A Quantitative Site-Specific Classification Approach Based on Affinity Propagation Clustering

Sayed Moustafa
National Research Institute of Astronomy and Geophysics
Farhan Khan
King Saud University
Mohamed Metwaly (mmetwaly70@yahoo.com)
King Saud University https://orcid.org/0000-0001-8770-0306
Eslam A.Elawadi
King Saud University
Nassir Al-Arifi
King Saud University

Research Letter

Keywords: seismic site classification, HVSR, clustering, affinity propagation

Posted Date: July 20th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-44214/v1

Version of Record: A version of this preprint was published at IEEE Access on January 1st, 2021. See the published version at https://doi.org/10.1109/ACCESS.2021.3128284.

Abstract

Investigations made to evaluate the site effect characteristics and develop a reliable site classification scheme have received the paramount importance for the planning of urban areas and for a reliable site-specific seismic hazard assessment. This paper presents a new approach for site classification based on affinity propagation (AP) along with a selected set of representative horizontal to vertical spectral ratio (HVSR) curves inside King Saud University (KSU) campus. Measurements of the ambient vibrations were performed to cover the entire campus area by about 307 stations with 20 minutes recording length and sample rate of 128 Hz for each station to satisfy the criteria for reliable and unambiguous HVSR results. Predominant period values were used for identifying of site response and subsequent site classification. Empirical equations from the literature relating frequency of HVSR peak to average shear wave velocity in the upper 30m, commonly used as a proxy for site classification, were found to be unreliable, making site classification difficult. To overcome this problem, Affinity propagation clustering algorithm is used. The obtained results illustrated that microtremors spectral ratios can be remarkably robust tool in determining site effects. The survey results concluded to the preliminary seismic site classification map for the mapped area, which would be useful for future safe design of structures. Finally, the results presented in this study are encouraging prolongation of this type of study in other parts of Saudi Arabia using the microtremors data and site response functions.

1. Introduction

Site effect:

Effects of near surface soil conditions on ground motion and the consequent seismic response of the structures are common phenomena and produce huge effects on the characteristics of ground shaking during an earthquakes (Nakamura, 1989; Kockar andAkgun, 2012). Site effect analysis, is an indispensable part for the urban areas planning. It is associated with the surface geology and geotechnical characteristics of soil deposits and has a paramount importance on seismic ground motion (Bonnefoy-Claudetet al., 2006). Generally, site effects include the modification of the characteristics (in terms of amplitude, frequency content and duration) are controlled by anomalies in the mechanical properties of the shallowest layers of subsoil, when it consists of soft sediments, or by the shape of surfaces layer discontinuity close to or coincident with the topographic surface (Nakamura, 1989; Aki, 1993; Bard, 1999). Local site response can be estimated by theoretical and empirical approaches. Explanations of various procedures to estimate local site effects based on the geology and topography properties of the studied area is given in (Sánchez-Sesma et al., 2004; Atakan, 1996; Bonnefoy-Claudet et al., 2006). The evaluation of the site effects by using microtremors records as a tool has gained widespread popularity in the recent researches (Mucciarelli and Gallipoli, 2001). Microtremor studies initially proposed by Nagoshi and Igarashi (1971), and strongly emphasized by Nakamura (1989) and Nakamura et al. (2000), for inferring the dynamic site characteristics and associated subsurface soil structure at an observation point. Observation of microtremor measurements can give useful information on dynamic properties of the site such as predominant period and amplitude. Microtremor observations

are easy to perform, inexpensive and can be applied to regions with low seismicity as well; hence, predominant period and amplitude measurements can be used conveniently for seismic microzonation (e.g., Nakamura, 1989). It is worth pointing out at that, after Nakamura numerous authors around the world (Lermo and Chávez-García, 1993; Lermo and Chávez-García, 1994; Lachetl and Bard, 1994; Field and Jacob, 1993; Mucciarelli, 1998; Lontsi et al., 2015) tested the validity of the technique experimentally and theoretically. They proved that it is successful for estimating the site response of surface deposits using ambient noise as a source. Applying the horizontal-to-vertical spectral ratio (HVSR), for instance (Ibs-von Seht and Wohlenberg, 1999; Tian et al., 2019), obtained map thickness of soft sediments and estimate the frequency of the fundamental resonant mode and correctly predict the amplitude level (Nakamura et al., 2000). Panah et al. (2002), obtained classification based on the HVSR method and validate the application for site dominant frequency estimation and site classification. In spite of its limitations, many researchers (e.g., Qaisar et al., 2008; Wallinget al., 2009; Fnais et al., 2010) have proved from the research that HVSR method is a robust technique and preliminary step toward site characteristics estimation and microzonation in the areas of interest (Ibs-von Seht and Wohlenberg, 1999; Choobbasti et al., 2013; Moustafa, 2015; Al-Malki et al., 2015; Alyousef et al., 2015; Hellel et al., 2019; Clementeet al., 2019; Tian et al., 2019; Anbazhagan et al., 2019).

Seismic site classification is the most widely accepted practical method in the design of seismic resistant infrastructure (Fukushima et al., 2007; Di Alessandroet al., 2012). The most elementary technique for site classification is borehole data. These soil classes are based on the average shear wave velocity at the upper 30 meters of the subsurface successive materials, and the dominant period. Both parameters also affect the normalized elastic response spectra (Fukushima et al., 2007; Di Alessandro et al., 2012; Yaghmaei-Sabegh and Rupakhety, 2020). In the recent studies on site characterization, the horizontal-to-vertical spectral ratio (HVSR) technique is one of the successfully used methods as it gives an accurate reading of the site's predominant frequency (Nakamura, 1989; Bard, 1999; Di Alessandro et al., 2012).

Need for clustering:

The conventional method employed in identifying seismic site effect is intuitive and simple, but its corresponding interpretation and classification is very subjective due to personal experience bias. Given the complexity of the problem, one of the most promising approaches is to develop alternative techniques for automatic identification of seismic site classification schemes. Therefore, it is necessary to seek a clustering method with rapid convergence, good global search capability, simple and convenient implementation for engineering applications.

In this study an alternative approach is proposed to characterize sites inside King Saud University (KSU) campus by integrating unsupervised machine learning particularly clustering and the horizontal vertical spectral ratio (HVSR) technique for analyzing ambient noise data that could be applied to quantify site effects in the estimation of seismic site classes associated with seismic hazards (Anbazhagan et al., 2019). Fortunately, machine-learning approaches offer great promise in seismological research as a

means for integrating numerous data into personalized indices of diagnostic and prognostic value (Kong et al., 2019; MacCarthy et al., 2020).

Machine learning:

The utilization of machine learning (ML) techniques are becoming increasingly widespread in seismology, with applications ranging from identifying unseen signals and patterns to extracting features that might improve our physical understanding. A good survey for the ML methods is given in Kong et al. (2019).

Clustering:

Finding groups in data is an important step in many fields of computer science analysis, and there already exist algorithms to solve such problem (Adelfio et al., 2012). Clustering, as unsupervised data mining technique, deals with the problem of dividing a given set of entities into meaningful subsets (Hartigan, 1975; Xu and Wunsch, 2009). A survey for the clustering methods is presented in Sabau (2012). Usually, clustering methods comes with the constraint that a user has to specify the initial amount of clusters, and is very sensitive to that parameter (Gan et al., 2007; Xu and Wunsch, 2009). One solution to this problem is the Popular Affinity Propagation Clustering Algorithm proposed by Frey and Dueck (2007). The affinity propagation (AP) algorithm has been successfully employed in applications including face recognition, gene discovery, text mining and image segmentation (Tian et al., 2013; Wang et al., 2008; Fujiwara et al., 2011; El-Samak and Ashour, 2015). It uses a message-passing model between data points to form a collection of exemplars and respective clusters. It tries to solve the problem without needing to know the number of clusters beforehand, by only supplying a similarity criterion. It has the advantage of identifying clusters faster and with lower errors than other methods (Dueck, 2009). Compared with conventional clustering methods, this method is insensitive to initial cluster centers and is able to achieve a global optimum.

Study Area:

The present area of study is the campus of King Saud University (KSU), which is located to the northwest of Riyadh city, Saudi Arabia (Figure 1). The campus area witness accelerated construction expansion and new engineering construction projects. These activities intensified the needs of site response evaluation. Many researchers utilized the microtremor measurements in many parts of Saudi Arabia for estimating the site response, in terms of predominant frequency and concluded that the technique is very promising particularly for the densely populated areas like the new urban planning and reducing the vulnerability on the existing civil constructions (Al-Malki et al., 2015; Alyousef et al., 2015). Hence, this study would be the first indispensable step, which will provide insights into the seismic site response of the KSU campus for designing and implementing phases of new and development expansions.

Goals:

The present study aims to perform a pilot site effects study on the KSU campus area by using the horizontal to vertical spectral ratio (HVSR) in conjunction with available geological information to estimate site-specific predominant frequency and amplification. This is followed by integrating HVSR predominant frequency with *Affinity Propagation (*AP*)* algorithm (Frey and Dueck, 2007) to develop a site classification scheme. To achieve the goal of this study, the authors implement a more efficient vectorized version of the Affinity Propagation algorithm. The performance of our algorithm is compared to that of the *K-means* clustering (Wu, 2012), and discuss some advantages and disadvantages. Codes and sample data will be available after the final acceptance of the article for publishing. The results of the current study is indispensable to understand the seismic properties of the site. The results are used to prepare preliminary seismic site classification map of KSU campus. Such a map is required in conjunction with the necessity of considering the bedrock acceleration and the proper design spectrum preference.

2. Materials And Methods

Geological setting:

The mapped area is part of the Riyadh region, the most important political, economic, and densely populated region in the Kingdom of Saudi Arabia with a population of 7.6 million people (Figure 1). The Riyadh region positioned at elevation around 600 meters above mean sea level and it covers almost 1,913km^2 area in the central part of the Najd Plateau (Figure 1). Geographically, the plateau extends from the Awanid scarp on the northern edge, to the Kharj rise on its southern edge, and from the Dahna sand belt on its eastern edge, to the Tuwaiq Mountains on its western edge (Searle, 2019). It is largely Jurassic to Quaternary sediments, mostly composed of sandy limestone, siltstones and shales (Al-Refeai and Al-Ghamdy, 1994). Najd Plateau has a great thickness of continental and shallow marine limestone deposits. Vaslet et al. (1991) studied the geological setting of Riyadh and stated that the sedimentary section of the region can be characterized into surface geological system composed of mixture of aeolian clay, silt, sand, and gravel deposits whereas subsurface geology is composed of great thickness of shallow marine limestone with shale and clay intercalations. The local geologic section of the study site (Figure 2) has main stratigraphical succession of Upper Jurassic Arab Formation comprises two main members: Arab-D member to the western side and mixture of Arab (C and D members) to the eastern side (Soleimani et al., 2016).

Microtremor HVSR Method:

Microtremor is introduced by Nagoshi and Igarashi (1971) and later enhanced by Nakamura (1989) and Lermo and Chávez-García (1994). It is defined as a low amplitude ambient vibration of the ground caused by man-made or atmospheric disturbances, like wind, sea or ocean-waves, and vehicle vibrations that can describe the geological conditions of an area. Nakamura, 1989 utilized a simple Horizontal to Vertical Spectral Ratio (HVSR) measurement in three orthogonal directions (two horizontal and one vertical). It is based on assumption that the ratio of the horizontal spectrum and vertical surface vibration is a function

displacement (Lermo andChávez-García, 1994; Mucciarelli, 1998). According to the method proposed by Nakamura (1989), the dominant vibration period of a site (or engineering structure) can be determined from microtremor record which is recorded as triple component in time domain and transformed to spectral domain (frequency or period) and by making use of the ratio of horizontal and vertical components in this environment utilizing the following definition:

$$HVSR\,(f) = \frac{\sqrt{S_{NS}(f) \times S_{EW}(f)}}{S_Z(f)} \qquad (1)$$

Where $HVSR(f)$ is the horizontal to vertical spectrum ratio, $S_{NS}(f)$, $S_{EW}(f)$ and $S_Z(f)$ are the Fourier amplitude spectra in the North-South, East-West and Vertical directions, respectively (Lermo and Chávez-García, 1994).

