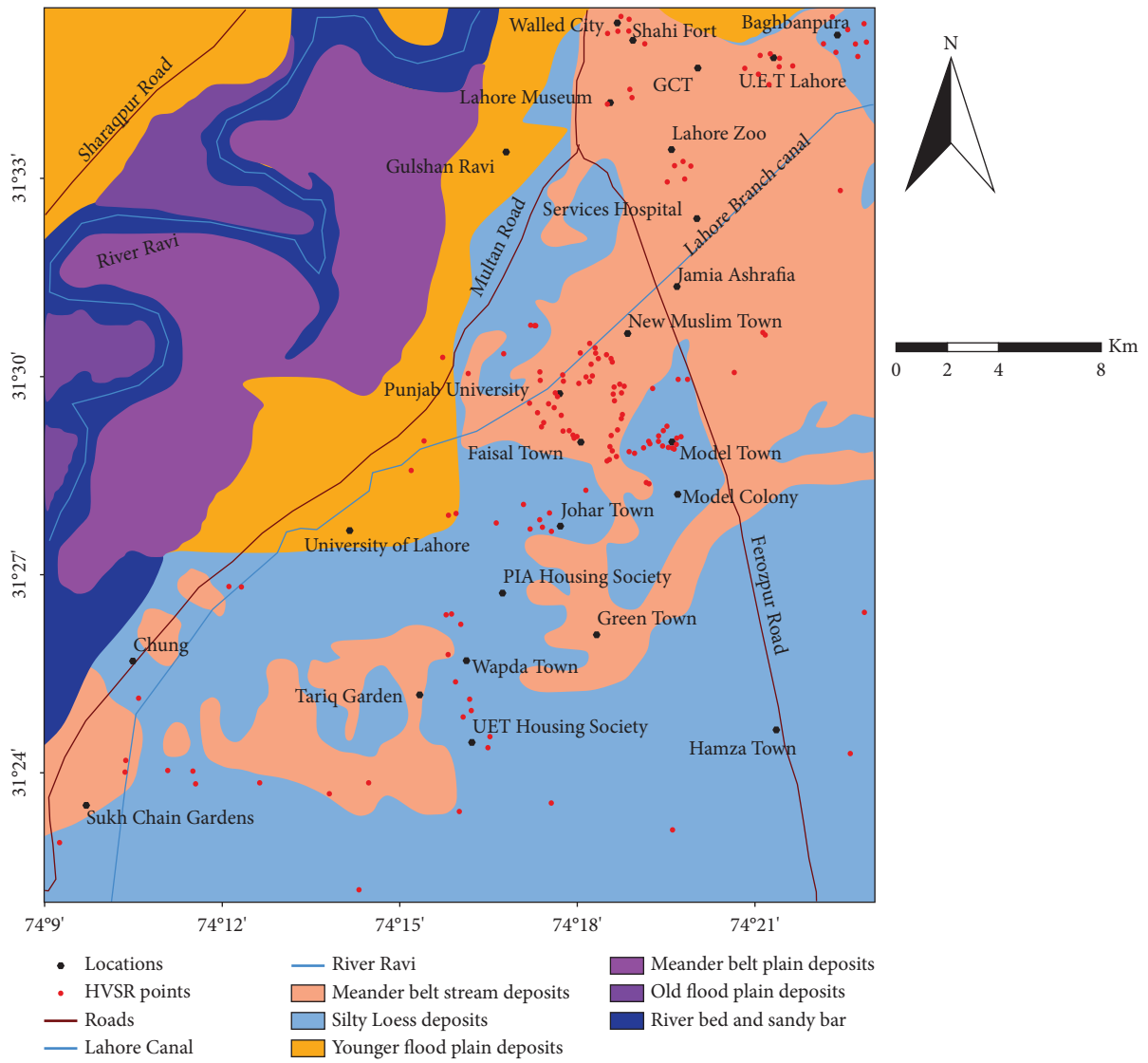

Figure 1: Geological map of the study area and acquired HVSR data points in the field.

using a standard deviation error value. Thus, a dataset of 151 records containing two variables was compiled for input in the clustering analysis stage.

2.1.2. Clustering Analysis. In this study, the processed dataset was assigned for clustering in XLSTAT statistical analysis software. XLSTAT is a flexible add-on for excel data analysis that allows operators to customize, analyze and share results with powerful features of machine learning, test validation, parametric/no parametric test, data visualization, clustering, etc. [40]. The Fuzzy C-mean (FCM) clustering algorithm developed by [41] and improved by [42] was used to find a number of cluster solutions in the HVSR data.

In the first step, the FCM algorithm attempts to divide a finite collection of n elements into a group of c fuzzy clusters according to some given principle. Given a finite set of data, this algorithm returns a list of cluster centers C and a partition matrix:

$$\text{Partition Matrix} \begin{bmatrix} 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 \end{bmatrix}. \tag{2}$$

In the second step, coefficients are assigned randomly to each data point for being in the clusters. Any point x_k has a set of coefficients giving the degree of being in the k^{th} cluster γ_{ik}. In the third step, the mean of all points, weighted by their degree of fitting to the cluster, is estimated by relation:

$$V_{ij} = \frac{\sum_1^n (\gamma_{ik}^m * x_k)}{\sum_1^n \gamma_{ik}^m}, \tag{3}$$

where m is the hyper-parameter, a fuzzier cluster will result at the end, if the value of m is higher.

For each observation, the clusters' coefficients are computed by updating the partition matrix using relation:

$$\gamma = \frac{1}{\sum_1^n \left(d_{ki}^2 / d_{kj}^2\right)^{1/(m-1).}} \tag{4}$$

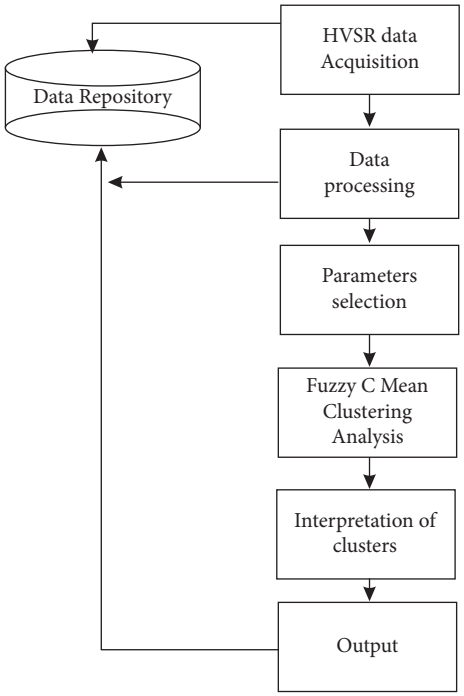

Figure 2: Flowchart indicating the stages performed for fuzzy C-mean clustering analysis in this research work.

