Originally published as:

Ullah, S., Bindi, D., Pittore, M., Pilz, M., Orunbaev, S., Modobekov, B., Parolai, S. (2013): Improving the spatial resolution of ground motion variability using earthquake and seismic noise data: the example of Bishkek (Kyrgyzstan). - *Bulletin of Earthquake Engineering*, 11, 2, 385-399

DOI: 10.1007/s10518-012-9401-8

Improving the spatial resolution of ground motion variability using earthquake and seismic noise data: the example of Bishkek (Kyrgyzstan)

S. Ullah, D. Bindi, M. Pittore, M. Pilz, S. Orunbaev, B. Moldobekov, S. Parolai

Abstract

Site response analysis plays an important role in seismic hazard and risk assessment, and in defining the optimal engineering design for civil structures. However, due to increasing urbanization, target areas are often too vast to be covered by standard approaches, resulting in large uncertainties in the spatial variability of the expected ground motion. Here, we propose a method to improve the spatial resolution of ground motion variability in terms of Standard Spectral Ratios (SSRs), using earthquakes recorded at a few selected sites for a relatively short amount of time, and seismic noise data collected over a denser grid, taking advantage of clustering and correlation analysis. The method is applied to Bishkek, Kyrgyzstan. Using the K-means clustering algorithm, three clusters of site response types have been identified, based on their similarity of SSRs. The cluster's site responses were adopted for sites where only single station noise measurements were carried out, based on the results of correlation analysis. The spatial variability of the site response correlates well with the main geological features in the area. In particular, variability is noted from south to north, consistent with both the changes in the thickness of the sedimentary cover over the basin and in the Quaternary material outcropping at the surface. This method has therefore the potential to improve the estimation of site effects at the local scale in the future.

S. Ullah, D. Bindi, M. Pittore, M. Pilz, S. Parolai

Helmholtz Centre Potsdam, GFZ German Research Centre for Geosciences, Helmholtzstraße 7, 14467 Potsdam, Germany

E-mail: ullah@gfz-potsdam.de

S. Orunbaev, B. Moldobekov

Central Asian Institute for Applied Geosciences CAIAG, Timur Frunze Rd.73/2, 720027 Bishkek, Kyrgyzstan

Keywords: microzonation, clustering analysis, correlation analysis, site response analysis

1 Introduction

Bishkek, which only developed as a city during the twentieth century, is the capital and main industrial center of the Kyrgyz Republic, with a population greater than 900,000 inhabitants. The country has witnessed several moderate to large earthquakes in the past. The most important include the August 3, 1885 M 7 Belovodski and January 3, 1911 M 8.2 Kemin events (Kalmetieva et al. 2009). These earthquakes generated in Bishkek macroseismic intensities as large as VII-VIII and VI MSK, respectively (Januzakov et al. 2003). Being located in a highly seismically active region and due to increase trends in urbanization, Bishkek is faced with a high seismic risk. In fact, according to Erdik et al. (2005), based on night-time occupancy and considering a 2% probability of exceedance in 50 years, the city could potentially experience an event that may cause 34,000 deaths with 90,000 injured persons who would need to be treated at hospitals. Furthermore, recent studies into possible worst case scenarios (Bindi et al. 2011) showed that for a magnitude 7.5 earthquake at the Issyk-Ata fault, located ~10 km from the city, over 48000 buildings are expected to be damaged with more than 22000 collapsing leading to 16624 casualties and 93447 injuries. Bindi et al. (2011) also outlined the importance of site amplification in determining the distribution of damages within the city. However, the authors could only use site amplification functions from 18 sites within the urban area, where seismic stations were deployed during a seismological experiment, while higher spatial resolution scenarios (including a high spatial resolution of building variability) require a much larger number of sites to gain an adequate knowledge of the variability of ground motion. We show in this study that this can be achieved by combining the results of earthquake data analysis and seismic noise recordings.

Seismic noise recordings have recently become an important tool for site effects estimation (e.g., Bard, 1999; Parolai et al. 2001; D'amico et al. 2004), using both single stations and arrays to retrieve information about the fundamental resonance frequency of a site and the S-wave structure of the uppermost layers. Picozzi et al. (2009) and Parolai et al. (2010), for example, used single station noise measurements to infer resonance frequency maps of Istanbul, Turkey, and Bishkek, Kyrgyzstan,

respectively. Parolai et al. (2005) and Boxberger et al. (2011) used inversions of surface waves from seismic noise array measurements to characterize the shallow geology in Cologne (Germany) and Bevagna (Italy), respectively, which in turn can be used to estimate the site response.

In several studies of urban site effects (Parolai et al. 2004; Pilz et al. 2009), temporary arrays of a few tens of stations have been used to record weak motion events for assessing the site response. Combining such results with those derived by seismic noise analysis (mainly Horizontal to Vertical Spectral Ratios, HVSRs), it is possible to obtain a better picture of ground motion variability over an investigated area. However, a quantitative analysis that allow a link between the site response estimated from earthquake recordings and the HVSR was not carried out and the possibility of exploiting this correlation, as a tool to improve the knowledge about the spatial variability of ground motion, was not investigated.