Affinity Propagation Approach:

Affinity Propagation creates clusters by sending messages between pairs of samples until convergence. A dataset is then described using a small number of exemplars, which are identified as those most representative of other samples. The messages sent between pairs represent the suitability for one sample to be the exemplar of the other, which is updated in response to the values from other pairs (Figure 3). This updating happens iteratively until convergence, at which point the final exemplars are chosen, and hence the final clustering is given (Frey and Dueck, 2007; Dueck, 2009).

Affinity Propagation can be interesting as it chooses the number of clusters based on the data provided. For this purpose, the two important parameters are the preference, which controls how many exemplars are used, the damping factor, which damps the responsibility and availability messages to avoid numerical oscillations when updating these messages (Dueck, 2009). Clustering starts by estimating the Euclidean distance as a measure of the *similarity matrix*. In the similarity matrix, the similarity $s(i, k)$ (for two distinct points, indexed as i and k) indicates how well the data points with index k is suited to be the exemplar (i.e., point that serves as a cluster center) for data point i. It is calculated using negative squared Euclidean distance in our implementation: $s(i, k) = -\|X[i] - X[k]\|2$ for a data set X. For the similarity of a point to itself, i.e. the "self-preference" for being an exemplar, we provide two options in our algorithm: setting all equal to the minimum $s(i, k)$ or to the median $s(i, k)$ for all i, k with $i \neq k$. This is followed by evaluating the *responsibility matrix* $r(i, k)$ to quantify how well-suited k is to serve as the exemplar for i, relative to other candidates. To begin with, the availabilities are initialized to be zero. In later iterations, as a data point is assigned to an exemplar, the availability drops below zero and reduce the effect of similarity (Frey and Dueck, 2007).

$$r(i,k) \leftarrow s(i,k) - \max_{k' s.t. k' \neq k} \{a(i,k') + s(i,k')\} \qquad (2)$$

For points on the diagonal, $r(k, k)$ is calculated as the input preference that point k is chosen as an exemplar $s(k, k)$, minus the largest similarity between the point and all other candidate exemplars. The

availability matrix a(*i*, *k*) is used to represents how appropriate it is for *i* to pick *k* as its exemplar, taking into account other points' preference for *k* as an exemplar.

$$a(i,k) \leftarrow min\{0, r(k,k) + \sum_{i^J \text{s.t.} i^J \notin i,k} \max\{0, r(i^J,k)\}\} \quad (3)$$

For points on the diagonal, the following equation is used.

$$a(k,k) \leftarrow \sum_{i's.ti' \neq k} \max\{0, r(i,k')\} \quad (4)$$

Quality of the estimated clusters is determined using the sum of the responsibility matrix and the availability matrix c(i, k) ← r(i, k) + a(i, k) which is known as the criterion matrix c(i, k).

Data Collection and Analysis:

HVSR measurements from ambient noise recordings imply both reliability of the results and rapidity of data collection. In order to initiate the HVSR technique, microtremor measurements were carried out across majority of the KSU campus. Measurements of microtremor data were performed in a total of 307 points on KSU campus as represented by the red circles in Figure (4). HVSR measurements have been performed, quite uniformly distributed with a mean spacing of about 200-500 m. The utilized equipment for the free-field single-station microtremor data collection is a highly portable three-component seismic station called Tromino 3G ENGY from Micromed (Micromed, 2018) equipped with three velocity transducers. For each conducted measurement point, twenty minutes of ambient noise were recorded at the sampling rate of 128 Hz. The non-stationary portion of the recorded noise was exclude, thus considering only the low-amplitude part of signal, for the computation of the average HVSR function. Seismic station localities were carefully chosen to evade the impact of trees, sources of monochromatic noise and strong topographic landscapes.

To prevent data from industrial sources, the measurements were taken from late night to early morning hours along the study area. Examples of the collected data is given in figure (5). The whole measurements were achieved in accordance with the internationally accepted the Site Effects Assessment using Ambient Excitation (SESAME) Project guidelines and precautions (Bard and Participants, 2004; Acerra et al., 2004). For experimental, all the site conditions and parameters were recorded at each station (e.g. recording parameters, recording duration, measurement spacing, in-situ soil-sensor coupling, artificial soil-sensor coupling, sensor setting, nearby structures, weather conditions, and the available geological information). The Tromino data were checked for noise levels and then processed and interpreted. The quality of the obtained microtremor records in our study could be categorized into main four categories, high and intermediate quality raw signal of microtremor recording. In the current study, the SESAME guidelines and precautions (Bard and Participants, 2004; Acerra et al., 2004) were fulfilled in the 273 sites. Thirty-four sites did not fulfill the standard criteria and were rejected.

Rejections were mainly due to the presence of artificial noise or non-clear HVSR peak as the amplitude of the peak is too small. Example of each category is given in figure (5).

The data processing to get the HVSR curves at each individual site was implemented to determine the peak frequency of the soft sedimentary layer above a harder layer, providing a strong impedance contrast, and from that, estimate the depth of the observed interface between the soft and hard layers (Nakamura, 1989). The origin of the identified peaks of predominant frequency has been tested first to check whether it is industrial or natural and then only natural peaks were considered for interpretation. Prior to conducting the HVSR analysis, the GEOPSY damping toolbox (Wathelet, 2005) is adopted to detect the presence of any data originating from an industrial source utilizing a random decrement technique (Dunand et al., 2002; Koller et al., 2004). An industrial origin is concluded if the damping is much lower than 3% and the frequency is sustained. This detection is important to justify the validity of the recorded ambient noise data used in the HVSR analysis. For each microtremor waveform in the database, HVSR is calculated using the geometric mean of the 5% damped acceleration waveforms of the two horizontal components divided by the corresponding spectral ordinates of the vertical component. Fourier amplitude spectra was estimated and smoothed using the Konno and Ohmachi (1998) logarithmic window smoothing function before the horizontal spectra is divided by the vertical one. The frequency corresponding to the largest peak of the HVSR is taken to be representing the site predominant frequency. Complete HVSR analysis was performed using GEOPSY software (Wathelet, 2005) developed within the framework of the SESAME project.

3. Results

Site effect estimation:

HVSR data were properly processed to extract frequency and amplitude of peaks. The result of this analysis is a natural frequency and an amplification of the local site. Figure (6) shows examples of the estimated HVSR values at different sites in the study area. Graphs show sites amplifications as a function of frequencies.

In the current study, visual inspection were used to pick both the frequency and corresponding amplification, as we noticed that the utilized software in most of the cases could not pick the correct peak. So we did not relay on the plotted (gray) line, which indicate the peak selected by the software. Various clustering approaches will be implemented to help us in discriminating between observed peaks, which may be caused by source effects and those due to site effects (D'Alessandro et al., 2013; Capizzi et al., 2014; Martorana et al., 2018). Table (1) gives the general descriptive statistics of the estimated HVSR data.

Table 1: The estimated HVSR results and presented using median and 25th – 75th percentiles.

	minimum	maximum	mean (95%)
Frequency	1.37	23.0	8.50 (3.75–11.0)
Amplification	0.10	4.5	1.43 (0.90–1.65)

A bivariate distribution of frequency and amplification variables is shown in Figure (7). This multi-panel figure shows both the bivariate (or joint) relationship between the frequency and amplification variables along with the univariate (or marginal) distribution of each on separate axes.

After the identification of the peaks of the HVSR curves attributable to resonance effects, it is essential to delineate regions of similar features characteristics. As it is possible to identify on the HVSR curves, different peaks associable with different resonance frequencies of the investigated site, a data driven procedures are needed (Bragato et al., 2007; Martorana et al., 2018). For KSU campus, and to group peaks to be attributed to the same origin (e.g., stratigraphic, topographic, anthropogenic or other sources), a multi-parametric clustering procedure utilizing the *Affinity Propagation (*AP*)* clustering method (Frey and Dueck, 2007) has been adopted for better quantitative data-driven interpretation. Zhu et al. (2009), shows that the error of this clustering algorithm is much smaller than that of other algorithms such as *K-means* as it does not require the number of clusters to be determined or estimated before running the algorithm. The tendency of any HVSR data sample to become an exemplar is highly dependent on an input parameter referred to as *preference* or *p*. During the iterative procedure utilized to choose the optimal value of *p*, information is spread among the frequency and amplification points. Those data points are seen as a network in which messages are sent back and forth between pairs of frequency and amplification samples (see Figure 3). The algorithm was very sensitive to the input parameters, and did not produce unique number of exemplars and a systematic way to determine the correct optimal number of clusters is needed.

The global Silhouette coefficient (Rousseeuw, 1987) as a validity index is utilized to overcome the above problem. It is calculated as an average of all samples in a clustering and is given by (Rousseeuw, 1987):

$$Silhouette = \frac{b-a}{\max(a,b)} \qquad (5)$$

where *a* is the mean intra-cluster distance, and *b* is the distance between a point and the nearest cluster.

The variation of the Silhouette score with the AP parameters is depicted in Figure (8). 200 iterations were performed to calculate the Silhouette score utilizing the Squared Euclidean distance between two HVSR data points. Each iteration of Affinity Propagation consisted of updating all responsibilities given the responsibilities, and finally, combining availabilities and responsibilities to monitor the exemplar decisions and terminate the algorithm when these decisions did not change for 15 iterations. When the AP algorithm fails to converge, the damping factor λ was increased to avoid numerical oscillations. The λ close to one has a greater capability to guarantee the computational stability but at the cost of the time. In our experiment in this paper, λ =0.95 was used to guarantee the stability of the AP algorithm. The peak

or maximum value of the Silhouette score for the entire clustering near the center of the distribution indicate the optimal parameter value as seen in Figure (8). Using the optimal λ and p as suggested by the Silhouette score peak values, AP clustering procedure has created three clusters from the total estimated observations. The suggested clusters are groups of observations with similar frequency and amplification characteristics (see Figure 9).

Before we start interpreting the results of the Affinity Propagation clustering algorithm, we calibrate the algorithm results by comparing the clustering results with the K-means clustering algorithm.

The K-means clustering algorithm creates clusters by separating data points into k number of clusters or groups. The value of k is needed to be inputted into the algorithm. The clusters are determined by minimizing the inertia, or the within-cluster sum-of-squares. The inertia is a measure of how coherent the clusters are. By minimizing the intertia, the algorithm tries to minimize the difference between the mean value of a cluster and the values of points in the cluster. The inertia is not normalized, but lower values are better and zero is the optimum value.

For KSU data, to find the optimal number of clusters k, the elbow method is firstly utilized. In Figure (8) we can see that total within-clusters sum of squares have been plotted against the number of clusters K. The bend (elbow) in the graph is detected where the value of K is three. Therefore, for KSU data, the elbow method suggests three cluster solutions. Secondly, Silhouette plot as a method of interpretation and validation of consistency within clusters of data also suggest that three clusters is an optimal number of clusters for HVSR data as can be seen in Figure (8).

