Bulletin of the Seismological Society of America, Vol. 101, No. 2, pp. 579–595, April 2011, doi: 10.1785/0120090355

On the Use of Kohonen Neural Networks for Site Effects Assessment by Means of H/V Weak-Motion Spectral Ratio: Application in Rio-Antirrio (Greece)

by G-Akis Tselentis and Paraskevas Paraskevopoulos

Abstract The investigated area, located around the Rio-Antirrio Strait, Central Greece, has been the target of a seismic microzonation campaign. Seventy seismic stations have been deployed for a period of 4 months, recording in continuous mode. Despite the high level of urban noise, we compiled a data set of 95 earthquakes recorded at most of the 70 sites. By employing the attributes of self-organizing maps (SOMs), a quality-control and signal-improving method is proposed. A SOM (Kohonen, 1997) is a type of unsupervised neural network. The main property of SOMs utilized is that while the competitive learning algorithm on whom this method is based maps the input data on an *n*-dimensional grid of neurons, the topological relations (proximity of patterns in input data) are preserved in the output space. SOM is applied to the horizontal-to-vertical spectral ratios (HVSR) of every weak event analyzed for each station separately and allows a better evaluation of the stability of the HVSR.

Introduction

The investigated region, surrounding the Rio-Antirrio Strait, is one of the most active seismic areas in Greece because it is under continuous tectonic deformation (Tselentis and Makropoulos, 1986; Tselentis *et al.*, 2007, 2010). The recent completion of a 3-km cable bridge spanning the Strait (the longest in the world of its kind) has resulted in an explosion in commercial activities and population settlement. Local variations of earthquake ground motion are expected in the region because the subsoil structure is characterized by important lithological heterogeneity and widespread soft-soil conditions. In this context, the evaluation of the local soil conditions and the study of the site effects are crucial steps towards a better seismic risk assessment for the area.

It is a well-documented phenomenon that earthquake ground motion can be amplified by local site conditions (e.g., Milne, 1898; Kanai, 1951; Borcherdt, 1970; Aki, 1988; Bard, 1999). When seismic waves emerge from the base of more competent rocks into the less-competent uppermost soils, ground motion can change dramatically. The impedance contrast near the surface affects the frequency-amplitude content of earthquake ground motion. It can also change the duration, one of the most important parameters controlling the damage of earthquake ground motion. Examples from Loma Prieta (1989), Guerrero Michoacan (1985), Northridge (1994), Kobe (1995), Armenia (1999), Kocaeli (1999), Athens (1999), Bhuj (2001), and Bam (2003) earthquakes have been extensively cited to illustrate the role of surface geology. This underlines how important it is to account for site effects in the design of new constructions, in the retrofitting of existing structures and in land-use planning.

Many investigators have evaluated site response functions from moderate to weak motions of earthquakes (e.g., Jongmans and Campillo, 1993; Carver and Hartzell, 1996; Hartzell *et al.*, 1996; Steidl *et al.*, 1996; Lachet *et al.*, 1996; Toshinawa *et al.*, 1997; Mucciarelli *et al.*, 2003; Zaslavsky *et al.*, 2005; Di Giulio *et al.*, 2005; Mandal *et al.*, 2005; Improta *et al.*, 2005; Tselentis *et al.*, 2010), exhibiting a rather good correlation with surface geology, and in many cases these studies are able to predict site resonance frequencies, whereas there is a lower reliability on the derived amplification factors (Bard, 1999; Haghshenas *et al.*, 2008). The task of the present experiment is to understand the seismic response of the Rio-Antirrio region and identify parts exhibiting similar ground motion behavior as part of a microzonation investigation. Due to the high urban noise, we examined the use of self-organizing maps (SOMs) to evaluate the stability of the calculated horizontal-to-vertical spectral ratios (HVSR).