This process is repeated until the algorithm has converged (that is, the coefficients' change between two iterations is no more than the given sensitivity threshold). Where for each element, γ tells the degree to which element belongs to a cluster.

For an overlapped data set, FCM results are quite better than the K-mean. However, the FCM algorithm responds poorly to data sets that contain unequal size clusters. It is also sensitive to noise and outliers.

2.2. Tukey-Kramer Post Hoc Test. To find which cluster means are exactly different, a most frequently used post hoc test of Tukey-Kramer was performed. In this test mean between pairs of samples is compared by calculating the absolute mean difference in each cluster. Then the critical Q value was determined using relation:

$$Q_{\text{critical value}} = Q^* \sqrt{\frac{\left(S^2\text{Pooled}\right)}{n}}, \quad (5)$$

where $Q^* = 5.99$ from the K-Wallis test; S^2pooled = Pooled variance across all groups; and n = Sample size for a given group = 10.

The pooled variance is calculated as the average of the variances for the clusters, which turns out to be 0.073. Thus, $Q_{\text{criticalvalue}}$ is calculated as 0.511.

In the last step, the comparison was made between the absolute mean-variance and each cluster to the Q critical value. The difference is significant if the absolute mean variance is larger than the Q critical value.

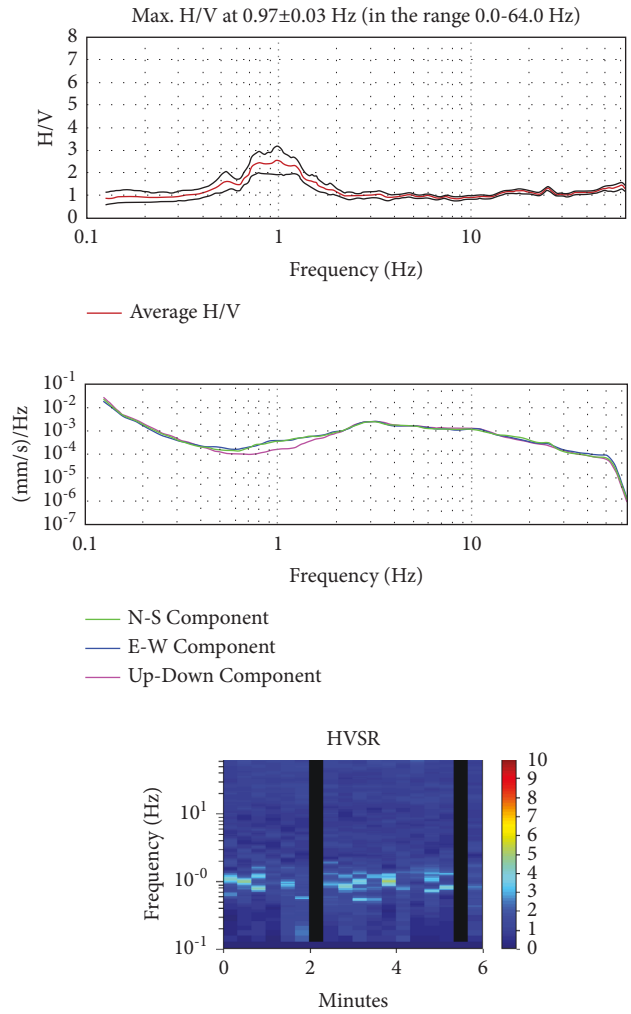

Figure 3: The amplitude of H/V spectral ratios as a function of fundamental frequencies representing the site response in the study area.

3. Comparison with Different Approaches

In this study, the cluster solution of the observed data was compared with four clustering techniques of K-mediod, Density-based spatial clustering of applications with noise (DBSCAN), and Hierarchical and K-mean algorithms [43–46]. Table 2 shows the comparison of these algorithms in terms of different criteria (distance measurement, granularity, initial and termination condition, etc.).

A comparison of p-value obtained using K-mean, Hierarchical, and FCM is presented in Table 3 in terms of cluster profiling of amplitude and frequency using a non-parametric classical one-way ANOVA Kruskal-Wallis test. The results are shown in Figures 5 and 6 as box plots. The Kruskal-Wallis test gives the interpretation of consistency in observed data samples. In all three approaches, the computed p-value is lower than the significance level of alpha = 0.05 for H/V amplitude, rejects the null hypothesis (Ho) for variable amplitude (Ao), and accepts the alternative

Table 1: SESAME criteria for HVSR data.

Criteria for a reliable HVSR curve
$fo > 10/lw$
nc $(fo) > 200$
$\sigma A(f) < 2$ for $0.5fo < f < 2fo$ if $fo > 0.5$ Hz or $\sigma A(f) < 3$ for $0.5fo < f < 2fo$ if $fo < 0.5$ Hz
Criteria for a clear HVSR curve peak (at least 5 out of 6 criteria fulfilled)
Exists $f-$ in $[fo/4, fo] \mid AH/V(f-) < Ao/2$
Exists $f+$ in $[fo, 4fo] \mid AH/V(f+) < Ao/2$
$Ao > 2$ $fpeak[AH/V(f) \pm \sigma A(f)] = fo \pm 5\%$
$\sigma f < \sigma\ (fo)$
$\sigma A\ (fo) < \sigma\ (fo)$

*"nc = lw * nw"; nw = number of windows selected for the average H/V curve; lw = window length; σf = standard deviation of H/V peak frequency; fo = number of significant cycles; f = current frequency; $\varepsilon\ (fo)$ = threshold value for the stability condition $\sigma f < \varepsilon\ (fo)$; Ao = H/V peak amplitude at frequency; AH/V (f) = H/V curve amplitude at frequency; $f+$ = frequency between fo and $4fo$ for which $AH/V(f+) < Ao/2$; $\sigma A\ (f)$ = "standard deviation" of AH/V (f), $\sigma A\ (fo)$ is the factor by which the mean AH/V(f) curve should be multiplied or divided; $f-$ = frequency between $fo/4$ and fo for which $AH/V(f-) < Ao/2$.