4. Discussion

Due to the increase of urban expansion and new engineering construction projects such as research centers, residential suites and hotels, the study area has intensified the importance of evaluation of site response study for detailed site investigation. To achieve these objectives, a pilot free-field site response study of the King Saud University Area (KSUA) is undertaken by using the horizontal to vertical spectral ratio (HVSR) in conjunction with available geological information to estimate dynamic soil properties and soil amplification ratio in the study area. However, the task becomes particularly difficult when it is possible to identify on the HVSR curves different peaks associable with different resonance frequencies of the investigated site. To overcome the aforementioned problems, we have implemented automatic procedures based on cluster analysis.

Affinity Propagation was implemented for a non-objective and data-driven site-specific characterization scheme. HVSR predominant frequency and corresponding amplification (Figure 7) were used as input data for clustering. The optimum number of clusters was found by the silhouette method. In this method, a graphical validation (Figure 8) was used for evaluating the number of clusters and comparing different scenarios (Rousseeuw, 1987). Silhouette plot suggest that three clusters is an optimal number of clusters for HVSR data as can be seen in figure (8) and figure (9). This was validated by comparing Silhouette plots of both Affinity Propagation and a well know K-means clustering algorithm (Figure 8).

Distribution of each cluster is shown in Figure (10), while histogram analysis of the three clusters is shown in Figure (11). The frequencies of the observed HVSR peaks are distributed in the wide range of 1.37–23 Hz (Figure 9 and Figure 11), but most of them (50%) are in the range 1.37–7 Hz, 32.32% are below 13 Hz and 17.68% above 13 Hz. The majority of HVSR spectral ratios calculated has first peak with the lowest fundamental frequency f_0 corresponding to the overall limestone thick deposits covering the area. The amplitudes of the HVRS peaks are distributed in a range of 0.5-4.5 for the delineated clusters. The observed amplitude peaks are related to the impedance contrast between the surface layer and the underlying bedrock, to the lateral heterogeneities, to the material damping of sediments and to the characteristics of the incident wavefield (Nogoshi andIgarashi, 1970). The spatial distribution of the observed frequency and corresponding amplification revealed subsurface columns below the recording stations consists of three layers; the first layer (shallow layer) is unsaturated overburden and rock fragments. The second layer is considered a compact saturated alluvium and/or fractured saturated limestone rocks. Hard and massive limestone bedrock (deep layer) is mapped below this layer.

Based on the results of clustering analysis and the collected geological information, it is quite clear that the first and second classless are mainly due to site-specific effect while the third may be due to the very thin soil layer brought to the area for agricultural proposes.

The first peak (first cluster) with the lowest frequency was interpreted to be the deepest compacted limestone layer while the other peaks corresponds to the highly weathered limestone and the superficial layers (see Figure 6). The second frequency peaks in the HVSR plot were observed at many sites. Identifying a second frequency peak (second cluster) is significant because amplification of ground motion may also occur at frequencies higher than the fundamental mode even when thick sediments are present (Parolai et al., 2004). There are two possible explanations for the second frequency peak. The first explanation is that it could be the first higher harmonic of the fundamental frequency of the site. This higher mode frequency would be expected to be at about three times the fundamental frequency (Anbazhagan et al., 2019; Parolai et al., 2004), which is what was observed (see Figure 6). The second explanation is that the second frequency peak could be a resonance of a soft soil layer over a shallow stiffer layer (Bodin et al., 2001).

Frequency Maps:

It is commonly recognized from different studies that the frequency of the HVSR peak replicates the predominant frequency of overburden sedimentary rocks. Its spectral amplitude mainly depends on the impedance contrast with deep seated bedrock and cannot be used as a quantify of amplification of the mapped site. However, the comparison with the standard reference site spectral ratio procedures results has revealed that the maximum amplitude of HVSR underestimates the actual site amplification (Acerra et al., 2004; Bard and Participants, 2004). Hence, we depend mainly on fundamental frequency rather that amplitude in our characterization of KSU campus.

The results of the Affinity Propagation clustering procedure, used to group peaks of HVSR average curves related to different sites, to identify areas characterized by site effects probably caused by the same buried structure. Considering the parameters of such peaks as sampling of spatial trends that are continuous on the KSU campus, it is possible to estimate the expected peak amplitude and frequency at each point of the area by two-dimensional interpolation techniques. The choice of the interpolation method depends on the analysis of the variability scale and the parameter to be studied. In our case, the context is relatively simple, where the geological variations are relatively mild and do not present a strong discontinuity. The natural neighbor interpolation method is used for spatial analysis of resonance frequencies distribution and to build the frequency map shown in Figure (12).

Interpretation:

The fundamental soil frequency map for the KSU campus provides valuable information for assessing soil-structure resonance in potential danger zones using the relationship between fundamental building frequency and the height of RC (Reinforced Concrete) structures .To investigated the validity of seismic site response characteristics estimated from microtremors by comparing them with the nearby KACST and Almalqa area studies. There is almost in good agreement. On closer comparing of the KACST study, the first peak varies from 7.85 to 8.72 Hz which clarifies the impedance contrast between the uppermost soil surface and the underlying completely weathered limestone, while the second peak ranges between 1.41 and 1.46 Hz that correspond the impedance contrast between the completely weathered limestone and the underlying hard limestone rocks. Whereas, based on the frequency of the H/V peak, the present study area has first peak with the lowest fundamental frequency for ranging from 3.0 to 8.25 Hz while the second peak f1 and the third peak f2 varies from 9.5 to 13.0 Hz and from 13.0 to 20 Hz respectively. As within the study area, the majority of H/V spectral ratios calculated has two peaks (f0 and f1) data which indicating agreement with the nearby KACST microtremor studies. The same comparison for the other nearby Almalqa area microtremor studies yielded results as: three zones of different frequencies, zone-1 Up to 1.7 Hz, then zone-2 from 1.7 Hz to 3.5 Hz and the lastly zone-3 from 3.5 Hz to 10 Hz. These results suggest that the microtremor H/V spectrum is a reliable tool to estimate the fundamental frequency.

5. Conclusions

Microtremor data collected for the present study suggests that Nakamura (1989) method of using H/V spectral ratios can be remarkably robust tool in determining site effects. Nevertheless, it is suggested that the microtremor technique be used alone or in conjunction to other geotechnical and geophysical techniques. As earlier, there is no site effect analyses have been carried out in the KSU campus, there is a great confidence that the outcome from present study can be highly significant and should be used as a guide to estimate the earthquake losses and related scenarios for designing and implementing phases of new constructions. Effect of surface geology and topography were noticed on the measured microtremor data sets. The majority of HVSR spectral ratios calculated in the study area has two peaks (fo and f1) data. Based on the obtained frequencies, the area was classified into three classes named as standard grounds, soft rocks and stiff rocks. Affinity propagation clustering, one of the unsupervised learning

techniques, is adopted to automatically determine the number of clusters and the corresponding cluster centers. The cluster center generates the strongest interference compared to other cluster members. A successive clustering procedure has been used to group the main HVSR clusters of peaks and categorized with areas characterized by site effects reasonably caused by the same lithological features. Assuming that the three main identified clusters contain peaks produced by resonance effects of layers with varying thickness, the possible trend of the top of the seismic bedrock was reconstructed by inversion of the HVSR curves constrained with geological and lithological information and considering a minimum lateral variability of the physical and geometrical parameters.

Results of the current microtrmor studies are consistent with previous work in the nearby KACST and Almalqa areas and suggest that the HVSR method is useful in evaluating ground seismic response. In addition, it worth to emphasis that the site effect analysis can help the decision makers to set the priorities in managing land uses, conducting programs for reducing the vulnerability of existing structures, enforcing building codes and planning for emergency response and long-term recovery. Finally, the results presented in this study are encouraging prolongation of this type of study in other urban areas of Saudi Arabia using the microtremors data and site response functions. As demonstrated, the site response variations are significant over very short distances, thus, strongly suggesting that estimation of any future earthquake loss scenarios for KSU or Riyadh city should be based on the site response functions obtained over a relatively dense grid of measurement points.

Declarations

Availability of data and materials:

Codes and the original acquired data will be available after the final acceptance of the article for publishing.

Competing interests:

The authors declare that they have no competing interests.

Funding:

This work is supported by the Research Supporting Project number (RSP-2020/89), King Saud University, Riyadh, Saudi Arabia.

Authors' contributions:

Sayed Moustafa planned and organized the article idea, acquiring the field data and did most of data analysis. Farhan Khan and Islam Elawadi contributed in the fieldwork, data analysis, exporting the figures in the final form and computing the mathematical parts. Mohamed Metwaly and Nassir Al-Arifi researched and wrote different parts of the paper, with contributions to text and figures. All authors read and approved the final manuscript.

Acknowledgements:

The authors would like to thank the Research Supporting Project number (RSP-2020/89), King Saud University, Riyadh, Saudi Arabia for funding this work.

References

Acerra, C., Aguacil, G., Anastasiadis, A., Atakan, K., Azzara, R., Bard, P.-Y., Basili, R., Bertrand, E., Bettig, B., Blarel, F., et al. (2004). Guidelines for the implementation of the h/v spectral ratio technique on ambient vibrations measurements, processing and interpretation. European Commission–EVG1-CT-2000-00026 SESAME.

Adelfio, G., Chiodi, M., D'Alessandro, A., Luzio, D., D'Anna, G., and Mangano, G. (2012). Simultaneous seismic wave clustering and registration. Computers & Geosciences, 44:60–69.

Aki, K. (1993). Local site effects on weak and strong ground motion. Tectonophysics, 218(1-3):93–111.

Al-Malki, M., Fnais, M., Al-Amri, A., and Abdelrahman, K. (2015). Estimation of fundamental frequency in Dammam city, eastern saudi arabia. Arabian Journal of Geosciences, 8(4):2283–2298.

Al-Refeai, T. and Al-Ghamdy, D. (1994). Geological and geotechnical aspects of Saudi arabia. Geotechnical and Geological Engineering, 12(4):253–276.

Alyousef, K., Aldamegh, K., Abdelrahman, K., Loni, O., Saud, R., Al-Amri, A., and Fnais,M. (2015). Evaluation of site response characteristics of king Abdulaziz city for science and technology, saudi arabia using microtremors and geotechnical data. Arabian Journa of Geosciences, 8(7):5181–5188.

Anbazhagan, P., Srilakshmi, K., Bajaj, K., Moustafa, S. S., and Al-Arifi, N. S. (2019). Determination of seismic site classification of seismic recording stations in the Himalayan region using HVSR method. Soil Dynamics and Earthquake Engineering, 116:304–316.

Atakan, K. (1996). A review of the type of data and the techniques used in empirical estimation of local site response. In International conference on seismic zonation, pages 1451–1460.

Bard, P. and Participants, S. (2004). The sesame project: an overview and main results. In Proc. of 13th World Conf. on Earthquake Engineering, Vancouver, BC, Canada, August, pages 1–6.

Bard, P.-Y. (1999). Microtremor measurements: a tool for site effect estimation. The effects ofsurface geology on seismic motion, 3:1251–1279.

Bodin, P., Smith, K., Horton, S., and Hwang, H. (2001). Microtremor observations of deep sediment resonance in metropolitan memphis, tennessee. Engineering Geology, 62(1):159–168.

Bonnefoy-Claudet, S., Cotton, F., and Bard, P.-Y. (2006).The nature of noise wavefield and its applications for site effects studies: A literature review. Earth-Science Reviews, 79(3):205–227.

Bragato, P., Laurenzano, G., and Barnaba, C. (2007). Automatic zonation of urban areas based on the similarity of h/v spectral ratios. Bulletin of the Seismological Society of America, 97(5):1404–1412.