Geology

The study area covers the Rio-Antirrio strait (Fig. 1), which is located in the northern part of the Peloponnesus. The Corinth and Patras rifts are linked by transfer faults in the Rio-Antirrio Strait (Doutsos and Poulimenos, 1992). After Middle Miocene, the whole northwestern Peloponnesus area was uplifted and extended. During this extension,

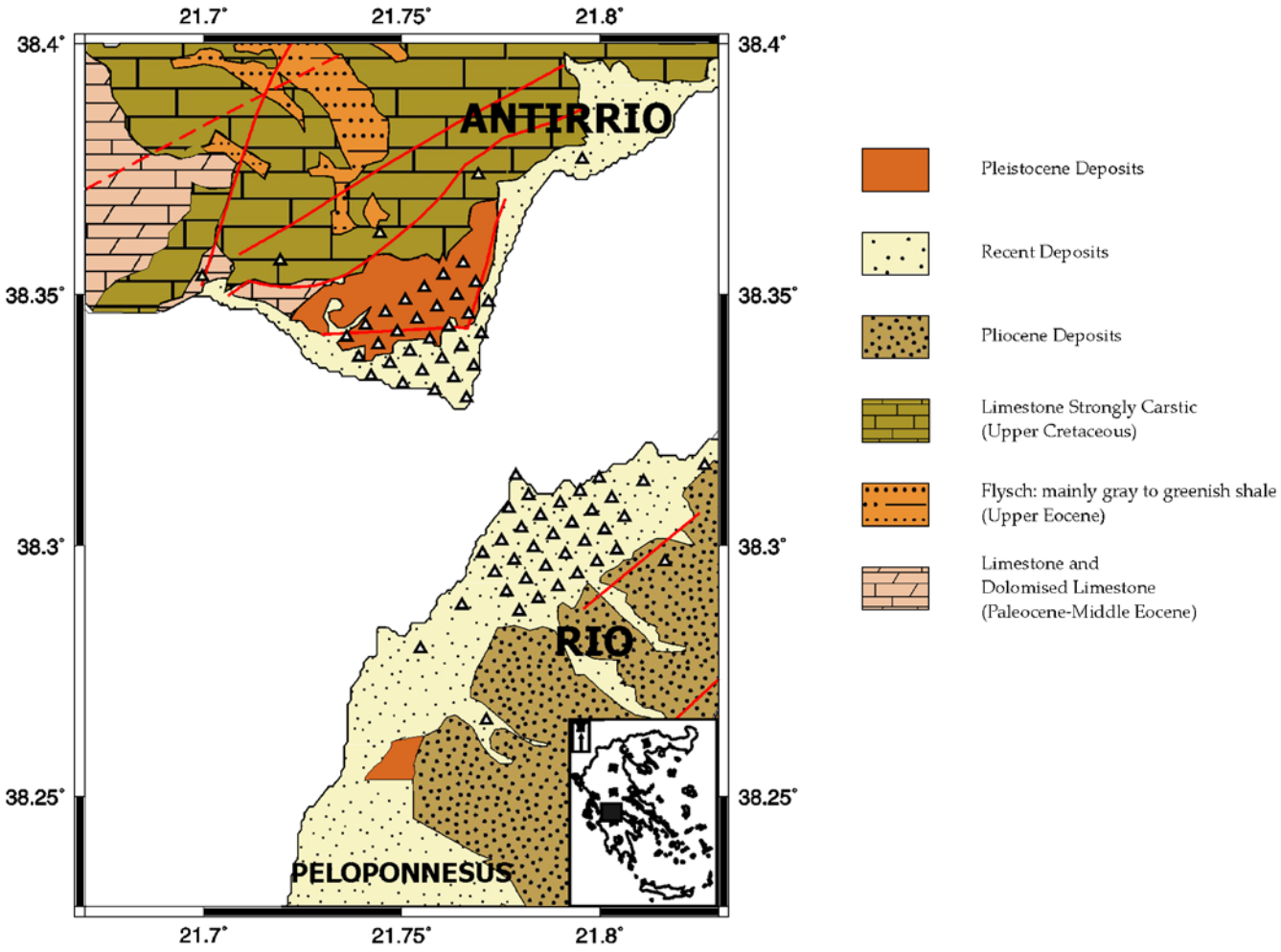

Figure 1. Geological map of the region with the station positions indicated. The color version of this figure is available only in the electronic edition.

three asymmetric grabens were formed: the Corinth, Rio-Antirrio, and Patras grabens.