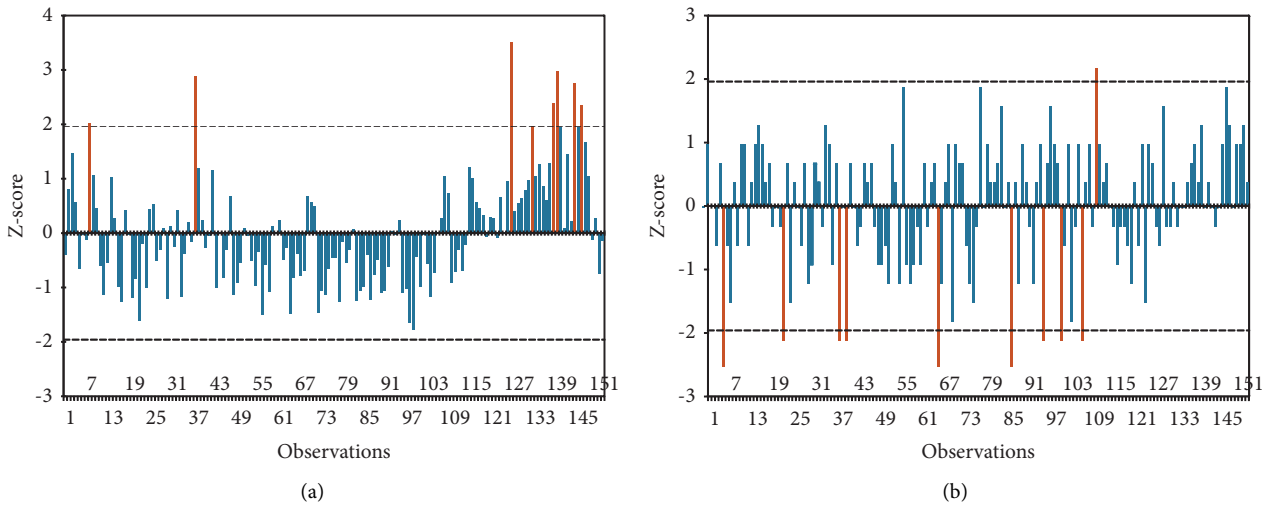

Figure 4: Z-standardization test performed on observed data for parameters: (a) amplitude and (b) frequency.

hypothesis (Ha). For variable fo the computed p-value is greater than the significance level alpha = 0.05, indicating that Kruskal-Wallis cannot reject the null hypothesis Ho.

The results of the Silhouette Index suggest that the FCM clustering solution is closer to +1, indicating the cluster is more consistent in differentiating the characteristics of observation data points. In contrast, Hierarchical and K-mean clustering solutions have negative Silhouette values, which tell about some misclassified locations of observed data.

4. Results and Discussion

This study is based on 159 sites. Some locations were very restricted and not allowed to carry out a survey. The official geological map on a local scale is also not available; the geological maps are only available from different research works in parts. The reason is the lack of studies that incorporate local soil conditions. But in this study, we try to map local soil and categorize it into different soil classes.

In this study, a meaningful cluster solution of the observed data was evaluated using the clustering algorithm of fuzzy C-mean. To see the effect of various centroids, the observed data set was iterated ten times for the optimal number of clusters by applying the fuzzy C-mean technique.

The profiling of a cluster for a variable plays a significant role in defining the differences between the clustering solutions while finding its characteristics. The clustering solutions generated by FCM were further analyzed in terms of cluster profiling of amplitude and frequency based on the median, mean, standard error, and standard deviation values using a non-parametric classical one-way ANOVA Kruskal-Wallis test. The profiling results for these two variables are listed in Table 4. In this table, the four rows for each variable fo and Ao contain information about the values of a number of observations without missing data, minimum, maximum, mean, and standard deviation for each cluster. The Kruskal-Wallis test gives the interpretation that the observed data samples either come from the same location (Ho) or do not come from the same location (Ha). In each table, the computed p-value is lower than the significance level of alpha = 0.05 for H/V amplitude rejecting the null hypothesis (Ho) for variables amplitude (Ao) and accepting the alternative hypothesis (Ha). For variables fo, the computed p-value is greater than the significance level alpha = 0.05, indicating that Kruskal-Wallis cannot reject the null hypothesis Ho.

Table 2: Comparison of different clustering algorithms.

Criteria	FCM	Hierarchical	*K*-mean	*K*-mediod	DBSCAN
Initial condition	YES	YES	YES	YES	YES
Termination condition	Precise	Not precise	Precise	Precise	Precise
Effect on size of data set	Good	Not good	Good	Not good	Not good
Shape of data set	Convex	Arbitrary	Convex	Convex	Arbitrary
Granularity	Flexible	Flexible	K and initial point	K and initial point	Threshold
Result optimization	Optimization	Optimization	Rebuild optimization	Rebuild optimization	Rebuild optimization
Distance measurement	Centroid	Any	Distance at normal space	Distance at normal space	Density
Accuracy without noise	90%	88%	76%	80%	71%

Table 3: A comparative Kruskal-Wallis test for fuzzy C-mean, Hierarchical, and *K*-mean clustering of two parameters.

Parameters	FCM *p*-value	Hierarchical *p*-value	*K*-mean *p*-value	Significance
Amplitude (Ao)	<0.0001	<0.0001	<0.0001	0.05
Fundamental frequency (*fo*)	0.157	0.447	0.286	0.05

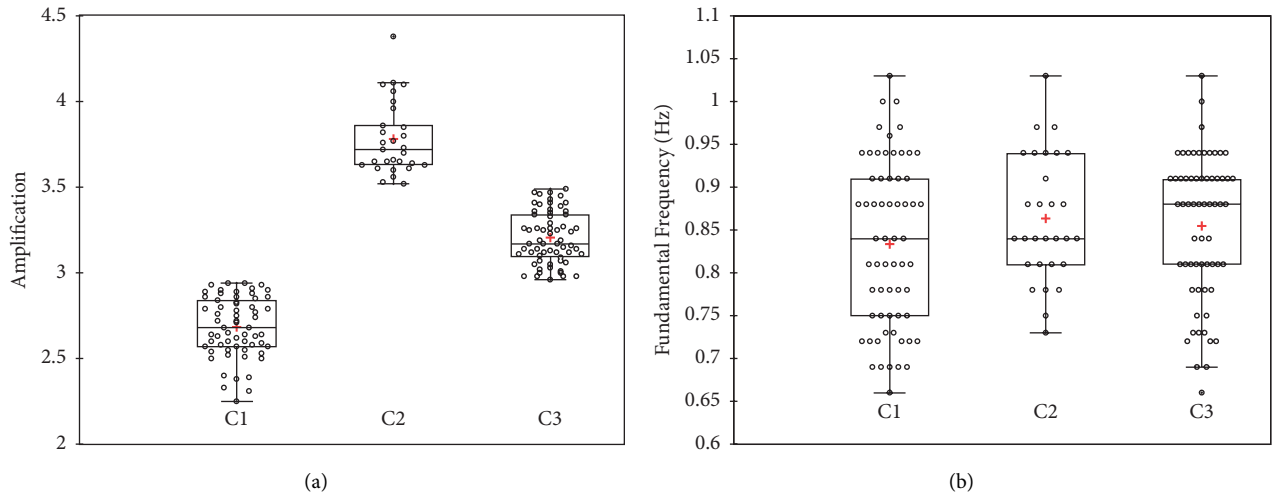

Figure 5: Box-plots showing the distribution of variables (a) Amplitude (Ao) and (b) Frequency (fo) by K-mean clustering technique.