Different methods of clustering analysis have been recently applied, either to seismic noise or to earthquake HVSRs, to identify areas presenting similar HVSRs characteristic and likely site amplification. Bragato et al. (2007), following the approach of Rodriguez and Midorikawa (2002), cluster homogeneous areas based on the similarity of seismic noise HVSRs, following a Bayesian approach. Tselentis et al. (2011) used the Self-Organizing Map (SOM) technique for cluster analysis, based on the similarity of earthquake HVSRs. In both studies, clustering was identified, but since neither the HVSRs of noise nor of earthquakes represents the true site response (Field and Jacob 1995; Bindi et al. 2000; Bard 2004; Parolai et al. 2004), it is not possible to assign to a site an appropriate amplification function that could then be used when calculating ground motion and risk scenarios. Recently, Strollo et al. (2012) used correlation analysis between HVSRs of Single station Seismic Noise measurement Points (SSNP) and seismic noise of a few stations for which spectral intensity ratios (SIRs) had been calculated using earthquake records to spatially extend the SIRs to the large number of SSNPs.

In this study, starting from the approach of Strollo et al. (2012) and using earthquake recordings collected by 19 temporary stations installed on sites of interest within the Bishkek basin and seismic

noise data recorded at nearly 200 SSNP sites, we estimate the site response variability with a high spatial resolution in Bishkek. First, using the K-means algorithm for clustering analysis, we show that sites having similar SSRs also show similar noise HVSRs. Second, starting from this analogy, a correlation analysis is carried out between the noise HVSRs of SSNPs and those of the temporary stations. Using the results of the correlation analysis, the SSRs are finally extended to all SSNPs, allowing us to assign a site response over a much denser grid than was originally defined by the 19 temporary stations.

2 Geology of the region

Bishkek is located in the northwestern part of the Tien Shan Mountains. The region is known for its high level of seismicity, with the occurrence of large earthquakes within the context of a continent to continent collision (Bullen et al, 2001; Kalmetieva et al. 2009). This seismicity is associated with a complex faulting system resulting from the collision of the Indian sub-continent with Eurasia. Note that some of the world's largest earthquakes with $M > 8$ occurred in the Tien Shan region during the twentieth century (Bullen et al. 2001). The city of Bishkek lies over the central part of the Chu Basin, one of the largest depressions of northern Tien Shan. The basin extends about 50 km north-south and 150 km east-west (Bullen et al. 2001). The basin is bounded by the Kyrgyz Range to the south and the Chu-Ili mountains in the north. Much of the Kyrgyz Range is made up of Ordovician granite and fine grained Devonian sedimentary rocks. It is cut by many short, irregular imbricate thrust faults that are laterally truncated by strike-slip faults. The Paleozoic bedrock is thrust northward over ~ 4 km of younger Cenozoic molasses at the margin of the Chu basin. This deformation front is stepped northward into the Chu basin which exposes a thick succession of Neogene strata along the Issyk-Ata fault, located ~10 km north of the basin's boundary. The Paleozoic basement depth below the urban area of Bishkek generally decreases from the north (~1 km) to the south (~3 km). The following seven Tertiary formations overlay the Paleozoic bedrock (from bottom to top): (1) Kokturpak (siltstones), (2) Kokomeren (sandstones), (3) Sera Firma (claystones), (4) Dzhel'dysu (mudstones) and (5) Saryagach (sandstones and pebbly conglomerate), all belonging to the Paleogene, and (6) Chu

(mudstones and sandstones) and (7) Sharpyldak (coarse pebble conglomerate), belonging to the Neogene (Bullen et al. 2001, their Fig. 2). Quaternary sediments overlay the Cenozoic deposits with a thickness of about 200-300 m. Figure 1 shows the presence of different types of Quaternary covers with some tertiary outcropping (i.e., Paleogene and Neogene). The geology of the study area in the south consists of alluvial material of rubble and gravel forming 15-40 m thick outcrops. The southern half of the Bishkek area, which consist of alluvial gravels, rubble and sandy materials with a total thickness of about 25-50 m, constitute the shallowest geological layers. The northern part of Bishkek consists of Quaternary sediments made up of rubbly-bench gravel with clay-sand lenses and break stone bench gravel outcrop. In the north, the alluvial material becomes finer with gravels alternating to silt layers. In the northernmost part of the city, eolic sediments (loess) with thickness reaching tens of meters are found and are occasionally cut by recent alluvial deposits. This information, not reported in the available geological maps, is the result of a recent geological survey carried out by the authors. Fig. 2 shows the available large scale NS geological cross section of the study area modified after Baeva 1999.