Capizzi, P., Martorana, R., Stassi, G., D'alessandro, A., and Luzio, D. (2014). Centroid-based cluster analysis of hvsr data for seismic microzonation. In Near Surface Geoscience 2014- 20th European Meeting of Environmental and Engineering Geophysics, volume 2014, pages 1–5. European Association of Geoscientists & Engineers.

Choobbasti, A. J., Rezaei, S., and Farrokhzad, F. (2013). Evaluation of site response characteristics using microtremors. Gradevinar, 65:731–741.

Clemente, P., Delmonaco, G., Puzzilli, L. M., and Saitta, F. (2019). Stability and seismic vulnerability of the stylite tower at umm ar-rasas. Annals of Geophysics, 61:49.

D'Alessandro, A., Capizzi, P., Luzio, D., Martorana, R., and Messina, N.(2013). Improvement of HVSR technique by cluster analysis. In FIST GEOITALIA 2013–IX Forum di Scienze della Terra, pages 193–193. FIST.

Di Alessandro, C., Bonilla, L. F., Boore, D. M., Rovelli, A., and Scotti, O. (2012). Predominant-period site classification for response spectra prediction equations in Italy. Bulletin of the Seismological Society of America, 102(2):680–695.

Dueck, D. (2009). Affinity propagation: clustering data by passing messages. Citeseer. Dunand, F., Bard, P., Chatelain, J., Guéguen, P., Vassail, T., and Farsi, M. (2002). Damping and frequency from randomdec method applied to in situ measurements of ambient vibrations evidence for effective soil structure interaction. In 12th European conference on earthquake engineering, London.

El-Samak, A. F. and Ashour, W. (2015). Optimization of traveling salesman problem using affinity propagation clustering and genetic algorithm. Journal of Artificial Intelligence and Soft Computing Research, 5(4):239–245.

Field, E. and Jacob, K. (1993). The theoretical response of sedimentary layers to ambient seismic noise. Geophysical Research Letters, 20(24):2925–2928.

Fnais, M., Abdelrahman, K., and Al-Amri, A. (2010). Microtremor measurements in Yanbu city of western Saudi Arabia: A tool for seismic microzonation. Journal of King Saud University-Science, 22(2):97–110.

Frey, B. J. and Dueck, D. (2007). Clustering by passing messages between data points. Science, 315(5814):972–976.

Fujiwara, Y., Irie, G., and Kitahara, T. (2011). Fast algorithm for affinity propagation. In Twenty-Second International Joint Conference on Artificial Intelligence.

Fukushima, Y., Bonilla, L. F., Scotti, O., and Douglas, J. (2007). Site classification using horizontal-to-vertical response spectral ratios and its impact when deriving empirical ground-motion prediction equations. Journal of Earthquake Engineering, 11(5):712–724.

Gan, G., Ma, C., and Wu, J. (2007). Data Clustering - Theory, Algorithms, and Applications. SIAM.

Hartigan, J. A. (1975). Clustering algorithms. John Wiley & Sons, Inc.

Hellel, M., Oubaiche, E., Chatelain, J.-L., Bensalem, R., Amarni, N., Boukhrouf, M., and Wathelet, M. (2019). Efficiency of ambient vibration HVSR investigations in soil engineering studies: backfill study in the Algiers (Algeria) harbor container terminal. Bulletin of Engineering Geology and the Environment, pages 1–12.

Ibrahim, E., Kassem, O., and Al Bassam, A. A. (2012). Aeromagnetic data interpretation to locate buried faults in Riyadh region, Saudi Arabia. Scientific Research and Essays, 7(22):2022–2030.

Ibs-von Seht, M. and Wohlenberg, J. (1999). Microtremor measurements used to map thickness of soft sediments. Bulletin of the Seismological Society of America, 89(1):250–259.

Kockar, M. and Akgun, H. (2012). Evaluation of the site effects of the Ankara basin, Turkey. Journal of Applied Geophysics, 83:120–134.

Koller, M. G., Chatelain, J.-L., Guillier, B., Duval, A.-M., Atakan, K., Lacave, C., and Bard, P. (2004). Practical user guidelines and software for the implementation of the h/v ratio technique: measuring conditions, processing method and results interpretation. In Proceedings of the 13th world conference in earthquake engineering, Vancouver. Citeseer.

Kong, Q., Trugman, D. T., Ross, Z. E., Bianco, M. J., Meade, B. J., and Gerstoft, P. (2019). Machine learning in seismology: Turning data into insights. Seismological Research Letters, 90(1):3–14.

Konno, K. and Ohmachi, T. (1998). Ground-motion characteristics estimated from spectral ratio between horizontal and vertical components of microtremor. Bulletin of the Seismological Society of America, 88(1):228–241.

Lachetl, C. and Bard, P.-Y. (1994). Numerical and theoretical investigations on the possibilities and limitations of Nakamura's technique. Journal of Physics of the Earth, 42(5):377–397.

Lermo, J. and Chávez-García, F. J. (1993). Site effect evaluation using spectral ratios with only one station. Bulletin of the Seismological Society of America, 83(5):1574–1594.

Lermo, J. and Chávez-García, F. J. (1994). Are microtremors useful in site response evaluation? Bulletin of the Seismological Society of America, 84(5):1350–1364.

Lontsi, A. M., Sánchez-Sesma, F. J., Molina-Villegas, J. C., Ohrnberger, M., and Krüger, F. (2015). Full microtremor h/v (z, f) inversion for shallow subsurface characterization. Geophysical Journal International, 202(1):298–312.

MacCarthy, J., Marcillo, O., and Trabant, C. (2020). Seismology in the cloud: A new streaming workflow. Seismological Research Letters, 91(3):1804–1812.

Martorana, R., Capizzi, P., D'Alessandro, A., Luzio, D., Di Stefano, P., Renda, P., and Zarcone, G. (2018). Contribution of hvsr measures for seismic microzonation studies. Annals of Geophysics.

Micromed (2018). Manuale tromino eng tr-engy plus october 2018. Micromed, S.p.A, Treviso ITALY.

Moustafa, S. S. (2015). Microtremor analysis of marsa matrouh industrial area using horizontal to vertical spectral ratio method. EJGE, 20:1591–1602.

Mucciarelli, M. (1998). Reliability and applicability of Nakamura's technique using microtremors: an experimental approach. Journal of Earthquake Engineering, 2(04):625–638.

Mucciarelli, M. and Gallipoli, M. R. (2001). A critical review of 10 years of microtremor HVSR technique. Boll. Geof. Teor. Appl, 42(3-4):255–266.

Nagoshi, M. and Igarashi, T. (1971). On the amplitude characteristics of microtremors (part 2). Zisin, pages 26–40.

Nakamura, Y. (1989). A method for dynamic characteristics estimation of subsurface using microtremor on the ground surface. Railway Technical Research Institute, Quarterly Reports, 30(1).

Nakamura, Y., Gurler, E. D., Saita, J., Rovelli, A., and Donati, S. (2000). Vulnerability investigation of roman colosseum using microtremor. In 12th WCEE.

Nogoshi, M. and Igarashi, T. (1970). On the propagation characteristics of microtremors. J. Seism. Soc. Japan, 23:264–280.

Panah, A. K., Moghaddas, N. H., Ghayamghamian, M., Motosaka, M., Jafari, M., and Uromieh, A. (2002). Site effect classification in east-central of Iran. Journal of Seismology and Earthquake Engineering, 4(1):37.

Parolai, S., Richwalski, S. M., Milkereit, C., and Bormann, P. (2004). Assessment of the stability of h/v spectral ratios from ambient noise and comparison with earthquake datain the cologne area (Germany). Tectonophysics, 390 (1):57–73.

Qaisar, M., Karam, K., Talat, I., Tariq, M., and Daud Shah, M. (2008). Fateh Jang (Pakistan) earthquake of february 17, 1993: source mechanism and intensity distribution. J Himalayan Earth Sci, 41:45–52.

Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. Journal of Computational and Applied Mathematics, 20:53–65.

Sabau, A. S. (2012). Survey of clustering based financial fraud detection research. Informatica Economica, 16(1):110.

Sánchez-Sesma, F. J., Palencia, V. J., and Luzón, F. (2004). Estimation of local site effects during earthquakes: An overview. From Seismic Source to Structural Response: Contributions of Professor Mihailo D. Trifunac, pages 44–70.

Searle, M. (2019). Arabia: Geography, history and exploration. In Geology of the Oman Mountains, Eastern Arabia, pages 3–25. Springer.

Soleimani, B., Brumand, S., and Khoshbakht, F. (2016). Petrophysical evaluation of Arab formation using multimin, petrography and petrography carbonate methods in one of Iranian oilfields, Persian gulf. The International Journal of Science and Technoledge, 4(2):75.

Tian, B., Du, Y., You, Z., and Zhang, R. (2019). Measuring the sediment thickness in urban areas using revised h/v spectral ratio method. Engineering Geology, 260:105223.

Tian, Z., Tang, X., Zhou, M., and Tan, Z. (2013). Fingerprint indoor positioning algorithm based on affinity propagation clustering. EURASIP Journal on Wireless Communications and Networking, 2013(1):272.

Vaslet, D., Al-Muallem, M., Maddeh, S., Brosse, J., Fourniquet, J., Breton, J., and Le Nindre, Y. (1991). Explanatory notes to the geologic map of the ar Riyad quadrangle, sheet 24th, kingdom of Saudi Arabia. Saudi Arabian Deputy Ministry for Mineral Resources, Jeddah, Geosciences Map, GM-121, 54.

Walling, M. Y., Mohanty, W. K., Nath, S. K., Mitra, S., and John, A. (2009). Microtremor survey in talchir, india to ascertain its basin characteristics in terms of predominant frequency by Nakamura's ratio technique. Engineering Geology, 106(3):123–132.

Wang, K., Zhang, J., Li, D., Zhang, X., and Guo, T. (2008). Adaptive affinity propagation clustering. arXiv preprint arXiv:0805.1096.

Wathelet, M. (2005). Geopsy geophysical signal database for noise array processing. Software, LGIT, Grenoble.

Wu, J. (2012). Cluster analysis and k-means clustering: an introduction. In Advances in K-means Clustering, pages 1–16. Springer.

Xu, R. and Wunsch, D. (2009). Clustering. IEEE/Wiley. Yaghmaei-Sabegh, S. and Rupakhety, R. (2020). A new method of seismic site classification using HVSRcurves: Case study of the 12 November 2017 mw 7.3 Ezgeleh earthquake in Iran. Engineering Geology, page 105574.

Zhu, Y., Yu, J., and Jia, C. (2009). Initializing k-means clustering using affinity propagation. In 2009 Ninth International Conference on Hybrid Intelligent Systems, volume 1, pages 338–343. IEEE.