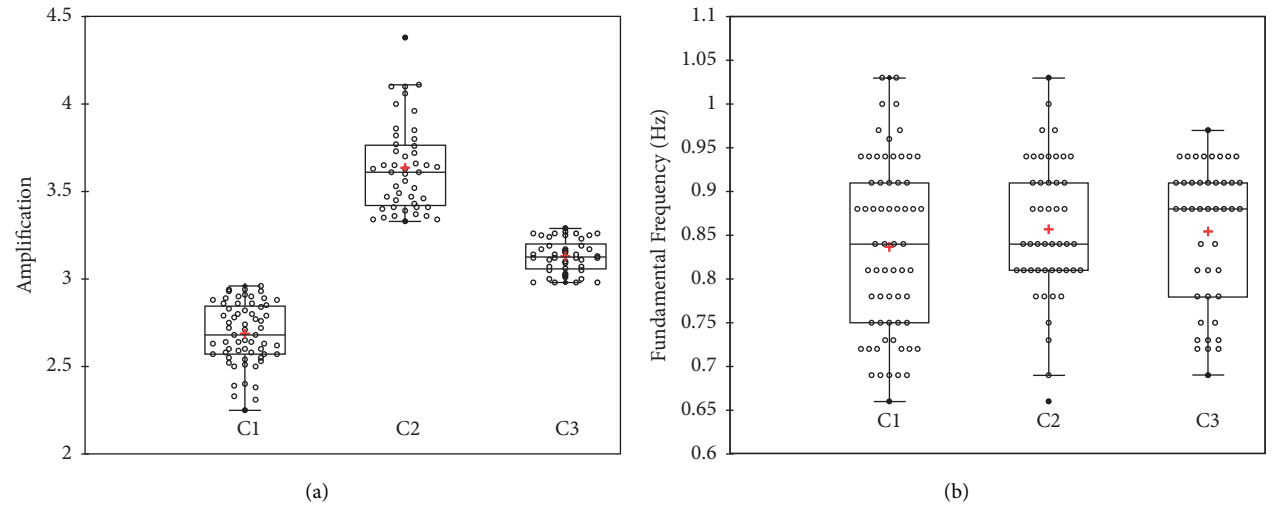

Figure 6: Box-plots showing the distribution of variables (a) Amplitude (Ao) and (b) Frequency (fo) by Hierarchical clustering technique.

Table 4: Kruskal-Wallis test for fuzzy C-mean clustering.

Parameters	Cluster 1	Cluster 2	Cluster 3	Degree of freedom	p-value	Significance
Amplitude (Ao)	63*	61	29	2	<0.0001	0.05
	2.64†	2.25	3.02			
	4.38‡	3.49	4.11			
	3.23**	2.71	3.64			
	0.29††	0.22	0.28			
Fundamental frequency (fo)	63	61	29	2	0.157	0.05
	0.66	0.66	0.73			
	1.03	1.03	1.03			
	0.85	0.831	0.87			
	0.09	0.095	0.07			

*: number of observations without missing data; †: minimum; ‡: maximum; **: mean; ††: standard deviation.

This indicates the cluster solutions of variables are from the same distribution (for Ao) and from different distributions (for fo). In the statistical analysis, the Silhouette Index of clusters has been found to be very effective in the interpretation of the cluster solutions [47]. The results of the Silhouette Index suggest that the FCM clustering solution is closer to +1 (average Silhouette width of cluster 1 = 0.480; cluster 2 = 0.498 and cluster 3 = 0.533), indicating the cluster is more consistent in differentiating the characteristics of observation data points.

The distribution of two variables (frequency and amplitude) in clusters 1, 2, and 3 is also shown as box-plots (in Figure 7) and scatter plot (in Figure 8) for the FCM clustering algorithm. The solution of cluster results shows that two variables H/V amplitude and frequency are very significant for discriminating among clusters. Cluster 1 is the largest cluster (63 observations), having a high H/V amplitude of 2.64–4.38 mm and low to medium range frequencies (0.66–1.03 Hz).

Cluster 2 has 61 observations with low to medium range frequencies of 0.66–1.03 Hz and moderate amplitude values (2.25–3.49 mm). The frequencies of HVSR peaks in Cluster 3 range from 0.73–1.03 Hz with an amplitude of 3.02–4.11 mm, resembles of soft to compact rocks overlaying the bedrock (Table 5). The fundamental frequency of cluster 1 ranges from 0.73 to 1.03 Hz with an amplification of 3.02–4.11 mm also indicates the presence of soft to hard rock and more sediment. The entire three clusters indicate the sites are moderately vulnerable in case of any seismic event. The observations in all clusters represent more sediment thickness, which indicates the bedrock is not shallow in the area.

Based on the one-way ANOVA Kruskal-Wallis test, the null hypothesis is not rejected for frequency data, while it is rejected for amplitude (Ao) data since the p-value is lower than the significance level of 5%. The rejection tells at least one sample is from a different location, and the means between the three clusters are not equal. Based on the Tukey-Kramer post hoc test, the mean difference between cluster 1-cluster 2, and cluster 2-cluster 3 is not statistically significant, while it is noted only between cluster 3 and cluster 1.

The contour maps of these variables, i.e., resonance frequency and amplification data, were prepared using the Kriging interpolation tool in ArcGIS 10.4 software developed by ESRI. In the contour map of frequency, the lowest dominant frequency ranging from 0.1–1.0 Hz and the highest ranging from 1.01–2.0 Hz is represented by green and red colors, respectively. Figure 9 also represents the correlation between topography and distribution of predominant frequency in the area. High topography has high frequencies, and low topography possesses low predominant frequencies. This reveals a consistent deposition of sediment layers. Low topography resembles the thickness of soft sediment layers corresponding to very low predominant frequencies, while high topography is related to weak layer deposition corresponding to high frequencies represented by red contours.

The study area is comprised of Quaternary sediments [48] (Figure 1). Almost 60% of the study area (for the central part, green color) has soil with amplification ranging from 2 to 3 mm. While 40 percent area, i.e., the northern, southern and southwest portions of the city, is covered by soil with amplification ranging from 3 to 4 mm (Figure 10 red contours). The peak amplitude is associated with the impedance contrast of bedrock and soil cover. Some locations with high soil amplification (i.e., Lahore Zoo, Lahore Museum, UET housing society, Hamza Town, Sukh Chain gardens, and Tariq gardens) have the tendency of high risk to earthquakes, while soil with a low amplification factor bears a low risk of damage in some locations including Model Town, Johar Town, PIA colony, Faisal Town, University of Punjab, Baghbanpura, Green town. Stiff soil is characterized by low soil amplification, while soft soils are characterized by high soil amplification values. The townships of Baghbanpura and model town are built over sandy gravel, sandy hard clay, and loam type soil that can weaken the soil strength against earthquakes greater than 5.0 M. About 90% of the city is built on deltaic-mud soil up to 30 meters in depth. The locations of the University of Punjab, Lahore Zoo, Lahore Museum, and many towns (Model town, Johar town, Garden town, Faisal town, etc.) are very sensitive to larger earthquakes if proper geotechnical and geophysical testing are not followed prior to construction. The construction along Lahore Branch Canal needs special study in terms of soil liquefaction and groundwater seepage. This canal needs proper maintenance on a yearly basis, and cleaning of sediments is required.