3 Data set and analyses techniques

A temporary seismological network of 19 stations (Fig. 1), which operated in continuous mode, was installed in Bishkek from 16 August 2008 until the middle of November 2008 (Parolai et al., 2010). 15 stations used Earth Data Logger (EDL) 24-bit acquisition systems with 13 of them employing L4C-3D sensors having a 1 Hz corner frequency, one Güralp CMG-ESPC 60 and one IO-3D sensor with a corner frequency of 4.5 Hz. The other 4 stations used Reftek-130 digitizers with Le3D-5_s sensors. At all stations, data was recorded at 100 samples per second. The network recorded 56 events, including 50 crustal earthquakes with magnitudes between M_L 1.6 and M_W 6.6, occurring at distances between 35 and 1527 km from station BI08 (see Fig.1). The dataset includes 5 deep earthquakes with magnitudes between m_b 5.1 and M_W 6.8, which occurred in the Pamir-Hindu Kush region, and an event of magnitude M_W 7.3 which occurred at a distance of 5685 km in the Okhotsk Sea (Russia) at a depth of 492 km.

In order to estimate the SSRs, 50 seconds S-wave window length has been selected to include enough cycles of the low frequencies of interest. In fact, the resonance frequency of the studied area is expected to be around 0.1-0.2 Hz, considering the thickness of the sedimentary cover underlying the city. Strollo et al. (2008) has shown the reliability of the H/V results down to frequencies of about 0.1 Hz when using 1 Hz sensors combined with EDL digitizers. Also, the comparison of the self-noise of the instrument with the measured seismic noise (not shown here) confirmed the reliability of the results down to frequencies about 0.1 Hz. The selected windows have been de-trended for base line correction, 5% tapered at both ends using cosine windows and corrected for instrumental response. No filter is applied to the selected windows. The analysis was performed considering frequencies with a signal-to-noise-ratio greater than 3. The pre-event seismic noise window length, used to calculate the HVSRs at the 19 stations of the temporary seismological network, was selected to be equal to the S-wave window. The spectra have been smoothed using a 10% triangular window of the central frequency equal-spaced in logarithmic scale. Logarithmic averages of pre-event noise HVSRs and of earthquake SSRs have been estimated for the sites (Parolai et al. 2010). Along with the network of 19 stations, single station seismic noise was also recorded in the urban area covering the whole city at nearly 200 points. The seismic noise was recorded using an EDL 24-bit digital acquisition system, equipped with a 1 Hz Mark L4C-3D sensor at 100 samples per second. Seismic noise was recorded for about 30 minutes at each point. The data was split into 60 s window lengths, de-trended for base line corrections and tapered using a 5% cosine window at both ends.

4 Clustering Analysis

Cluster analysis has been used in many fields to organize objects into different groups based on the similarity of various selected parameters, in order to obtain a simple understanding of a given data set by grouping together those elements with similar attributes. Clustering is found under different names in different fields, for example unsupervised learning in the field of pattern recognition, numerical taxonomy in the biological sciences, typology in social sciences and partition in graph theory (Halkidi

et al. 2001). Nowadays, various algorithms for clustering analysis are used in different fields for specific kinds of data and application.

We used 15 out of the 19 temporary stations, since some stations had not recorded the events observed by the reference station BI04, from which SSRs and pre-event noise HVSRs were calculated (see table 1). The analyses were carried out over a frequency range from 0.1-2 Hz for SSRs using logarithmically equal-spaced bins. The lower bound of the selected frequency range was chosen to allow for each site's fundamental frequency to be covered, and to be within the range where the used instruments work properly. The fundamental frequencies are expected to be low, owing to the sedimentary cover thickness in the area. The upper bound of the considered frequency range is selected by acknowledging that at higher frequencies, the HVSR of the earthquake data for the reference site (BI04) is showing site amplification (see Parolai et al. 2010). Although this band restriction toward higher frequencies might limit the ability to discriminate the shallow geological structural variations, it will be shown later that the changes in the thin and shallow Quaternary layers can still be successfully captured by using the chosen frequency band. We think, and will explain in the Discussion and Conclusions section, that this might indicate that the shallow geological variations are linked to the deeper ones.

We used the K-means clustering algorithm (MacQueen, 1967) to subdivide the dataset into groups based on their similarity of spectral ratios. This means that there is no a priori geographical location information. The following are the main steps involved in the K-means algorithm: (1) the algorithm requires that the number of clusters (k) be defined a priori. k centroids are then defined randomly to represent each cluster at each frequency. (2) The dissimilarity between an object (in this case the value of the spectral ratios at each frequency) and the centroid of each cluster is calculated. The squared Euclidean distance parameter is used here to determine this difference. (3) Based on the shortest distance, an object is assigned to the cluster whose centroid is nearest to the object. (4) When all objects are assigned in this way, a new series of k centroids for each frequency are re-calculated based on the already formed clusters. The objects are again assigned to the new centroids based on the

shortest distance. The process continues iteratively until there is no further update of the centroids. The aim of this algorithm is to minimize the sum of the squared error over the whole dataset given as,

$$E = \sum_{i=1}^k \sum_{l=1}^n \sum_{j=1}^m \|x_{lj,i} - c_i\|^2 \quad (1)$$

where m is the number of objects (in our case, the number of frequencies examined) in a data set, n is the number of observations (or spectra) in each cluster, k the number of clusters, c_i is centroid of the cluster i , and $x_{lj,i}$ is the value of j object in cluster i of observation number l . The random selection of the centroid of each cluster affects the results of the clustering analysis such that it might not converge to the global minimum of the chosen function. Therefore, the analysis is repeated 10 times, using different randomly chosen starting centroids and the resulting solution that provides the minimum of the sum of the squared error is chosen as the best one.