Figures

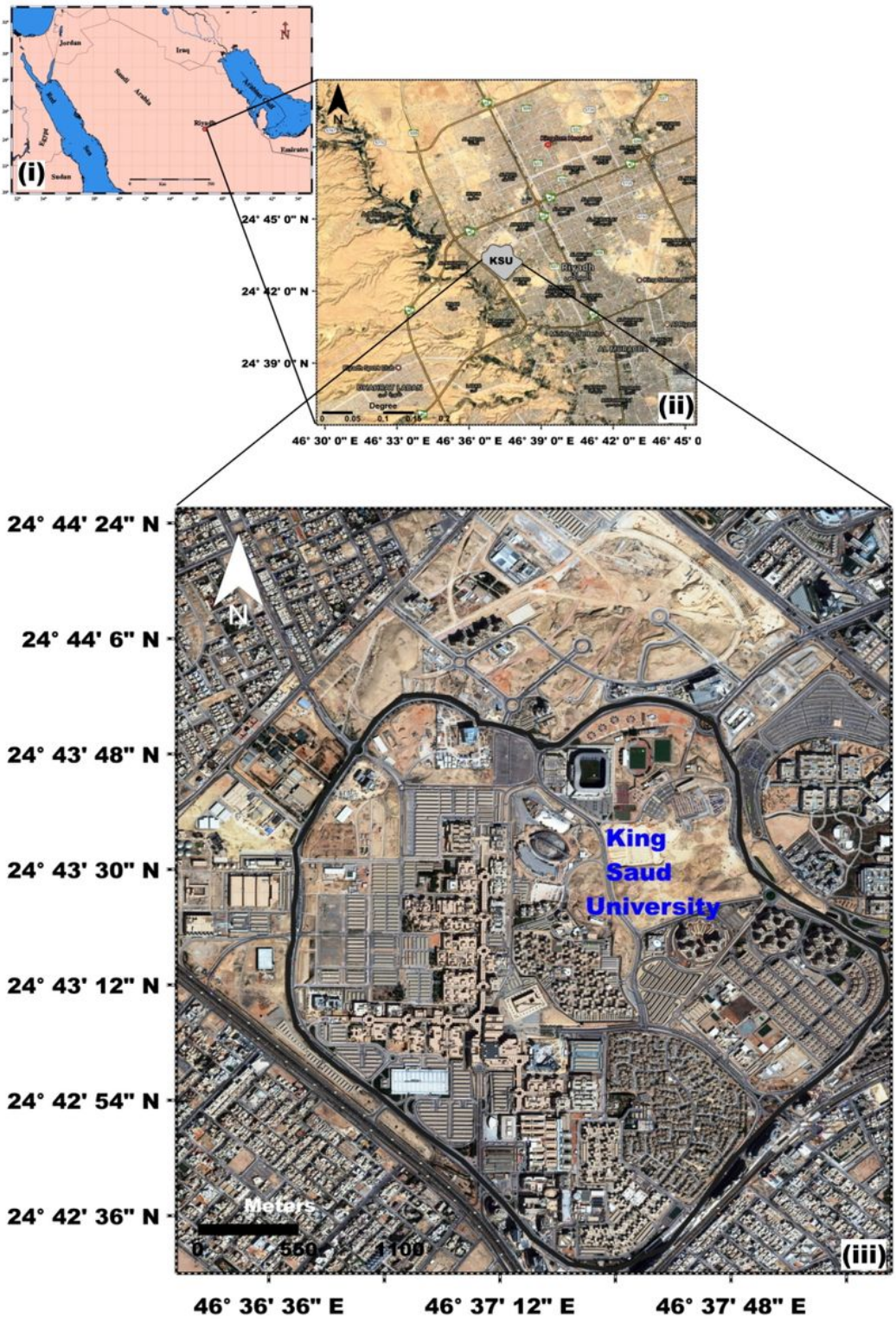

Figure 1

Location map of the study area. (i)- Location of Riyadh, (ii)- General location of King Saud University (KSU) inside Riyadh City, (iii)- General outline of KSU campus.

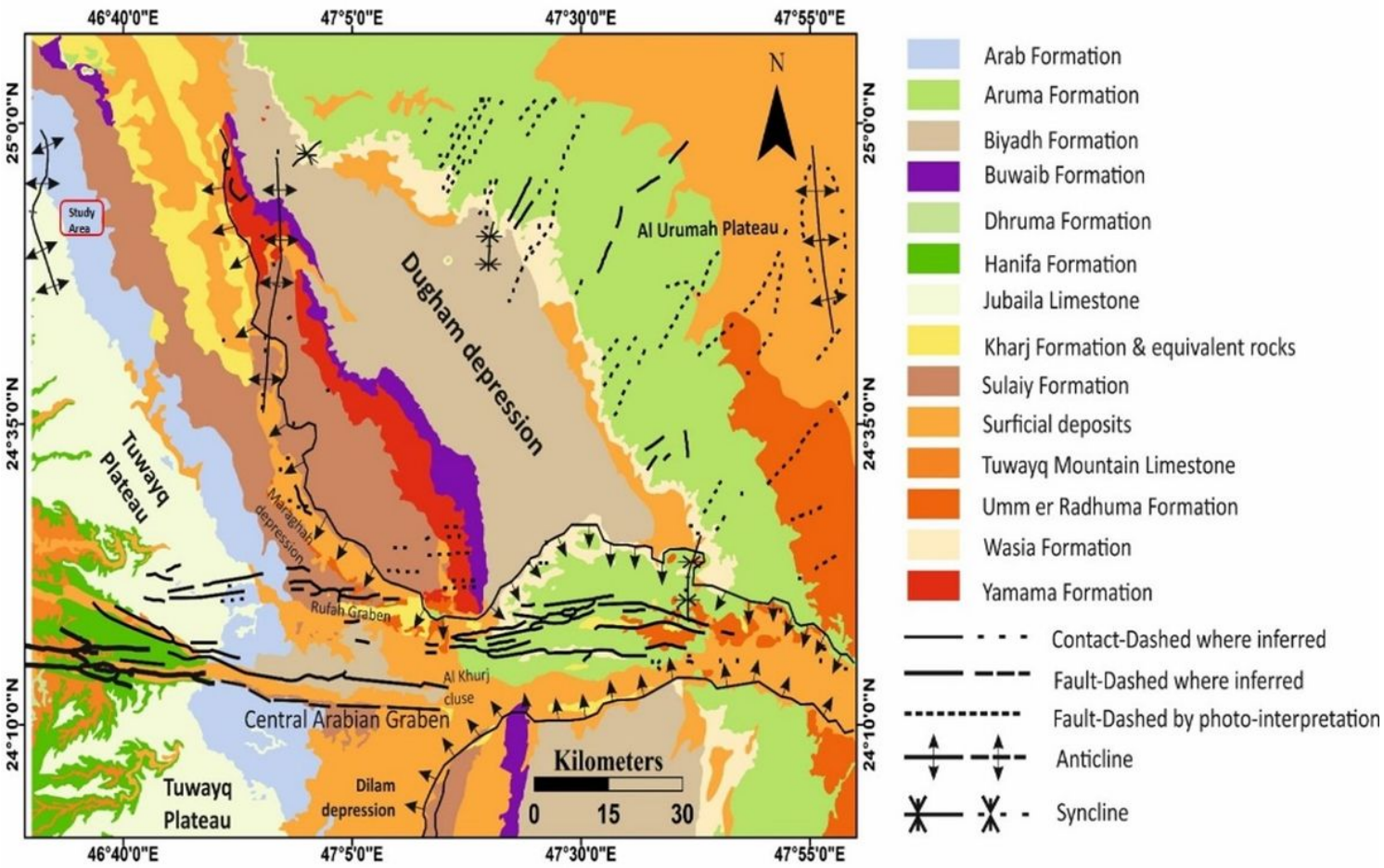

Figure 2

General Geological Setting of the Riyadh Region (Vaslet et al., 1991). Small red square indicate the location of KSU mapped area. Adopted and modified from Ibrahim et al. (2012).

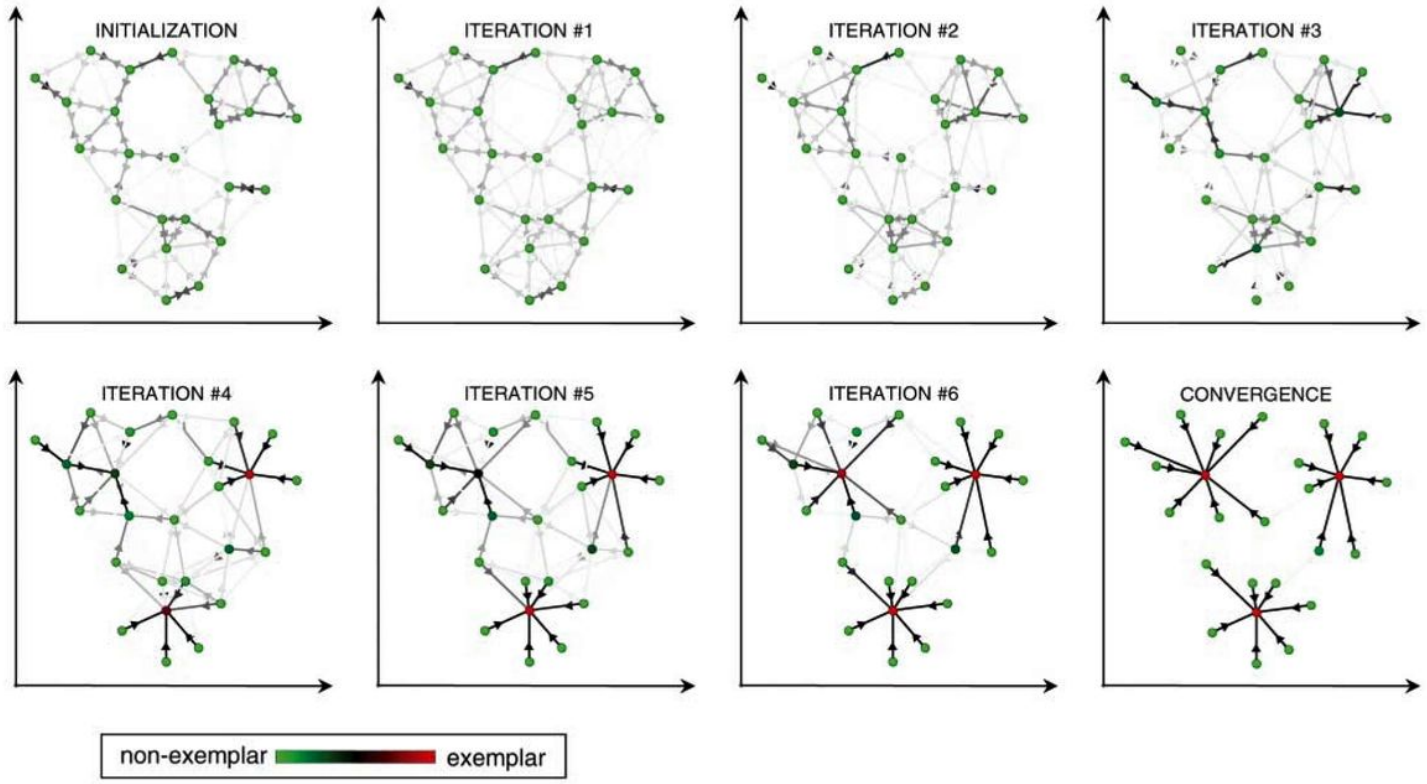

Figure 3

Visualized message-passing procedure (Frey and Dueck, 2007) showing the likelihood of a point being an exemplar, and its exemplar-relation to other points. A directed edge represents the likelihood for a point to choose another point as its exemplar, adopted and modified from Frey and Dueck (2007).