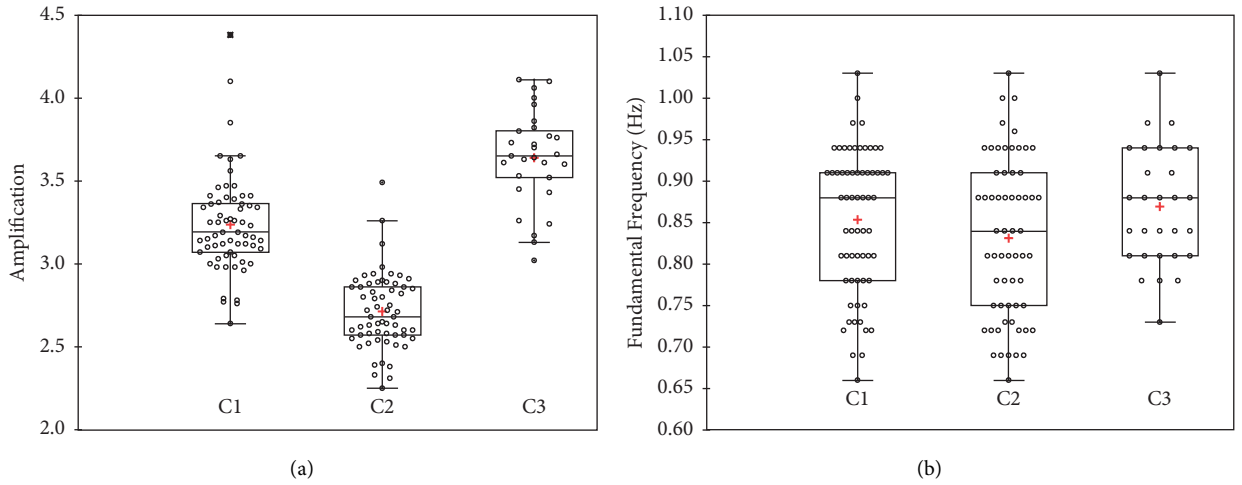

FIGURE 7: Box-plots showing the distribution of variables: (a) Amplitude (Ao) and (b) Frequency (fo), by FCM clustering technique.

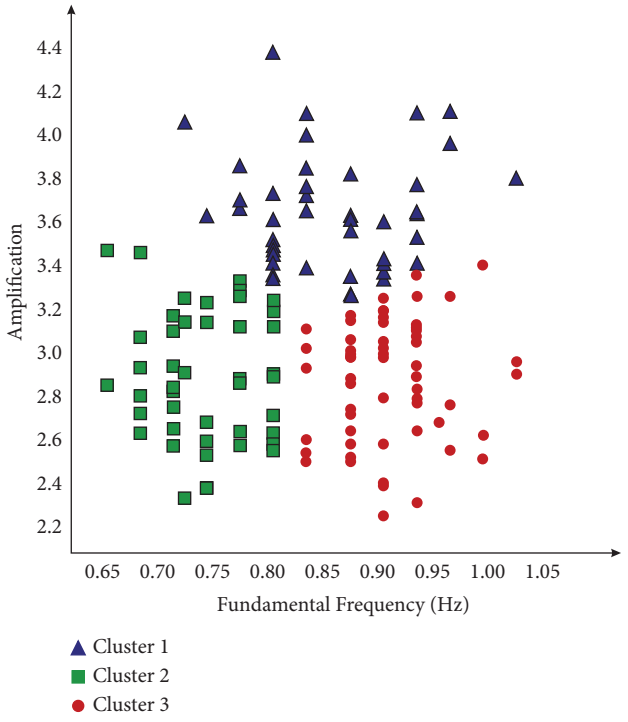

FIGURE 8: Scatter plot of frequency and amplitude by fuzzy C-mean clustering technique.

The horizontal-to-vertical spectral ratio technique is favorable for the locations where the stiff layer is underlying a low rigid layer. Likewise, a clear variation in the mechanical properties of the hard rocks and soft soils is seen. Besides this, HVSR also identifies the difference between soft rocks and stiff soils based on the shape of curves. Soft soils reveal a clear curve like a bell in shape, while stiff to soft soils have flat-type curves.

Several studies have examined the variations in predominant frequencies. After such studies, it is

Table 5: Fuzzy C-mean clustering analysis of variance of three clusters for frequency and amplitude.

Amplitude (mm)			Frequency (Hz)		
Cluster 1	Cluster 2	Cluster 3	Cluster 1	Cluster 2	Cluster 3
3.41	2.93	3.53	0.91	0.84	0.94
3.12	2.8	3.86	0.78	0.69	0.78
3.07	2.83	3.63	0.69	0.94	0.88
3.36	2.57	3.66	0.94	0.78	0.78
3.26	2.86	3.64	0.97	0.88	0.94
3.34	2.64	4.06	0.91	0.94	0.73
3.12	2.5	3.72	0.81	0.88	0.84
3.03	2.54	3.7	0.91	0.84	0.78
3.35	2.71	3.26	0.88	0.81	0.88
3.39	2.33	3.02	0.84	0.73	0.84
2.98	2.63	3.73	0.91	0.69	0.81
3.17	2.88	3.13	0.72	0.78	0.94
3.19	2.53	3.6	0.91	0.75	0.91
3.01	2.55	3.45	0.88	0.97	0.81
3.34	2.94	3.52	0.81	0.94	0.81
3.23	2.63	3.61	0.75	0.81	0.88
3.05	2.72	3.61	0.91	0.88	0.81
3.25	2.57	3.65	0.73	0.75	0.84
3	2.68	3.76	0.91	0.75	0.84
3.11	2.86	3.43	0.84	0.78	0.91
3.14	2.88	3.77	0.91	0.88	0.94
2.98	2.65	3.82	0.91	0.72	0.88
3.47	2.38	4.11	0.81	0.75	0.97
3.17	2.84	3.17	0.72	0.72	0.88
3.11	2.59	3.24	0.94	0.75	0.81
2.96	2.89	4	1.03	0.81	0.84
3.19	2.39	4.1	0.81	0.91	0.94
3.14	2.72	3.8	0.75	0.69	1.03
3.25	2.94	3.96	0.91	0.72	0.97
3	2.74		0.88	0.88	
3.47	2.79		0.66	0.94	
3.41	2.4		0.94	0.91	
3.37	2.6		0.91	0.78	
3.05	2.57		0.94	0.72	
2.98	2.8		0.88	0.69	
3.16	2.9		0.91	0.81	
3.15	2.9		0.88	1.03	
3.12	2.5		0.94	0.84	
3.25	2.86		0.73	0.88	
2.64	2.51		0.78	1	
2.77	2.6		0.94	0.84	
3.14	2.64		0.73	0.88	
3.65	2.93		0.94	0.69	
2.78	2.52		0.94	0.88	
2.79	2.75		0.91	0.72	
3.63	2.89		0.75	0.94	
3.41	2.58		0.81	0.88	
3.36	2.6		0.81	0.81	
3.29	2.82		0.78	0.72	
3.1	2.58		0.72	0.91	
3.27	2.62		0.88	1	
3.26	2.31		0.78	0.94	
3.09	2.25		0.94	0.91	
3.46	2.91		0.69	0.73	
4.38	3.12		0.81	0.94	
3.33	2.85		0.78	0.66	
3.4	2.55		1	0.81	
3.56	3.49		0.88	0.81	
4.1	2.68		0.84	0.96	
3.85	2.98		0.84	0.88	
3.65	3.26		0.84	0.94	
3.07			0.94		
2.76			0.97		