We applied the K-means clustering algorithm separately to the SSRs and the seismic noise HVSRs of examined 15 stations of the temporary seismic network. The analyses were performed by varying the number of clusters (from 2 to 7). As a guide for the selection of the optimum number of clusters, we evaluated the Calinski & Harabasz index (CH Index) (Calinski and Harabasz 1974) and Silhouette index (Rousseeuw, 1987) for both the SSRs and seismic noise HVSRs. The CH Index is a variance ratio criterion that provides insight into the structure of points. In this method, the dispersion of a group of n -points is measured by the sum of the squared (SS) distances of the points from the centroid of their clusters. It is a measure of the between-groups SS over the within-groups SS, mathematically given as,

$$CH \text{ Index} = \frac{BGSS(k)}{(k-1)} \bigg/ \frac{WGSS(k)}{(n-k)} \quad (2)$$

where BGSS is the between-groups sum of the squared distance, WGSS is the within-group sum of the squared distance, k (2, 3, 4...) is the number of clusters and n (15) is the number of points (in this

case the spectral ratios of the 15 temporary stations). For the best clustering, the value of CH Index is maximized.

The mathematical form of the Silhouette index is given as,

$$S(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (3)$$

where $a(i)$ is the average dissimilarity (distance) of object i to all other objects in the same cluster, and $b(i)$ is the minimum of the average dissimilarity of object i to all other objects in the closest neighboring cluster. The value of $S(i)$ ranges from -1 to +1 with +1 representing a well-clustered sample, whereas -1 shows the sample is misfit. This is a graphical method for selecting the optimum number of clusters which shows the width of silhouette for each sample, the average silhouette width for each cluster and the overall average silhouette width for the complete data set. The overall average silhouette width is the average of the $S(i)$ for all objects in the whole dataset.

Fig.3 gives the CH Index (a, b) and Silhouette index (c, d) for various numbers of clusters for the SSRs (a, c) and seismic noise HVSRs (b, d). Both evaluation methods indicate that 2 is the optimal number of clusters. Cluster 1 is composed of the four stations in the north (BI01, BI14, BI15 and BI18) while cluster 2 is composed of the rest of the stations. However, in order to allow a smoother spatial variation of ground motion, we have preferred to choose 3 clusters that, anyway, provide the second best CH and Silhouette index results.

5 Clustering Results

The three selected clusters (for both SSRs and noise HVSRs) are depicted in Figures 4 and 5, respectively. They show a very similar spatial distribution, independent of whether seismic noise or earthquake data are used, although at each station, the SSR and HVSR might look quite different.

Fig. 4 (a) shows the resulting SSRs for the stations clustered in the north of the city. It reveals amplifications generally between 5-9 over the examined range of frequencies, with fairly constant

values after around 0.2 Hz. Fig. 4 (b) is the same for the stations in the center of the city, and shows a clear amplification of up to 5 at lower frequencies, in particular at 0.2 Hz and 0.4 Hz. For frequencies higher than 0.6Hz, the amplification remains between 2 and 3. Fig. 4 (c) presents the SSRs for the stations in the southern-most and south-east of the city, showing a lower amplification compared to the first two clusters, with a fairly constant amplification of between 1.5 and 3 over the examined frequency range, with occasional peaks at 0.2, 0.4 and 1.5Hz. Fig. 4 (d) shows the distribution of the different clusters superimposed on a geological map of the city (see Fig. 1 for details and the geological units).

Fig. 5 is the same as for Fig. 4, but for the seismic noise HVSRs. In HVSRs, the position and amplitude of peak plays an important role in the classification of the geological structure (Parolai et al. 2005; Picozzi et al. 2009). Fig. 5 (a) clusters the stations with higher amplitudes at around 0.2 Hz and a second prominent peak at around 0.9 Hz. Fig. 5 (b) clusters the stations with clear peaks at less than 0.2Hz, with a second peak, although not as prominent as in the first cluster. The amplitude of the peaks is also lower than in the first cluster, which possibly results from the deeper bedrock with a lower impedance contrast. Fig. 5 (c) shows the cluster of stations having a lower flat first peak at around 0.2 Hz, where the amplitude being lower than that for the first two clusters. The trend of lower resonance frequencies towards the southern part of the city is in agreement with the thickening of sedimentary cover to Paleozoic bedrock, as proposed by Bullen et al. (2001). Fig. 5 (d) shows the distribution of stations over a geological map, clustered together based on the noise HVSRs (see Fig. 1 for details and the geological units).