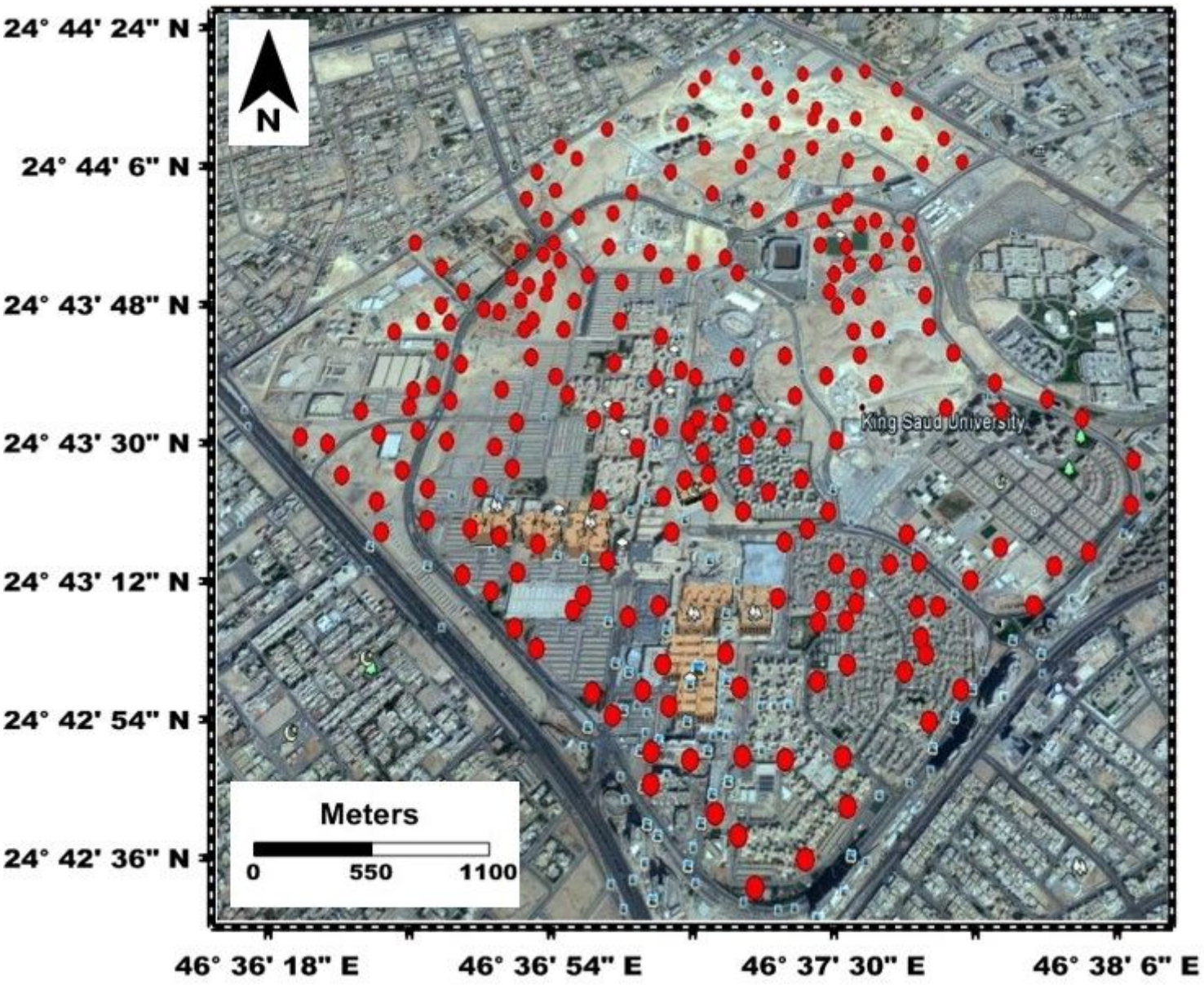

Figure 4

KSU map showing the location of microtremor sites surveyed.

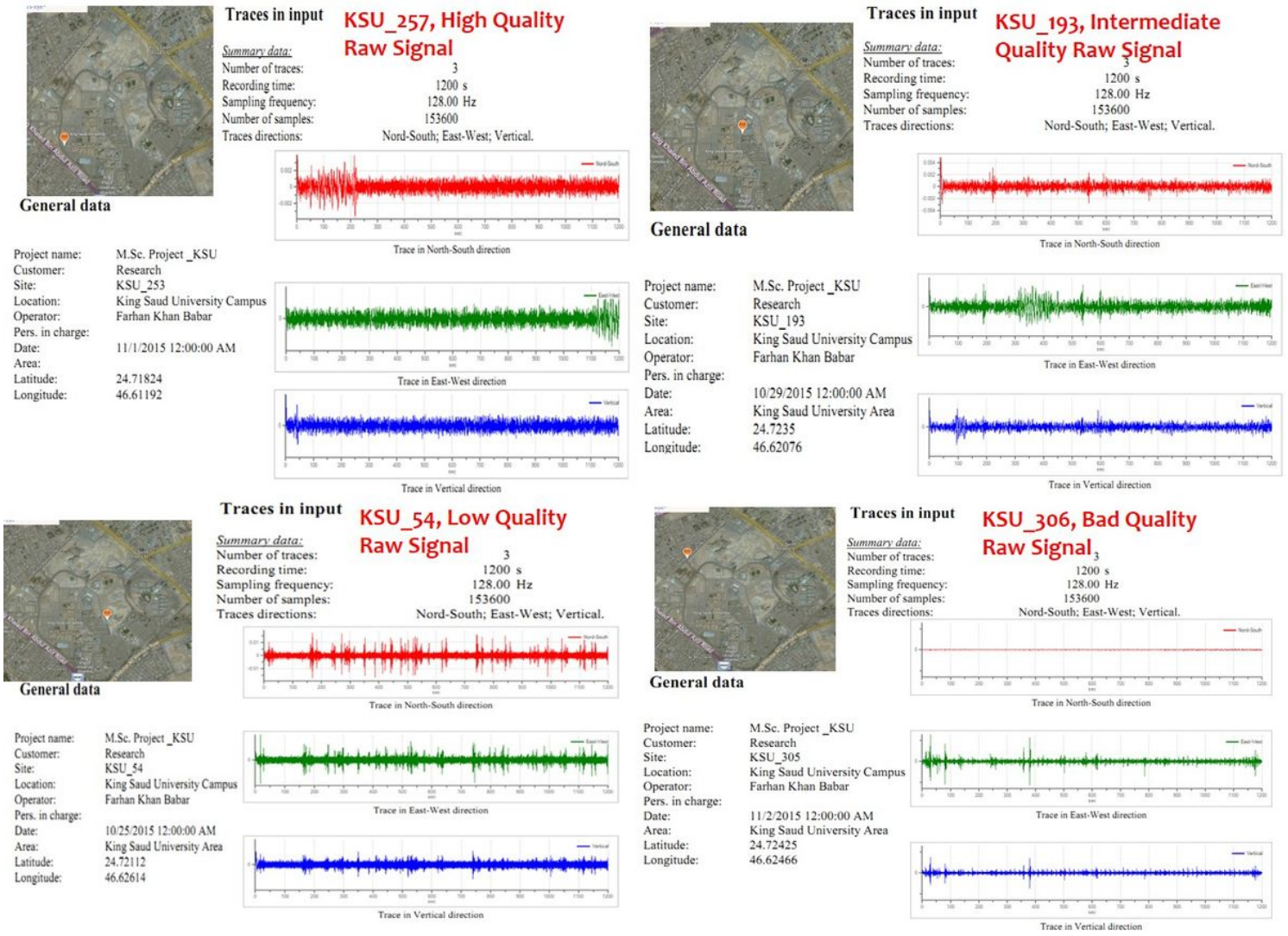

Figure 5

Example of high and intermediate quality (top panel) raw signal of Microtremor recording at site KSU257 and KSU193. Example of low and bad quality (bottom panel) raw signal of Microtremor recording at site KSU054 and KSU306.

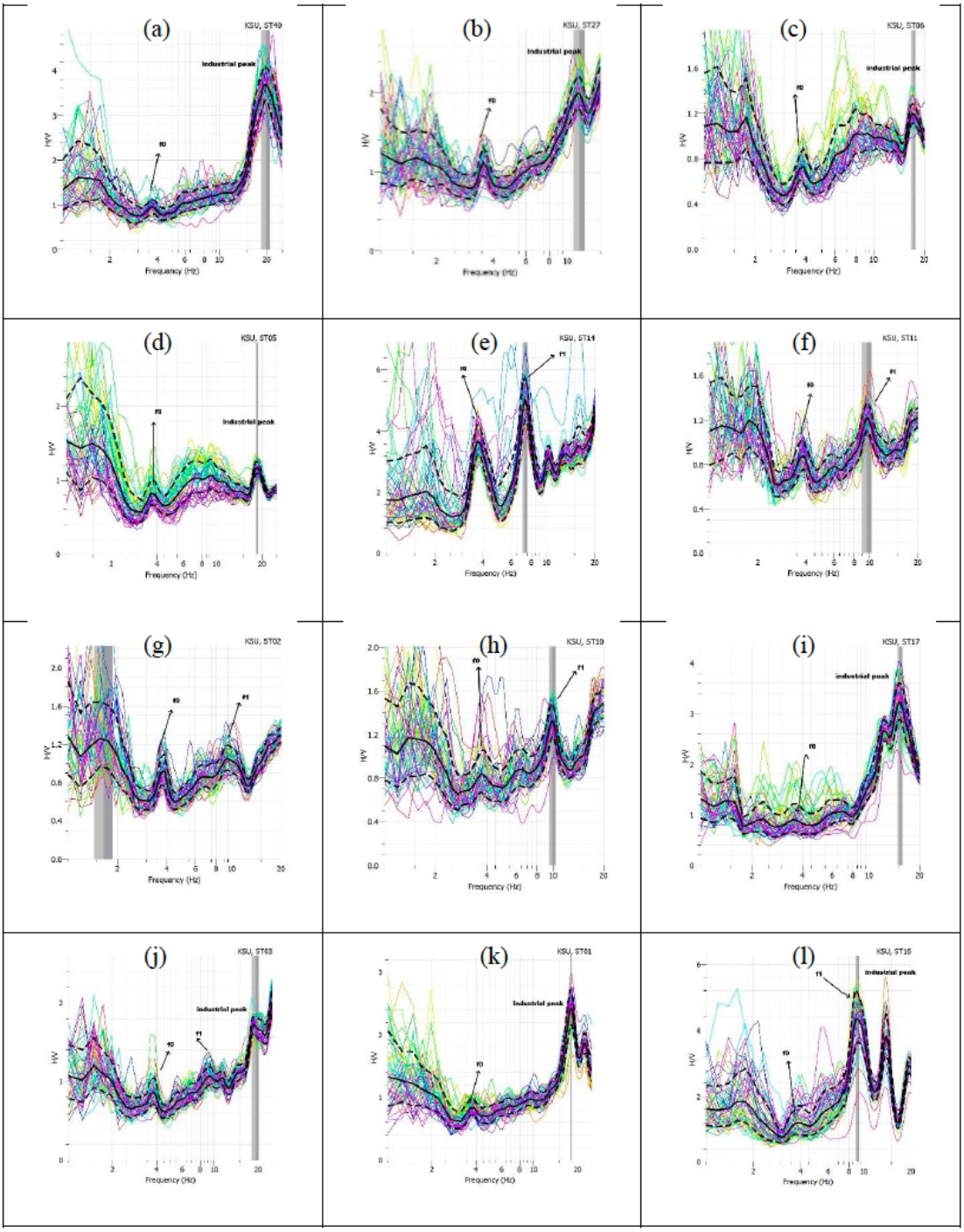

Figure 6

Examples of the estimated HVSR values at different sites in the study area. Graphs show sites amplifications as a function of frequencies.