The asymmetry of the grabens is largely induced by north-dipping master faults, which trend parallel to the coastline of the north Peloponnesus (Doutsos and Poulimenos, 1992). In the Quaternary, rifting propagated westward until it reached the Rio-Antirrio graben. The Rio graben, formed by movement of preexisting northeast–southwest trending faults, reactivated in the Pliocene (Doutsos *et al.*, 1985). Changes in predominant stress directions at this time led to the Rio graben acting as a transfer zone between the existing Patras and Corinth grabens (Tselentis and Makropoulos, 1986; Doutsos *et al.*, 1988).

The three-dimensional (3D) tomographic velocity inversion in this area has been extensively studied by Tselentis *et al.* (2007). The primary structural features are the following. Low velocity top geological layers corresponding to Quaternary and Neogene sedimentary formations. The thickness of these formations in the target area is approximately 500–600 m. The Alpine Basement is characterized by limestone outcroping in the northwestern part of the study area, where the limestone Klokova Mountain sits, and also in the southern part (Panahaikon Mountain). In the northern and northeastern part of the study, the flysch formations of the Pindos are present.

Data

The acquisition campaign was carried out between December 2003 and May 2004. A total of 70 stations were used in the whole network. Sixty of the stations were deployed in the Rio-Antirrio area all over an approximately 500-m grid (Fig. 1). In addition, ten peripheral stations were deployed for the better location control of the events occurring at the outskirts of the investigated area.

The instrumentation consisted of Earthdata PR24 dataloggers connected to LandTech S100 three-component sensors with 1 Hz natural frequency and flat response between 1 and ~70 Hz.

A total of 220 events were recorded, and 95 were selected (Fig. 2, Table 1) for the present investigation on the basis of the best signal-to-noise ratio (SNR) and number of

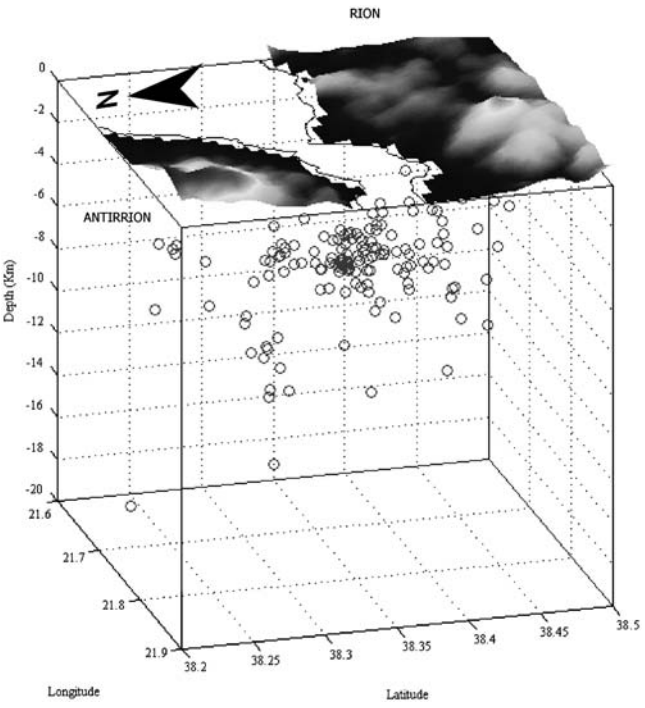

Figure 2. Hypocenters of used events.

stations that have been recorded. In weak-motion studies of this kind, it is important to use as many data as possible, but it is well known that instabilities might be induced when we involve in the analysis low SNR events that have been recorded only in a subset of the stations.

The distribution of the back-azimuths spans a broad circular sector covering almost all directions (Fig. 3). For example, Figure 3a illustrates the coverage as a function of the event's magnitude for the selected station 08. Figure 3b depicts the azimuthal coverage provided by the recorded events as a function of their epicentral distances, which varies between 0.5 and 10 km. Examples of recorded seismograms at 8 different locations for event 1 in Table 1 (M ~1.70) are shown in Fig. 4.