It is interesting that sites clustered together by similar SSRs are also clustered based on their similarity of seismic noise HVSRs. This means that, irrespective of the difference between the SSRs and seismic noise HVSRs for a station, the variability in the noise HVSRs is consistent with that of the SSRs. This analogy provides the opportunity for SSRs to be extended to SSNPs where earthquake data is not available. In general, stations having higher amplitude values of SSRs in the northern part of the city also show higher peaks in the seismic noise HVSRs, while stations in the southern part that have a

lower amplitude in the SSRs also show lower amplitudes in the seismic noise HVSRs. Considering the seismic noise HVSRs clustering, only station BI16 shows anomalous behavior (highlighted in a different color in Fig. 5 a, d) that agrees better with cluster 3 than with that from the nearby stations in cluster 1 because of the low HVSRs peak at lower frequencies. A possible reason for this deviation is that the station at this site was equipped with an EDL 24-bit acquisition system with 4.5 Hz sensor. According to Strollo et al. (2008), the HVSR of the noise peak at lower frequencies obtained by such a sensor is always biased due to the strong filtering effect of the geophone, leading to lower spectral ratio amplitudes at frequencies less than 0.5 Hz. Hence, this station is not considered in further analysis.

6 Correlation Analysis

Since sites having similar SSRs also share similar noise HVSRs, we can take advantage of the SSNPs to improve the spatial resolution of the microzonation (intended here as assessing the site response variation over short distances) in the city. We used the following correlation scheme to relate the seismic noise HVSRs of 177 SSNPs out of the 200 (this was because some time series were discarded, as a preliminary analysis of the ratios suggested serious installation problems) with the remaining 14 seismic noise HVSRs (for which SSRs are available).

First of all, the Pearson cross-correlation coefficients (Davis 1986) were calculated between 177 SSNPs and the 14 stations using seismic noise HVSRs. The Pearson correlation coefficient is the covariance of the two variables divided by the product of their standard deviations, given as,

$$\rho_{jk} = \frac{\left(\frac{\sum_{i=1}^n (x_{ij} - \bar{x}_j) \cdot (x_{ik} - \bar{x}_k)}{(n-1)} \right)}{\left(\sqrt{\frac{\sum_{i=1}^n (x_{ij} - \bar{x}_j)^2}{(n-1)}} \right) \cdot \left(\sqrt{\frac{\sum_{i=1}^n (x_{ik} - \bar{x}_k)^2}{(n-1)}} \right)} \quad (4)$$

where ρ represents the correlation coefficient, j is an index that goes from 1 to the number of SSNPs (177), k is an index that goes from 1 to the number of temporary seismic stations (14), i is an index of the frequency (0.1-2Hz), x is the seismic noise HVSR value at frequency i , and $\bar{x}$ is the mean value of x . The value of ρ ranges from -1 to +1.

To quantify the degree of similarity, the Pearson correlation coefficients are multiplied by the degree of fit, which is given by,

$$f_{jk} = \frac{1}{\sqrt{\frac{\sum_{i=1}^n (x_{ij} - x_{ik})^2}{n}}} \quad (5)$$

This is basically the standard error between the seismic noise HVSRs of the SSNPs and the permanent stations. Equation (5) is used to take into account the difference in amplitude between the noise HVSRs of the SSNPs and the temporary seismic stations. The greater the difference, the lower the value of equation (5), while multiplying equation (4) by equation (5) will suppress the correlation coefficient and vice versa.

$$CC = \rho \cdot f \quad (6)$$

The Matrix CC in equation (6) has dimensions of [177 x 14]. In order to select only those SSNPs having good similarity of seismic noise HVSRs with the stations of the temporary network in terms of CC coefficient, a lower threshold of 0.5 for CC is selected. Changing the value of this threshold affects only the total number of SSNPs to be considered for further analysis (but not their assignment to a certain cluster). This leads to about 81% of the 177 stations to be robustly classified. Furthermore, a SSNP is assigned to a cluster based on the condition that the maximum CC coefficient between the SSNP and the temporary stations of the cluster is greater than 15% of the maximum CC coefficient when it is compared to the stations of the second best cluster. We ran the analysis with different threshold levels, and we noticed that while it affects the number of SSNPs assigned to a particular

cluster/group, mainly in the transition zone, the overall distributions of SSNPs remain the same. About 79% of the classified SSNPs (i.e., 79% of the classified 81%) are finally assigned through this process to different clusters. Noise measurement points that could not satisfy the second criterion were assigned to the group based on the nearest distance and geology, provided their CC coefficients were within 15% of the maximum. This leads to a total of 143 out of 177 noise measurement points being assigned to the three clusters in this way. Fig. 6 shows examples of higher to lower CC coefficients values for the seismic noise HVSR of station BI18 compared to the SSNPs in the first cluster. Note that larger the value of the CC, the greater is the resemblance to the NHVSRs. Fig. 1 shows the distribution of SSNPs and the permanent stations over the geological map of the area.