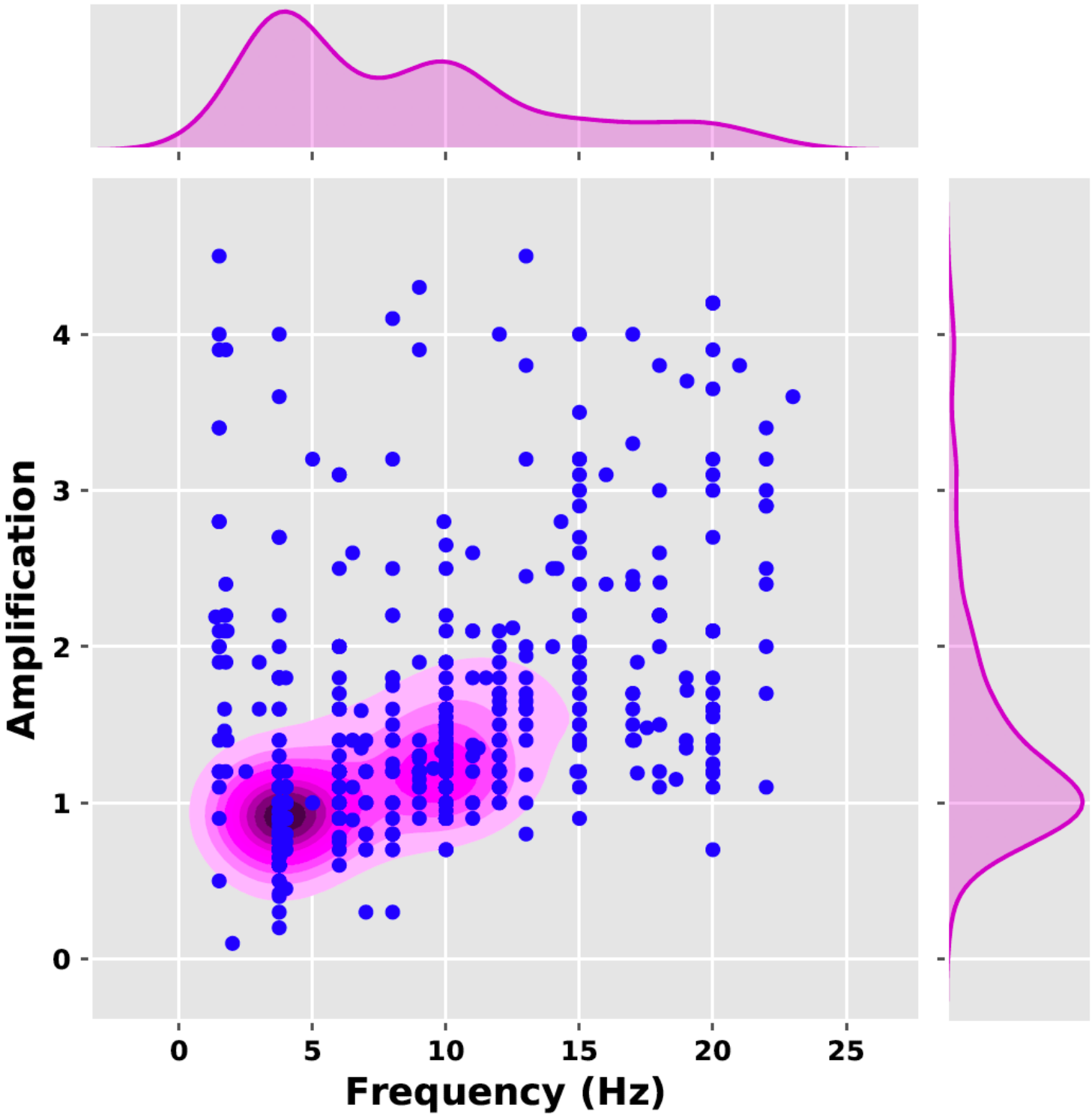

Figure 7

Bivariate distribution of HVSR frequency and amplification.

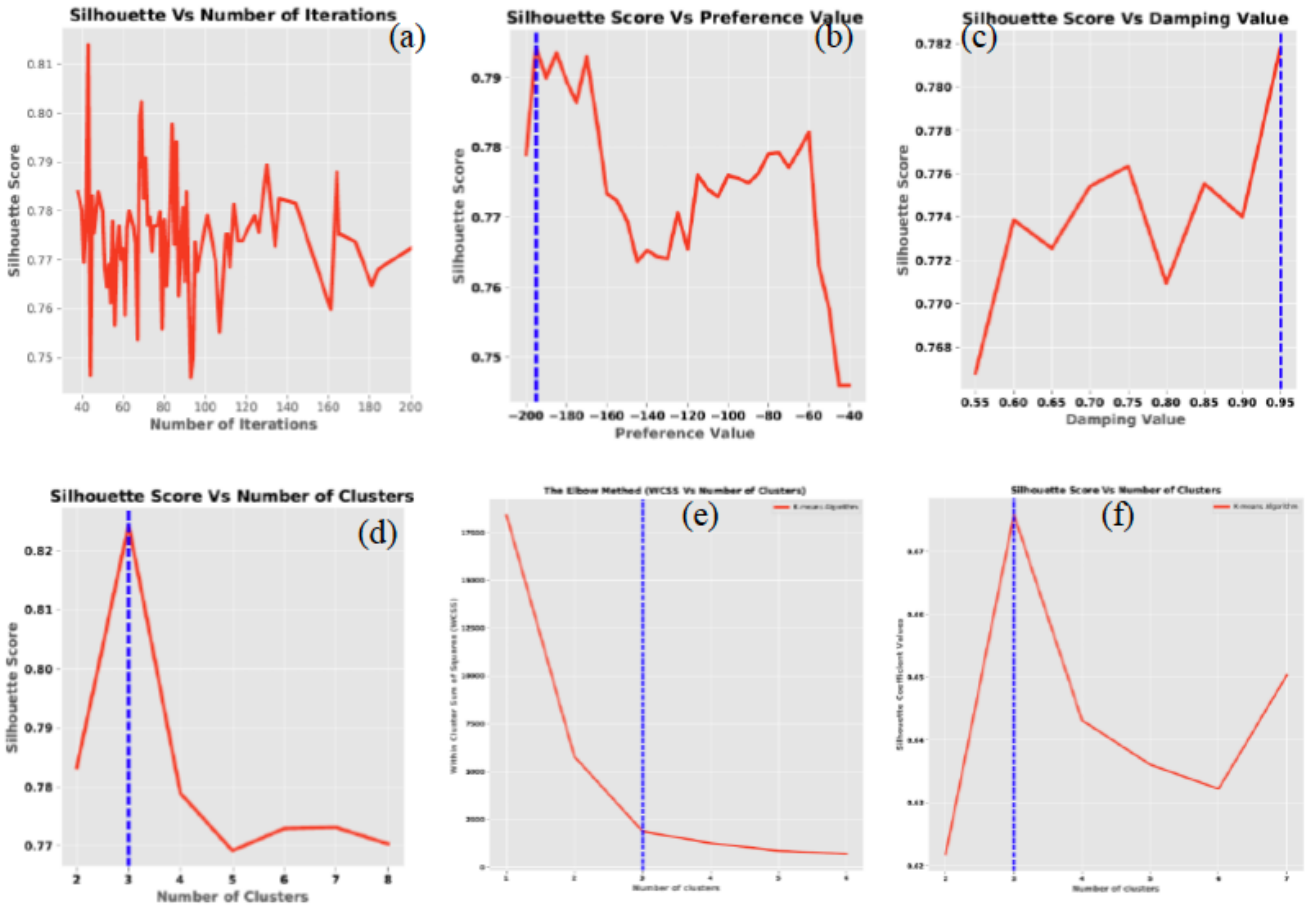

Figure 8

Variation of Silhouette score with various parameters.

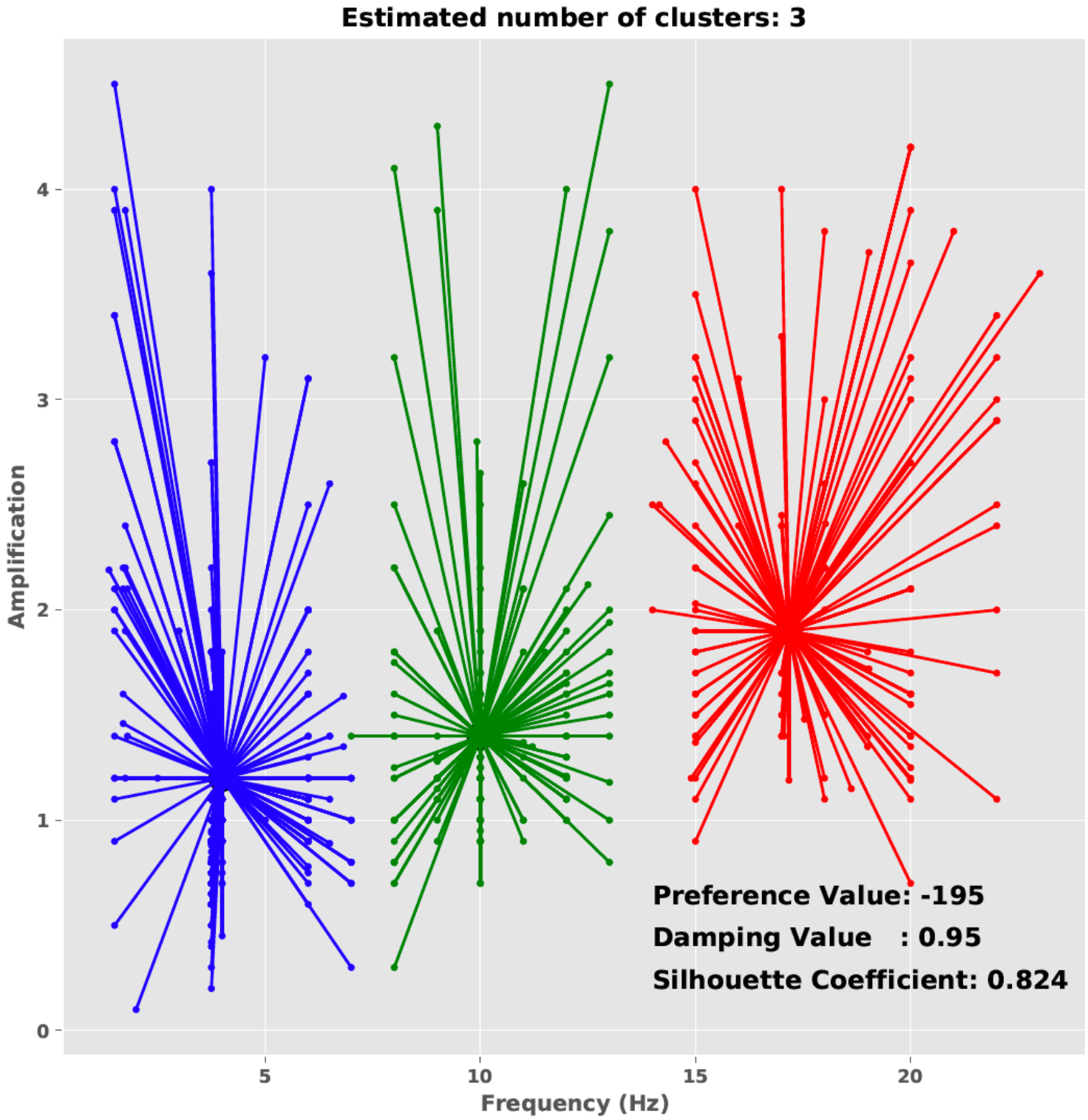

Figure 9

Three color-coded exemplars identified by Affinity Propagation clustering algorithm of HVSR data. Optimal clustering parameters are given.