Methodology

Spectral Analyses

Because there is no downhole array in the investigated area, we have decided to use single-station estimates of the HVSR. This approach is quite useful in estimating the site fundamental frequency (Nogoshi and Igarashi, 1971; Nakamura, 1989; Lermo and Chavez Garcia, 1993; Mandal *et al.*, 2005; Tselentis et al., 2010).

The method consists of calculating the spectral ratio (HVSR) between the combined smoothed horizontal components and the smoothed vertical component. Bonilla *et al.* (2002) noted that a significant site response can be associated with the vertical component resulting from *S*-to-*P*-wave conversion at the weathered granite boundary, and that this violates the basic assumption behind the HVSR method. Nevertheless, this methodology has been widely used to estimate site response of many areas, although there is not unanimous agreement yet among seismologists regarding the accuracy of the methodology.

Figure 5a shows the normalized, vertical-component seismograms as recorded at 15 selected stations for event 72 in Table 1. For the analysis, a time window of 10.24 s (consisting in this case of 1024 samples) was used and started 0.5 s before the *S* arrival for all three components (Fig. 5b). The *S* arrivals were manually picked, and even picks with lesser accuracy were utilized as only the approximate onset of *S* waves was needed; also the time length was chosen in order to contain most of the high amplitude direct *S*wave energy. Bonilla *et al.* (1997) pointed out that using longer times results in better spectral resolution at the cost of including in the spectra scattered and reflected energy as well as surface waves. Field and Jacob (1995) found no statistical variations in site response computed with spectra of different time-window lengths. On the other hand, Castro *et al.* (1997) suggested that *S* waves can be contaminated by surface waves at larger epicentral distances, which recommends the use of variable time windows for the estimation of HVSR using *S* waves.

The selected part of the time traces was detrended, a 5% tapering was applied, and for each component the spectrum was calculated. Each spectrum was smoothed using a 0.4-Hz length Parzen window. In order to obtain a site response for each site, a resultant horizontal component was calculated following Mandal *et al.* (2005) as the average of the spectra of the two horizontal components and divided by the spectrum of the vertical component following equation (1):

$$Rhv = \frac{\left[\left(\sqrt{S_{\mathrm{NS}}(f)^2 + S_{\mathrm{EW}}(f)^2}\right)/2\right]}{S_v(f)}. \qquad (1)$$

The final step is gathering all the spectra for all the events in each site and the calculation of the average HVSR (Fig. 6). In order to have a quality control and calculate a weighted average of the HVSRs, we decided to use the SOMs methodology (Kohonen, 1997).

Kohonen Neural Networks

SOMs, or Kohonen neural networks, is a type of unsupervised artificial neural network (Kohonen, 1997). The structure of a SOM neural network is composed of two layers: an input layer and an output one (or Kohonen layer). The output layer consists of nodes known as neurons. The aim of this method is to transform the input vectors with arbitrary dimensions into generally 2D simplified maps (Klose, 2006).

One basic distinction between "classical" neural networks and SOMs is their ability to utilize an unsupervised learning method, known as competitive learning. As a result

Table 1
Parameters of the Microearthquakes That Were Used in the Present Investigation