The site response associated with each SSNP is the logarithmic average of the SSRs of each cluster to which the SSNP was assigned after the correlation analysis. Fig. 7 shows the SSRs for each cluster. Different dashed colored lines represent different clusters, while the bold continuous line in each cluster represents the logarithmic average of that cluster. The grey area shows the frequency range where analyses are not considered because of the spectra ratios being affected by the site response at the reference site (see Parolai et al. 2010).

In order to better depict the spatial variation of amplification in the study area, the site response amplitude in terms of SSRs has been interpolated using natural neighbors spatial interpolation. For each temporary station, we decided to retain its own logarithmic average site response as estimated from the earthquake analysis (Parolai et al. 2010). Fig. 8 shows the site response estimated for different frequencies over the study area. The figure shows that the largest amplification mainly affects the northern part of the urban area of Bishkek. In particular, the whole frequency band from 0.2 to 2 Hz shows amplification values larger than 5, with maximum values reaching nearly 8 at around 0.4 Hz. The sudden increase of the amplification values corresponding to the geological contact between the different Quaternary materials is striking. The effect of the different Quaternary sediments, responsible for the change in the site response amplitudes between the north and south of the city, can be clearly seen around latitude 42.85 in Fig 1.

7 Discussions and Conclusions

In this study, the spatial resolution of ground motion variability in terms of SSRs has been improved upon in Bishkek, Kyrgyzstan, using earthquake and seismic noise data. The appropriateness of seismic noise as a proxy for extrapolating the earthquake site response is evaluated with the help of cluster analysis. Using the K-means algorithm, clusters of sites that display a similar response have been identified based on SSRs and seismic noise HVSRs. This indicates that, irrespective of the shape of the SSRs and of the seismic noise HVSRs, sites showing similar SSRs also share similar noise HVSRs. Using this analogy, the site response in terms of SSRs has been linked with the SSNPs recorded in the city for which earthquake recordings were not available. This is accomplished by using a correlation analysis of the noise HVSRs between the 14 stations of a temporary network and 177 SSNPs.

The three clusters identified in the study area, which cover the northern part of the city, a central transition zone and the southern part, follow well the geological structure of the basin. The depth distribution of the Paleozoic bedrock in the urban area of Bishkek is such that iso-lines are aligned east-west (Bullen et al. 2001, their Fig. 4). The clustering patterns identified in this work following the same depth distribution, as the spectral ratios reflect the deeper geological structure.

Note that although the clustering analysis performed on the SSRs is, in this study, only considers frequencies up to 2Hz (due to the amplification at the employed reference station), this procedure may also be considered to be valid at higher frequencies since all stations are equally affected by the site response at the reference site. The obtained results allow an efficient microzonation with a higher spatial resolution using seismological parameters alone, and the proposed approach could be used in general as a tool wherever a “good” rock reference site is available. However, in terms of its application for seismic hazard and risk scenarios, in the case at hand, a broader band site response is necessary to correctly evaluate the absolute value of ground motion in the different parts of the city. Suitable correction factors will thus have to be taken into account when considering the effect of the site amplification at the reference site.

Acknowledgements

Shahid Ullah carried out this work within the Global Earthquake Model (GEM) regional project Earthquake Model Central Asia (EMCA) while a doctoral student at the Technical University Berlin. His study at TU Berlin is supported by a developmental project “Strengthening of existing earthquake engineering center” University of Engineering & Technology, Peshawar Pakistan. Dr. K. Fleming kindly revised our English. We are grateful to the editor in chief A. Ansal and the two anonymous reviewers for their constructive comments. Some of the figures presented in this manuscript have been generated using the GMT software package (Wessel and Smith 1991).

References

- Baeva, N.B. (1999) Geological profile, Government Agency of geology and mineralogy of Kyrgyz Republik. Kyrgyz-Hydro-geological expedition.
- Bard PY (1999) Microtremor measurements: a tool for site effect estimation? In Proceedings of the Second International Symposium on the Effects of Surface Geology on Seismic Motion, Yokohama, Japan, December 1998, Vol. I–III, 1251–127
- Bard PY (2004): The SESAME project (2004) An overview and main results, In Proceedings of the 13th World Conference on Earthquake Engineering, Vancouver. Paper: 2207.
- Bindi D, Mayfield M, Parolai S, Tyagunov S, Begaliev UT, Abdrakhmatov K, Moldobekov B, Zschau J (2011) Towards an improved seismic risk scenario for Bishkek, Kyrgyz Republic. *Soil Dyn Earthquake Eng* 31:521-525
- Bindi D, Parolai S, Spallarossa D, Cattaneo M (2000) Site effects by H/V ratio: comparison of two different procedures. *J Earthquake Eng* 4:97-113
- Boxberger T, Picozzi M, Parolai S (2011) Shallow geology characterization using Raleigh and Love wave dispersion curves derived from seismic noise array measurements. *J Appl Geophys* 75:345-354
- Bragato PL, Laurenzano G, Barnaba C (2007) Automatic Zonation of Urban Areas Based on the Similarity of H/V Spectral Ratios. *Bull Seismol Soc Am* 97:1404-1412
- Bullen ME, Burbank DW, Garver JI, Abdrakhmatov KY (2001) Late Cenozoic tectonic evolution of the northwestern Tien Shan: New age estimates for the initiation of mountain building. *Geol Soc Am Bull* 113:1544-1559
- Calinski T, Harabasz J (1974) A dendrite method for cluster analysis. *Communication in Statistics-Theory and Methods*, 3:1-27