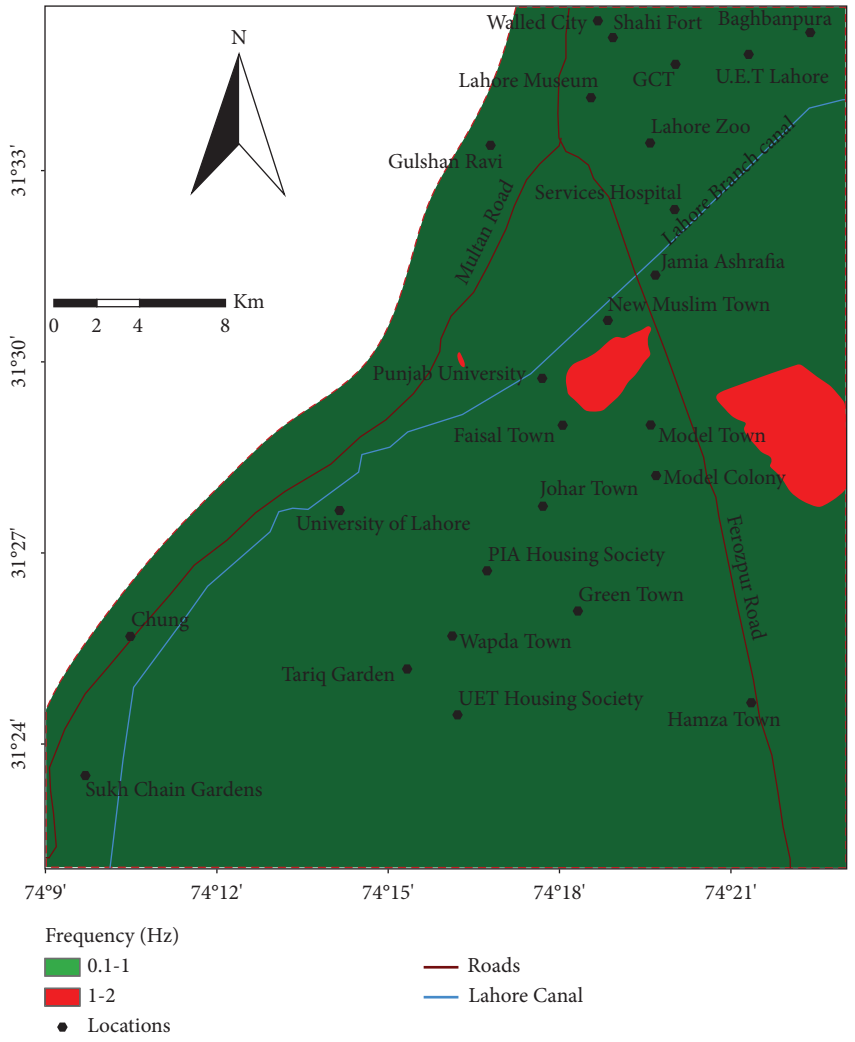

Figure 9: Map characterizing the study area into high (red color) and low (green color) reigns of predominant frequency.

concluded that predominant frequencies with low values mainly correspond to the presence of thick deposition and high-frequency values resemble thin deposition of sediments.

The maps produced could be used in the risk assessment (according to expected ground motion) of infrastructures in Lahore city. The results will be fruitful in designing underground constructions while considering subsurface soil.

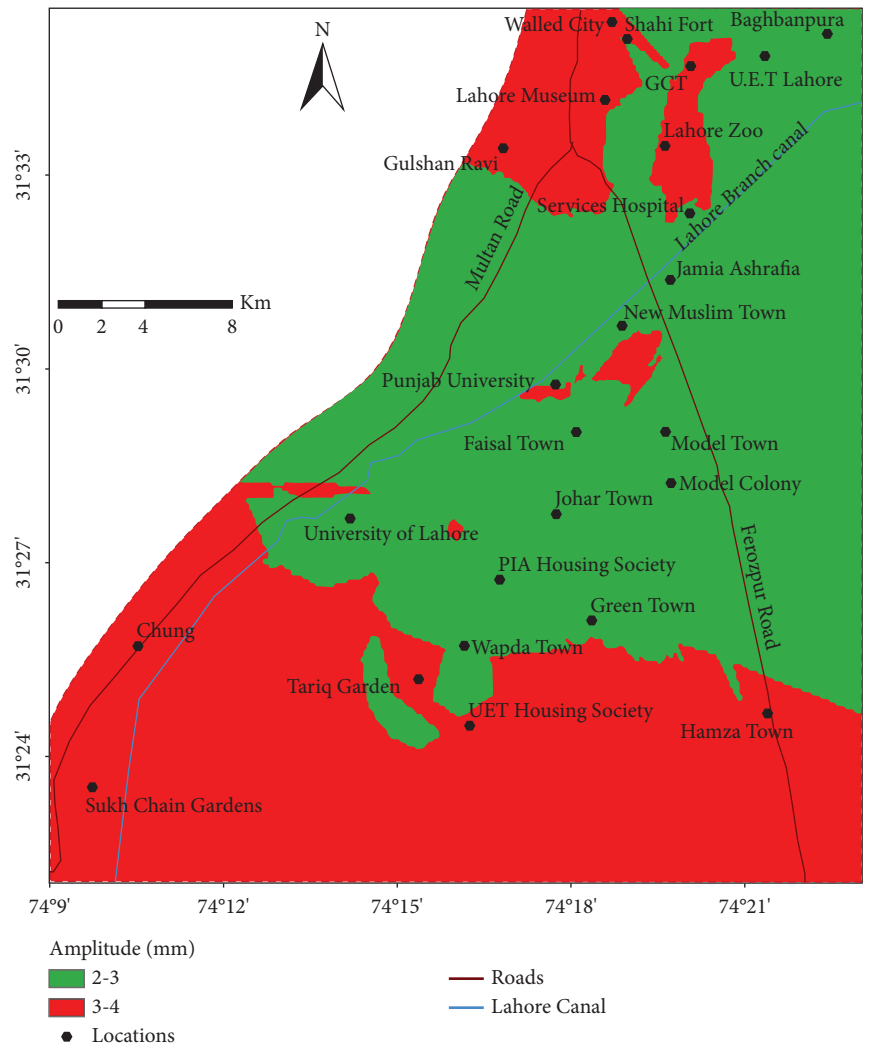

Figure 10: Spatial distribution of peak amplitude in the study area characterizing into low and high values.