Year	Month	Day	Hours	Minutes	Seconds	Latitude	Longtitude	Depth	Magnitude
2003	Dec	24	01	01	10.380	38.2985	21.7648	9.56	1.70
2003	Dec	25	10	30	48.030	38.3072	21.7597	9.23	1.35
2003	Dec	28	02	12	10.930	38.3538	21.7302	6.45	1.35
2004	Jan	09	15	05	21.740	38.3355	21.8243	8.29	1.07
2004	Jan	24	02	32	17.880	38.3118	21.7500	10.90	2.27
2004	Jan	24	08	20	40.730	38.3032	21.7385	10.67	1.76
2004	Jan	24	08	20	58.180	38.2993	21.7215	10.80	1.67
2004	Jan	24	09	41	50.980	38.3057	21.7460	11.97	1.62
2004	Jan	24	18	43	58.410	38.3063	21.7400	12.52	1.55
2004	Jan	24	21	04	34.640	38.2932	21.7277	9.19	1.88
2004	Jan	26	14	32	12.550	38.3122	21.7137	10.76	1.74
2004	Jan	27	08	35	3.250	38.3685	21.7935	4.85	1.38
2004	Jan	27	10	08	59.150	38.3130	21.7783	4.44	1.67
2004	Jan	30	21	48	40.500	38.3528	21.7643	4.55	2.19
2004	Jan	31	01	14	27.440	38.3180	21.7482	12.08	2.26
2004	Feb	01	22	52	21.460	38.3580	21.7380	5.87	1.46
2004	Feb	02	21	23	9.540	38.3780	21.7515	5.64	1.51
2004	Feb	03	22	54	2.060	38.3512	21.7352	6.51	1.44
2004	Feb	13	06	20	54.260	38.3953	21.7818	3.29	1.81
2004	Feb	13	18	41	53.540	38.3723	21.7597	5.10	1.49
2004	Feb	13	18	48	41.020	38.3717	21.7645	5.29	1.27
2004	Feb	14	21	04	51.950	38.3635	21.7623	4.78	1.27
2004	Feb	14	21	05	2.580	38.3622	21.7598	4.67	1.44
2004	Feb	19	01	10	22.280	38.3582	21.7943	4.61	1.96
2004	Feb	20	22	43	30.770	38.3552	21.7575	5.17	2.41
2004	Feb	20	22	59	19.130	38.3548	21.7475	4.69	0.97
2004	Feb	23	10	24	37.460	38.3512	21.7525	5.00	1.64
2004	Feb	26	21	41	1.910	38.3325	21.7600	5.27	1.51
2004	Feb	28	23	30	48.340	38.3337	21.7602	5.82	1.74
2004	Mar	01	02	52	11.010	38.3067	21.7787	4.53	1.41
2004	Mar	01	23	57	1.720	38.2612	21.7542	7.54	2.04
2004	Mar	02	02	50	15.420	38.3423	21.8080	3.37	1.41
2004	Mar	03	01	55	4.050	38.3048	21.7752	4.94	1.31
2004	Mar	03	20	37	59.170	38.3052	21.7725	5.00	1.93
2004	Mar	03	20	52	36.490	38.3072	21.7765	4.80	1.58
2004	Mar	03	22	21	32.740	38.3040	21.7727	5.07	1.47
2004	Mar	05	01	12	50.590	38.3003	21.7703	4.84	1.75
2004	Mar	05	02	19	53.930	38.2982	21.7722	5.24	1.72
2004	Mar	06	15	58	49.690	38.3057	21.7740	4.41	1.54
2004	Mar	09	02	03	51.050	38.3005	21.8432	0.30	1.26
2004	Mar	09	03	13	31.350	38.3022	21.8402	1.65	1.68
2004	Mar	10	09	22	29.760	38.3707	21.7957	4.86	1.88
2004	Mar	11	21	40	49.970	38.3270	21.8247	3.21	2.05
2004	Mar	12	01	02	13.800	38.3352	21.7663	6.02	1.62
2004	Mar	14	09	44	21.280	38.3620	21.7475	4.87	1.24
2004	Mar	14	10	16	41.150	38.3815	21.7500	3.54	0.96
2004	Mar	14	22	05	3.420	38.2933	21.7980	2.95	1.09
2004	Mar	15	23	05	23.630	38.3560	21.8193	4.33	1.53
2004	Mar	16	22	25	12.820	38.3065	21.7757	5.69	1.04
2004	Mar	16	22	25	12.940	38.3160	21.7857	5.15	0.85
2004	Mar	18	06	55	43.880	38.3127	21.7793	5.31	1.17
2004	Mar	18	17	19	22.270	38.3517	21.8233	4.32	2.33
2004	Mar	18	22	20	52.550	38.3478	21.8228	3.40	1.53
2004	Mar	18	22	35	55.190	38.3523	21.8280	4.72	1.62
2004	Mar	19	16	10	1.240	38.3492	21.8183	3.81	2.10
2004	Mar	19	22	35	31.150	38.3523	21.7438	6.39	2.49
2004	Mar	23	22	10	57.590	38.3295	21.8387	3.80	2.58
2004	Apr	05	12	04	58.570	38.3523	21.8190	4.95	1.68
2004	Apr	06	10	26	5.540	38.3618	21.7377	6.15	2.28
2004	Apr	08	02	25	4.360	38.2838	21.8092	4.24	1.54
2004	Apr	08	02	25	4.330	38.2892	21.8003	5.18	1.71