D'amico V, Picozzi M, Albarello D, Naso G , Tropenscovino S (2004) Quick estimates of soft sediment thicknesses from ambient noise horizontal to vertical spectral ratios: a case study in southern Italy. *J Earthquake Eng* 8:895-90

Davis JC (1986) *Statistics and Data Analysis in Geology*-2nd edition. John Wiley and Sons, Toronto

Erdik M, Rashidov T, Safak E, Turdukulov A (2005) Assessment of seismic risk in Tashkent, Uzbekistan and Bishkek, Kyrgyz Republic. *Soil Dyn Earthquake Eng* 25:473-486

Field EH, Jacob KH (1995) A comparison and test of various site response estimation techniques, including three that are not reference-site dependent. *Bull Seismol Soc Am* 85:1127-1143

Halkidi M, Batistakis Y, Vazirgiannis M (2001) On Clustering Validation Techniques. *J Intell Inf Syst* 17:107-145

Horike M, Zhao B, Kawase H (2001) Comparison of site response characteristics inferred from microtremor and earthquake shear waves. *Bull Seismol Soc Am* 91:1526-1536

Januzakov KJ, Omuraliev M, Omuralieva A, Ilyasov BI, Grebennikova VV (2003) Strong earthquakes of the Tien Shan (within the Kyrgyzstan territory and adjacent regions of the countries of Central Asia), Bishkek: Ilim, 216 pp, ISBN 5-8355-1335-6.

Kalmetieva ZA, Mikolaichuk AV, Moldobekov BD, Meleshko AV, Jantaev MM, Zubovich AV, (2009) *Atlas of Earthquakes in Kyrgyzstan*, Central Asian Institute for Applied Geosciences, Bishkek, ISBN 978-9967-25-829-7

Konno K, Ohmachi T (1998) Ground-Motion Characteristics Estimated from Spectral Ratio between Horizontal and Vertical Components of Microtremor. *Bull Seismol Soc Am* 88: 228-241

Lermo J, Chavez-Garcia FJ (1994) Are microtremors useful in site response evaluation?. *Bull Seismol Soc Am* 84:1350-1364

MacQueen JB (1967) Some Methods for classification and Analysis of Multivariate Observations, *Proceeding of 5th Berkley Symposium on Mathematical Statistics and Probability*, Berkley, University of California Press 1:281-297

Parolai S, Bormann P, Milkereit C (2001) Assessment of the Natural Frequency of the Sedimentary Cover in the Cologne Area (Germany) using Noise Measurements. *J Earthq Eng* 5:541-564

Parolai S, Orunbaev S, Bindi D, Strollo A, Usupaev S, Picozzi M, Di Giacomo D, Augliera P, D'Alema E, Milkereit C, Moldobekov B, Zschau J (2010) Site Effects Assessment in Bishkek (Kyrgyzstan) Using Earthquake and Noise Recording Data. *Bull Seismol Soc Am* 100:3068-3082

Parolai S, Picozzi M, Richwalski SM, Milkereit C (2005) Joint inversion of phase velocity dispersion and H/V ratio curves from seismic noise recordings using a genetic algorithm, considering higher modes. *Geophys Res Lett* 32, 1, L01303.