The present study can be significant for the researchers integrating the seismic risk assessment studies with the clustering techniques.

5. Conclusions

The comparison of partitioned and non-partitioned clustering techniques recognized three cluster solutions that combine the data in terms of low, high, and medium risky sites.

However, the FCM reveals further study of additional variables, vulnerability index, sediment thickness, soil liquefaction, etc., that need to be carried out at certain locations that reveal more than one cluster feature.

The predominant frequency *f*o corresponds to the topography of the study area. High *f*o values reveal deposition of a thin alluvial cover, and the low-frequency belongs to a thick alluvial cover.

A major part of the study area possesses a very low-frequency range (0.5–1 Hz) corresponding to a thick alluvial cover.

High amplification is noted in the southern and western parts of the study area, which shows a thick cover of sediments and high ground resonance potential. This part of the city might be harmful against medium to large earthquakes and has a high potential to damage engineering structures in the area.

Lahore city has 60% of alluvium cover exhibiting resultant amplification of 2–3 mm in the central area and about 40% of 3–4 mm in the northern, southern, and southwest portions.

A major part of the city has stiff-soft sediments, while a small part of the city is covered with soil comprising of a very dense nature of soils.

The locations of the University of Punjab, Lahore Zoo, Lahore Museum, and many towns (i.e., Model town, Johar town, Garden town, Faisal town, etc.) are very sensitive to larger earthquakes if proper geotechnical and geophysical testing are not followed prior to construction.

The construction along Lahore Branch Canal needs to carry out a special study in terms of soil liquefaction and groundwater seepage.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The authors are thankful to the National Centre of Excellence in Geology, University of Peshawar, for arranging field work for the data collection. This research received no external funding.

References

[1] S. Trevisani and J. Boaga, "Passive seismic prospecting in Venice historical center for impedance contrast mapping," *Environmental Earth Sciences*, vol. 77, no. 21, Article ID 733, 2018.

[2] R. K. Gupta, M. Agrawal, S. K. Pal, R. Kumar, and S Srivastava, "Site characterization through combined analysis of seismic and electrical resistivity data at a site of Dhanbad, Jharkhand, India," *Environmental Earth Sciences*, vol. 78, no. 6, Article ID 226, 2019.

[3] X. Zhao, K. Hu, S. F. Burns, and H. Hu, "Classification and sudden departure mechanism of high-speed landslides caused by the 2008 Wen-chuan earthquake," *Environmental Earth Sciences*, vol. 78, no. 5, Article ID 125, 2019.

[4] B. Aslam and F. Naseer, "A statistical analysis of the spatial existence of earthquakes in Balochistan: clusters of seismicity," *Environmental Earth Sciences*, vol. 79, no. 1, Article ID 41, 2020.

[5] A. A. Shah, T. Qadri, and S. Khwaja, "Living with earthquake hazards in south and southeast asia," *ASEAN Journal of Community Engagement*, vol. 2, no. 1, pp. 15–37, 2018.

[6] S. H. Sajjad, S. A. Waheed, T. Khan, N. Gilani, and S. A. Waheed, "Natural Hazards and related contents in curriculum of Geography in Pakistan," *Asian Journal of Natural and Applied Sciences*, vol. 3, no. 2, pp. 40–48, 2014.

[7] A. J. Durrani, A. S. Elnashai, Y. Hashash, S. J. Kim, and A. Masud, "The Kashmir earthquake of October 8, 2005: a quick look report," *MAE Center CD Release*, vol. 05-04, no. 1, pp. 5–7, 2005.

[8] K. Rehman, S. M. T. Qadri, A. Ali, A. Ali, and S. Ahmed, "Analysis of the devastating Kashmir earthquake 2005 aftershocks," *Arabian Journal of Geosciences*, vol. 9, no. 5, Article ID 379, 2016.

[9] A. Ansal, G. Tonuk, and A. Kurtus, "Microzonation for urban planning," in *Earthquakes and Tsunamis, Geotechnical, Geological, and Earthquake Engineering*, A. T. Tankut, Ed., vol. 11, pp. 133–152, 2009.

[10] S. Khan and M. A. Khan, "Seismic microzonation of islamabad–rawalpindi metropolitan area, Pakistan," *Pure and Applied Geophysics*, vol. 175, no. 1, pp. 149–164, 2018.

[11] J. Roser and A. Gosar, "Determination of Vs30 for seismic ground classification in the ljubljana area, Slovenia," *Acta Geotechnica Slovenica*, vol. 1, pp. 71–76, 2010.

[12] S. Bonnefoy–Claudet, C. Cornou, P. Y. Bard et al., "V ratio: a tool for site effects evaluation. Results from 1-D noise simulations," *Geophysical Journal International*, vol. 167, pp. 827–837, 2006.

[13] M. D. Trifunac, "Site conditions and earthquake ground motion –A review," *Soil Dynamics and Earthquake Engineering*, vol. 90, pp. 88–100, 2016.

[14] D. M. Boore, "Estimating s(30) (or NEHRP site classes) from shallow velocity models (depths < 30 m)," *Bulletin of the Seismological Society of America*, vol. 94, no. 2, pp. 591–597, 2004.

[15] MonaLisa, A. A. Khwaja, and M. Q. Jan, "Seismic hazard assessment of the NW himalayan fold-and-thrust belt, Pakistan, using probabilistic approach," *Journal of Earthquake Engineering*, vol. 11, no. 2, pp. 257–301, 2007.

[16] Z. Rafi, C. Lindholm, H. Bungum, A. Laghari, and N. Ahmed, "Probabilistic seismic hazard of Pakistan, Azad-Jammu and Kashmir," *Natural Hazards*, vol. 61, no. 3, pp. 1317–1354, 2012.

[17] Y. M. A. Hashash, B. Kim, S. M. Olson, and I. Ahmad, "Seismic hazard analysis using discrete faults in Northwestern Pakistan: Part I–methodology and evaluation," *Journal of Earthquake Engineering*, vol. 16, no. 7, pp. 963–994, 2012.

[18] M. Waseem, C. G. Lai, and E. Spacone, "Seismic hazard assessment of northern Pakistan," *Natural Hazards*, vol. 90, no. 2, pp. 563–600, 2017.