(*continued*)

Table 1 (*Continued*)

Year	Month	Day	Hours	Minutes	Seconds	Latitude	Longtitude	Depth	Magnitude
2004	Apr	08	16	15	49.410	38.3442	21.7778	5.70	2.56
2004	Apr	08	16	15	49.560	38.3390	21.7637	4.52	2.55
2004	Apr	09	14	05	24.420	38.3622	21.7862	6.45	2.53
2004	Apr	09	14	05	24.500	38.3540	21.7857	6.36	2.46
2004	Apr	11	21	24	23.130	38.3697	21.7410	6.43	1.81
2004	Apr	18	03	14	34.820	38.3350	21.7695	5.91	1.69
2004	Apr	21	15	34	9.760	38.3790	21.7993	4.68	1.73
2004	Apr	21	17	27	8.650	38.2805	21.7953	5.56	1.68
2004	Apr	22	18	25	14.230	38.3675	21.7462	6.03	1.30
2004	Apr	22	23	16	8.040	38.3423	21.7637	7.02	2.08
2004	Apr	28	07	26	58.750	38.3542	21.7472	6.69	4.19
2004	Apr	28	07	38	43.700	38.3562	21.7425	6.40	2.94
2004	Apr	28	07	45	19.130	38.3583	21.7385	6.38	1.97
2004	Apr	28	07	46	30.070	38.3590	21.7342	6.61	1.87
2004	Apr	28	08	04	38.780	38.3725	21.7417	6.03	1.38
2004	Apr	28	08	14	50.120	38.3660	21.7357	5.49	1.54
2004	Apr	28	08	45	38.120	38.3615	21.7437	6.32	1.88
2004	Apr	28	08	41	47.380	38.3597	21.7340	6.85	1.73
2004	Apr	28	08	50	46.410	38.3602	21.7345	6.71	1.97
2004	Apr	28	09	38	21.270	38.3687	21.7363	7.07	1.79
2004	Apr	28	10	23	47.400	38.3655	21.7310	6.97	1.63
2004	Apr	28	11	03	59.300	38.3648	21.7290	6.77	1.84
2004	Apr	28	17	13	30.470	38.3668	21.7240	6.92	1.51
2004	Apr	28	17	30	33.530	38.3607	21.7390	7.90	2.02
2004	Apr	30	02	43	0.710	38.3602	21.7423	6.53	1.87
2004	Apr	30	15	12	38.440	38.3632	21.7393	6.85	2.20
2004	May	01	08	01	34.310	38.3617	21.7245	6.74	2.03
2004	May	05	01	45	25.880	38.3285	21.7885	6.41	1.38
2004	May	05	14	30	6.240	38.3617	21.7408	6.83	1.07
2004	May	06	04	10	28.990	38.3700	21.7893	7.29	1.61
2004	May	09	10	15	45.980	38.3623	21.7888	6.77	2.31
2004	May	17	03	55	36.510	38.3918	21.7825	6.94	1.76
2004	May	22	22	44	50.990	38.2873	21.8197	13.67	2.43
2004	May	24	16	01	38.250	38.3745	21.7623	8.33	1.85

of this, SOMs do not require *a priori* information to function, and they excel at establishing unseen relationships in data sets (Penn, 2005). An important advantage of SOM is the ability to preserve topological relations, that is, patterns that are close in the input space will be mapped to neurons that are close in the output space and vice versa. SOM