- Parolai S, Richwalski S, Milkereit C, Bormann P (2004) Assessment of the stability of H/V spectral ratios from ambient noise and comparison with earthquake data in the Cologne area (Germany). *Tectonophysics* 390:57-73
- Picozzi M, Strollo A, Parolai S, Durukal E, Özel O, Karabulut S, Zschau J, Erdik M (2009) Site characterization by seismic noise in Istanbul, Turkey. *Soil Dyn Earthquake Eng* 29:469-482
- Pilz M, Parolai S, Leyton F, Campos J, Zschau J (2009) A Comparison of site response techniques using earthquake data and ambient seismic noise analysis in the large urban areas of Santiago de Chile. *Geophys J Int* 178:713-728
- Rodriguez VH, Midorikawa S (2002) Applicability of the H/V spectral ratio of microtremors in assessing site effects on seismic motion. *Earthquake Eng Struct Dyn* 31(2):261–279
- Rousseeuw PJ (1987) Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *J Comput Appl Math* 20:53-65
- Strollo A, Parolai S, Bindi D, Chiuzzi L, Pagliuca R, Mucciarelli M, Zschau J (2012) Microzonation of Potenza (Southern Italy) in terms of spectral intensity ratio using joint analysis of earthquakes and ambient noise. *Bull Earthquake Eng*. doi:10.1007/s10518-011-9256-4
- Strollo A, Parolai S, Jäckel KH, Marzorati S, Bindi D (2008) Suitability of Short-Period Sensors for Retrieving Reliable H/V Peaks for Frequencies Less Than 1 Hz. *Bull Seismol Soc Am* 98:671-681
- Tselentis GA, Paraskevopoulos P (2011) On the Use of Kohonen Neural Networks for Site Effects Assessments by means of H/V Weak-Motion Spectral Ratio: Application in Rio-Antirrio (Greece). *Bull Seismol Soc Am* 101:579-595
- Wessel P, Smith WHF (1991) Free software helps map and display data. *Eos Trans AGU* 72:441-461

Table 1

Agency	Station Name	Latitude (N)	Longitude(E)	Digitizer	Sensor
GFZ	BI01	42.9109	74.5869	EDL	MARK-L4C-3D
GFZ	BI02	42.8665	74.6392	EDL	MARK-L4C-3D
GFZ	BI03	42.8133	74.6564	EDL	MARK-L4C-3D
GFZ	BI04	42.6712	74.6438	EDL	MARK-L4C-3D
GFZ	BI06	42.8040	74.6036	EDL	MARK-L4C-3D
GFZ	BI07	42.8087	74.7074	EDL	MARK-L4C-3D
INGV	BI08	42.8543	74.5332	Reftek-130	LENNARTZ-LE3D-5s
INGV	BI09	42.8350	74.5238	Reftek-130	LENNARTZ-LE3D-5s
GFZ	BI10	42.7836	74.5182	EDL	MARK-L4C-3D
INGV	BI11	42.8722	74.5868	Reftek-130	LENNARTZ-LE3D-5s
GFZ	BI13	42.8400	74.6289	EDL	MARK-L4C-3D
GFZ	BI14	42.9282	74.6317	EDL	MARK-L4C-3D
GFZ	BI15	42.8901	74.7341	EDL	MARK-L4C-3D
GFZ	BI16	42.9368	74.6061	EDL	IO-3D 4.5Hz
GFZ	BI18	42.8683	74.4505	EDL	MARK-L4C-3D
GFZ	BI19	42.8537	74.6858	EDL	MARK-L4C-3D

Fig. 1

Fig. 2

Fig. 3

Fig. 4

Fig. 5

Fig. 6

Fig. 7

Fig. 8

Figure Captions

Fig. 1 The location of the Single station Seismic Noise measurement Points (SSNPs) and the temporary network used in this study, superimposed over the geological map of Bishkek, Kyrgyzstan. The bigger size shapes indicate the permanent stations, while the smaller sizes represent the SSNPs. Different colors represent the different clusters or groups. (The detail of method is presented in the clustering and correlation section)

Fig. 2 Geological cross-section of the Bishkek basin redrawn after Baeva (1999).

Fig. 3 Calinski & Harabasz (CH) and Silhouette Indexes vs the number of clusters. (a) and (b) shows the CH Index for the SSRs and seismic noise HVSRs clustering, respectively. (c) and (d) shows the Silhouette Index for the SSRs and seismic noise HVSRs clustering, respectively

Fig. 4 Results of the clustering analysis using the K-means algorithm on the SSRs. Fig. (a), (b), and (c) show the resulting spectra subdivided into different clusters. Fig. (d) shows the spatial distribution of the stations and their respective clusters over a geological map of the area (see Fig. 1)

Fig. 5 Results of the clustering analysis using the K-means algorithm on the seismic noise HVSRs of the temporary network stations. Fig. (a), (b), and (c) show the resulting spectra subdivided into different clusters. Fig. (d) Shows the spatial distribution of the stations and their respective clusters over a geological map of the area. The anomalous behavior of BI16 is discussed in the text

Fig. 6 Comparison of seismic noise HVSR for SSNPs with that from a temporary network station (BI18). Red line shows the noise HVSR of station BI18, magenta color shows HVSR of SSNP #166 having a CC coefficient of 3.7, the blue color is for SSNP 11 having a medium CC coefficient of 2.4 and the black line is for SSNP 164 having a lower CC coefficient of 1.05. The location of the SSNPs are shown in Fig. 1

Fig. 7 Logarithm average of SSRs of each temporary station. Different colors indicate the different clusters, while bold colors in each cluster show the logarithmic average of that cluster. The grey area represents the frequency range where analyses are not considered

Fig. 8 Spatial distribution of SSRs at selected frequencies (0.2, 0.4, 1 and 2Hz) using Natural Neighbours Interpolation

Table 1: Details of the temporary network stations and instruments.