[19] S. M. Talha Qadri, B. Nawaz, S. H. Sajjad, and R. A. Sheikh, "Ambient noise H/V spectral ratio in site effects estimation in Fateh jang area, Pakistan," *Earthquake Science*, vol. 28, no. 1, pp. 87–95, 2015.

[20] S. M. T. Qadri, S. H. Sajjad, R. A. Sheikh et al., "Ambient noise measurements in Rawalpindi-Islamabad, twin cities of Pakistan: a step towards site response analysis to mitigate impact of natural hazard," *Natural Hazards*, vol. 78, no. 2, pp. 1111–1123, 2015.

[21] S. Khan, M. Waseem, M. A. Khan et al., "Microzonation map of the Abbottabad basin and immediate surroundings," *Journal of Seismology*, vol. 24, no. 1, pp. 165–181, 2020.

[22] S. M. T. Qadri and O. A. Malik, "Establishing site response-based micro-zonation by applying machine learning techniques on ambient noise data: a case study from Northern Potwar Region, Pakistan," *Environmental Earth Sciences*, vol. 80, no. 2, Article ID 53, 2021.

[23] S. Khan, M. Waseem, and S. Jan, "Site response studies in Peshawar using the Nakamura technique of HVSR," *Arabian Journal of Geosciences*, vol. 14, no. 3, Article ID 193, 2021.

[24] V. Bolandi, A. Kadkhodaie, and R. Farzi, "Analyzing organic richness of source rocks from well log data by using SVM and ANN classifiers: a case study from the Kazhdumi formation, the Persian Gulf basin, offshore Iran," *Journal of Petroleum Science and Engineering*, vol. 151, pp. 224–234, 2017.

[25] A. Karpatne, I. Ebert-Uphoff, S. Ravela, H. A. Babaie, and V. Kumar, "Machine learning for the geosciences: challenges and opportunities," *IEEE Transactions on Knowledge and Data Engineering*, vol. 31, 2018.

[26] M. A. Shafiq, Z. Long, H. Di, and G. AlRegib, "A novel attention model for salient structure detection in seismic volumes," *Applied Computing and Intelligence*, vol. 1, no. 1, pp. 31–45, 2021.

[27] P. Fränti, "Efficiency of random swap clustering," *J Big Data*, vol. 5, no. 1, pp. 13–29, 2018.

[28] P. Fränti and S. Sieranoja, "How much can k-means be improved by using better initialization and repeats?" *Pattern Recognition*, vol. 93, pp. 95–112, 2019.

[29] P. Capizzi and R. Martorana, "Analysis of HVSR data using a modified centroid-based algorithm for near surface geological reconstruction," *Geosciences*, vol. 12, no. 4, Article ID 147, 2022.

[30] P. Capizzi, R. Martorana, G. Stassi, A. D'Alessandro, and D. Luzio, "Centroid-based cluster analysis of HVSR data for seismic microzonation," in *Proceedings of the Near Surface Geoscience 2014—20th European Meeting of Environmental*

and Engineering Geophysics, Athens, Greece, September 2014.
[31] A. D'Alessandro, P. Capizzi, D. Luzio, R. Martorana, and N. Messina, "Improvement of HVSR technique by cluster analysis," Geoitalia, Pisa, Italy, 2013.
[32] Y. Nakamura, *Clear Identification of Fundamental Idea of Nakamura's Technique and its Applications*, 12WCEE, 2656, Upper Hutt, New Zealand, 2000.
[33] Y. Nakamura, "A method for dynamic characteristics estimation of subsurface using microtremor on the ground surface," *Railway Tech Res Inst Q Rep*, vol. 30, no. 1, pp. 25–33, 1989.
[34] S. Yaghmaei-Sabegh and B. Hassani, "Investigation of the relation between Vs30 and site characteristics of iran based on horizontal-to-vertical-spectral ratios," *Soil Dynam Earthq Engg*, vol. 128, Article ID 105899, 2020.
[35] G. Lombardo, G. Coco, M. Corrao et al., "Results of microtremor measurements in the urban area of Catania," *Italy. Bollettino di Geofisica Teorica ed Applicata*, vol. 42, pp. 317–334, 2001.
[36] J. Chávez, X. Goula, A. Roca, F. Mañá, J. Presmanes, and A. López-Arroyo, "Preliminary seismic risk assessment for Catalonia (Spain)," in *Proceedings of the the 11th European Conference on Earthquake Engineering*, CD-ROM. Balkema, Rotterdam, Netherlands, September 1998.
[37] J. F. Deng, S. Su, X. X. Mo, G. Zhao, Q. H. Xiao, and G. Y. Ji, "The sequence of magmatic-tectonic events and orogenic processes of the Yanshan Belt, North China," *32nd International Geological Congress on Geology of China*, vol. 78, pp. 260–266, 2004.
[38] J. B. Scott, R. Tiana, L. Barbara et al., "Shallow shear velocity and seismic microzonation of the urban Las Vegas, Nevada, basin," *Bulletin of the Seismological Society of America*, vol. 96, no. 3, pp. 1068–1077, 2006.
[39] Sesame, "Guidelines for the implementation of the H/V spectral ratio technique on ambient noise vibrations measurements, processing and interpretation," *SESAME Euro Res Project WP*, vol. 12, pp. 1–62, 2004.
[40] Addinsoft, "XLSTAT Statistical and Data Analysis Solution," 2022, https://www.xlstat.com/en.
[41] J. C. A. Dunn, "A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters," *Journal of Cybernetics*, vol. 3, no. 3, pp. 32–57, 1973.
[42] J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*, Plenum Press, New York, NY, USA, 1981.
[43] M. Z. Rodriguez, C. H. Comin, D. Casanova et al., "Clustering algorithms: a comparative approach," *PLoS One*, vol. 14, no. 1, Article ID e0210236, 2019.
[44] S. S. Hussain, I. M. Javed, M. Bakhsh, and A. Iqbal, "Analysis of different clustering algorithms for accurate knowledge extraction from popular DataSets," *Information Sciences Letters*, vol. 9, no. 1, 2020.
[45] C. F. Christian, V. U. Constantino, B. E. Pilar, D. S. Maria Paz, and T. L. Javier, "A comparison of fuzzy clustering algorithms applied to feature extraction on vineyard," *Inteligencia artificial revista iberoamericana de inteligencia artificial*, vol. 1, Article ID 778, 2011.
[46] H. Suyal, A. Panwar, and A. Singh Negi, "Text clustering algorithms: a review," *International Journal of Computer Application*, vol. 96, no. 24, pp. 36–40, 2014.
[47] S. Yaghmaei-Sabegh, "A novel approach for classification of earthquake ground-motion records," *Journal of Seismology*, vol. 21, no. 4, pp. 885–907, 2017.
[48] NEHRP, *Commentary on the Guidelines for the Seismic Rehabilitation of Buildings*, Federal emergency management agency, Washington, DC, USA, 1997.