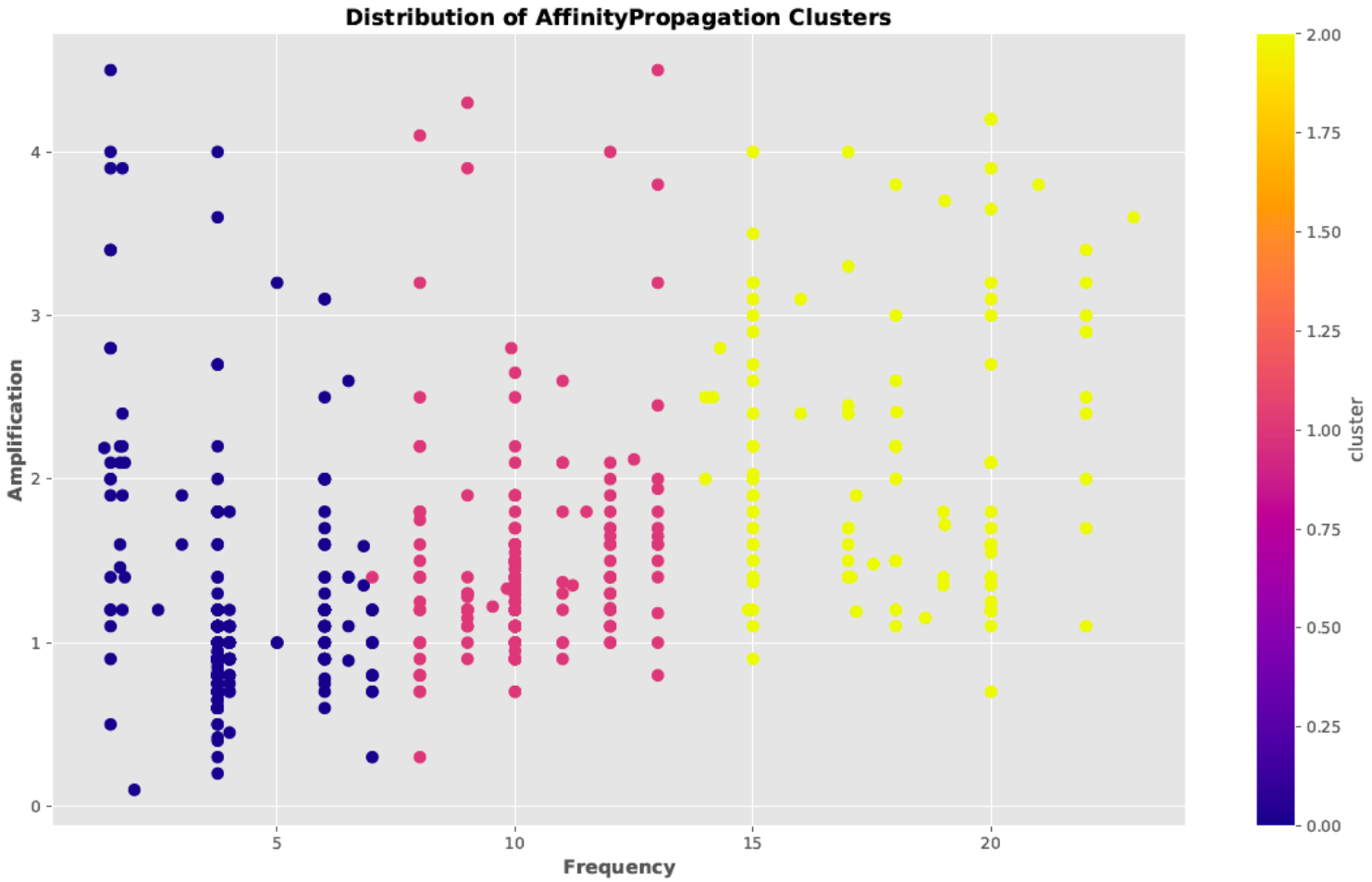

Figure 10

Three color-coded clusters distribution as identified by Affinity Propagation clustering algorithm of HVSR predominant frequency and amplification data.

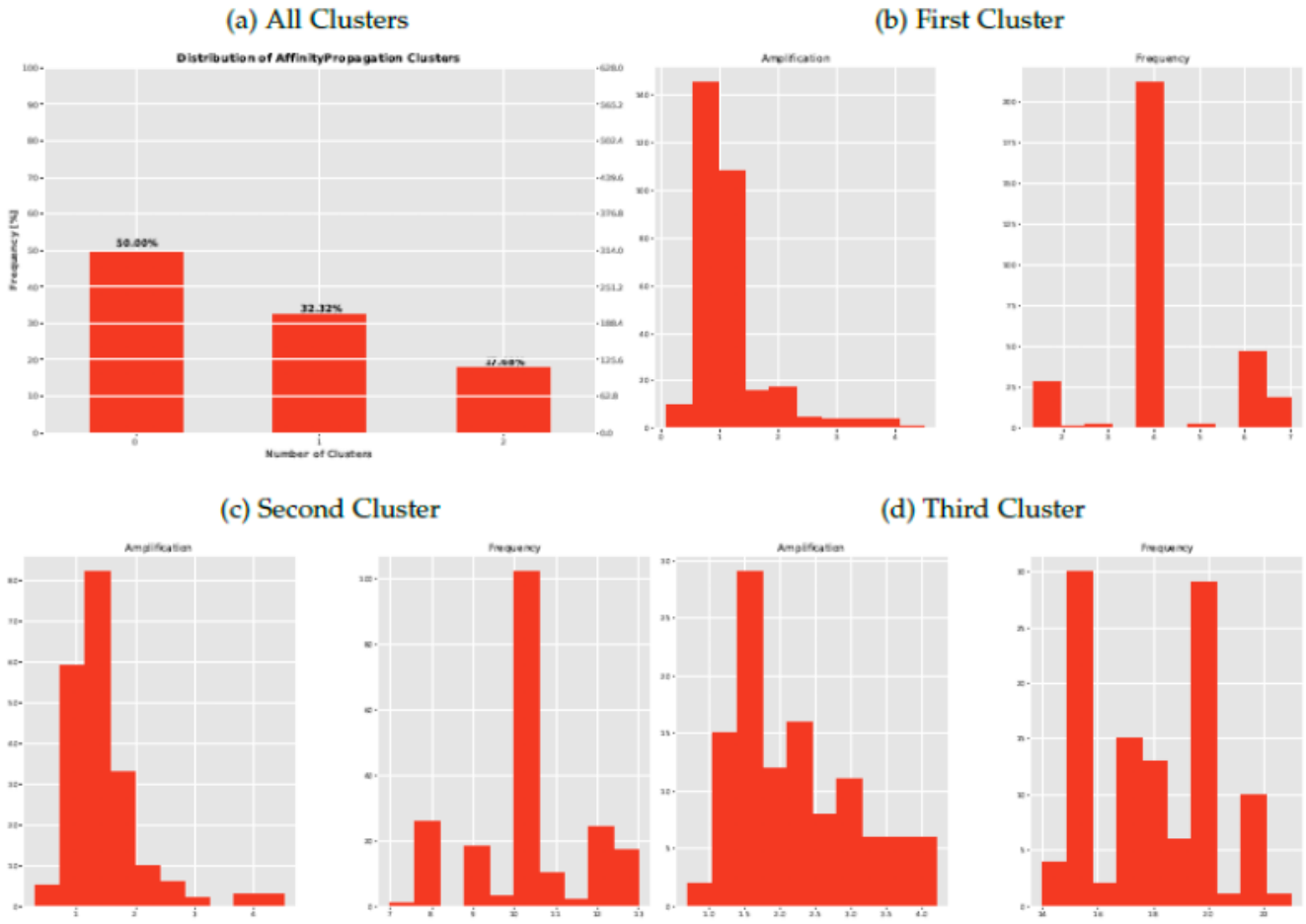

Figure 11

a- Spatial distribution of observed fundamental frequency and corresponding amplification of HVSR curves, b-Low class, c- Medium class, and d-High class of observed frequency and corresponding amplification inside KSU Campus.

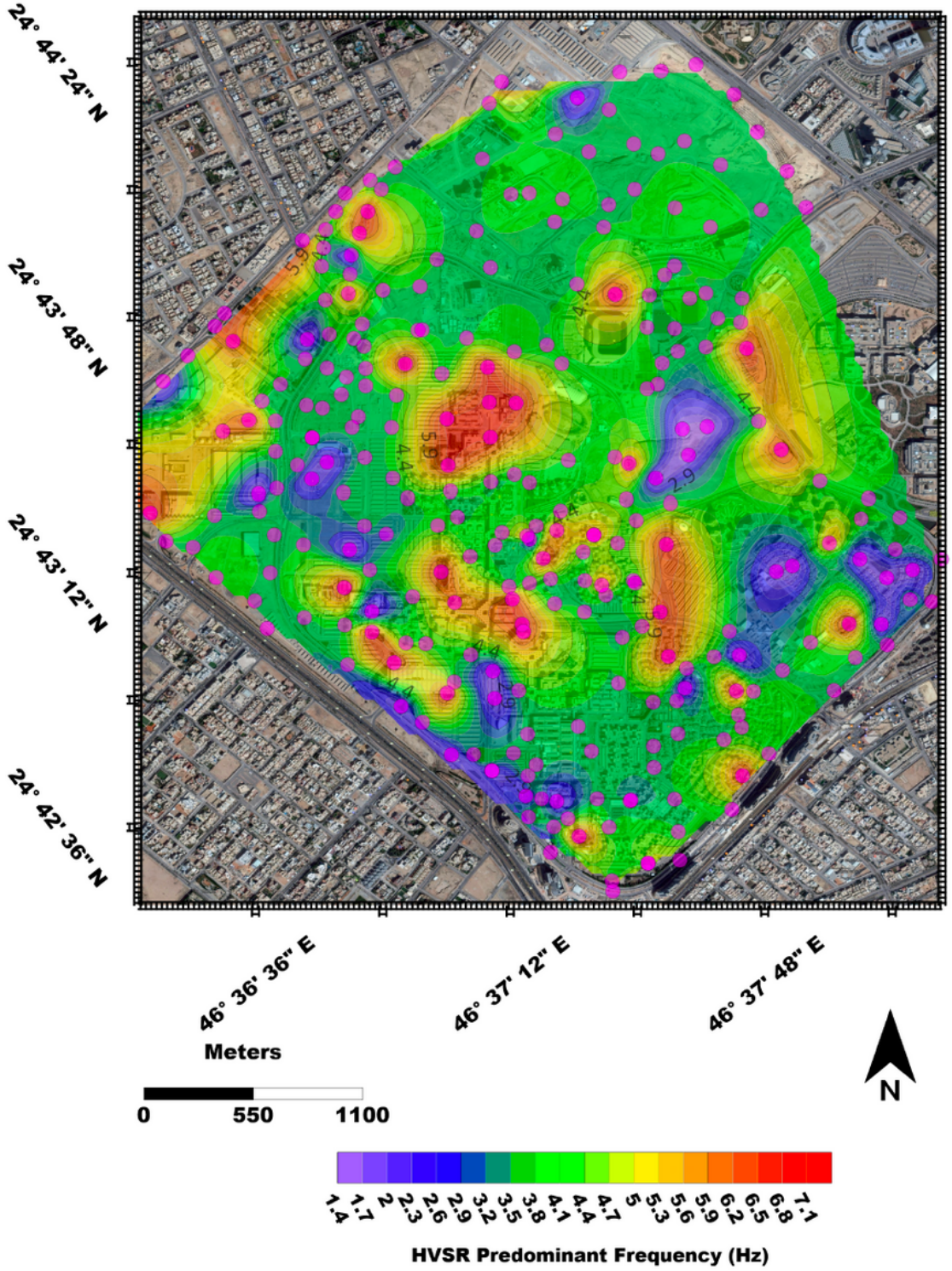

Figure 12

Map of the fundamental HVSR soil frequencies obtained for the KSU campus. Circles indicate microtremor measurements.