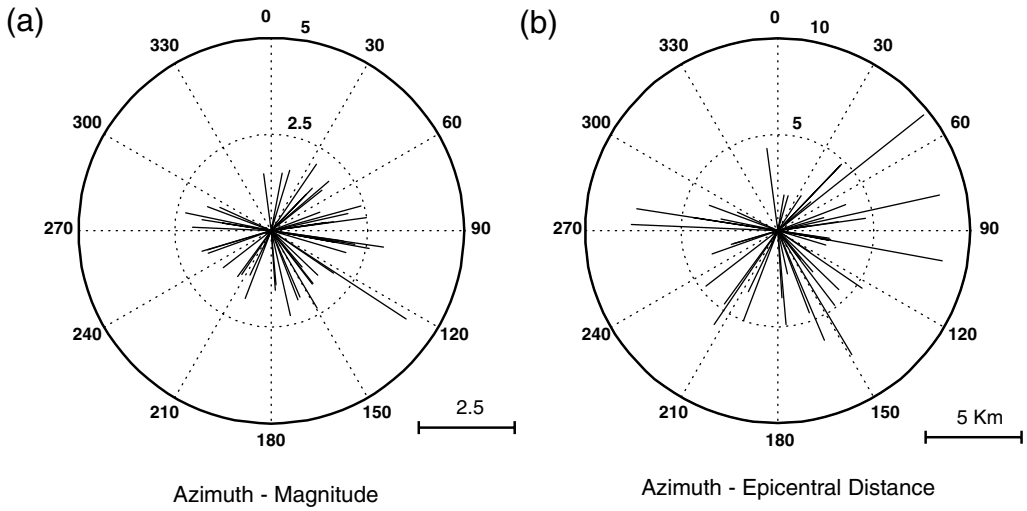

Figure 3. Azimuthal coverage for station 8 plotted in a polar diagram with (a) the magnitude of the events and (b) the epicentral distances in kilometers.

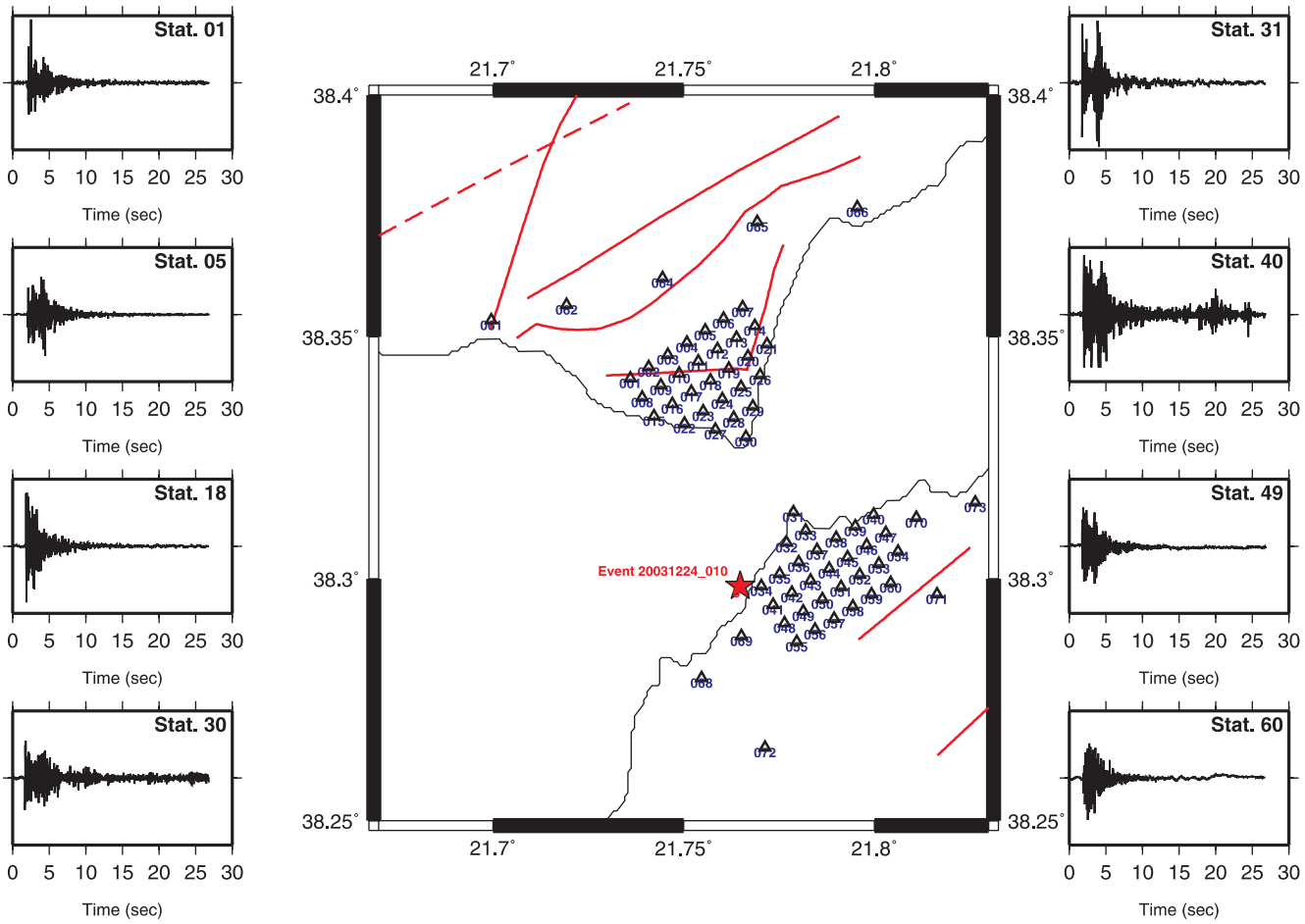

Figure 4. Example of recordings of the vertical components for event 1 at 8 stations. The color version of this figure is available only in the electronic edition.

capability in data clustering allows the analysis of multidimensional, nonlinear and highly noisy geophysical data (Klose, 2006).

Each neuron is described by an n-dimensional (n in this case is the dimensionality of input data) weight vector, w_i, with all the neurons arranged on a low-dimensional grid (Fig. 7). Every neuron of the network is connected with every other node of the data input layer. The topology of the map is determined by the interconnection of the neighboring neurons that are equidistant in the map space. The network represents a feed-forward structure with only one computational layer formed by neurons (Carniel *et al.*, 2009). The suitability of the self-organization process is ensured by initially using sufficient input vectors on the input neurons even though only the winning neuron and its neighbors adapt their connections.

In order to use the SOMs on our data, the HVSRs computed for each station will constitute the input vectors. Every input vector is introduced to the neural network creating a localized region of activity. This "bubble" will change its position and nature as the training process unfolds.

The basic stages for the formation of the SOM are, in brief, the following:

1. The values of the weight vectors, w_i, of the neurons are randomly initialized.
2. For every input vector, the value of the Euclidean distance between the input vector and the node's weight vector on the map is calculated for every neuron:

$$||\mathbf{x}_i - \mathbf{w}_j|| = \sqrt{\sum_{m=1}^{n} (\mathbf{x}_{i,m} - \mathbf{w}_{j,m})^2}, \qquad (2)$$

   where $\mathbf{x}_i = [x_{i,1} x_{i,2\ldots} x_{i,n}]$ is the input vector $i = 1, 2, \ldots, n$ with n being the dimension of the input space (in this case the number of the frequencies of the HVSR), and $\mathbf{w}_j = [w_{j,1} w_{j,2\ldots} w_{j,n}]^T$ the weight vector of a neuron $j = 1, 2, \ldots, m$, where m represents the total number of neurons.
3. After the values are calculated for all the nodes, the one that produces the smallest distance is found. This node will be the best matching unit (BMU).
4. The nodes in the neighborhood of the winning neuron are updated and "pulled" closer to the input vector. The

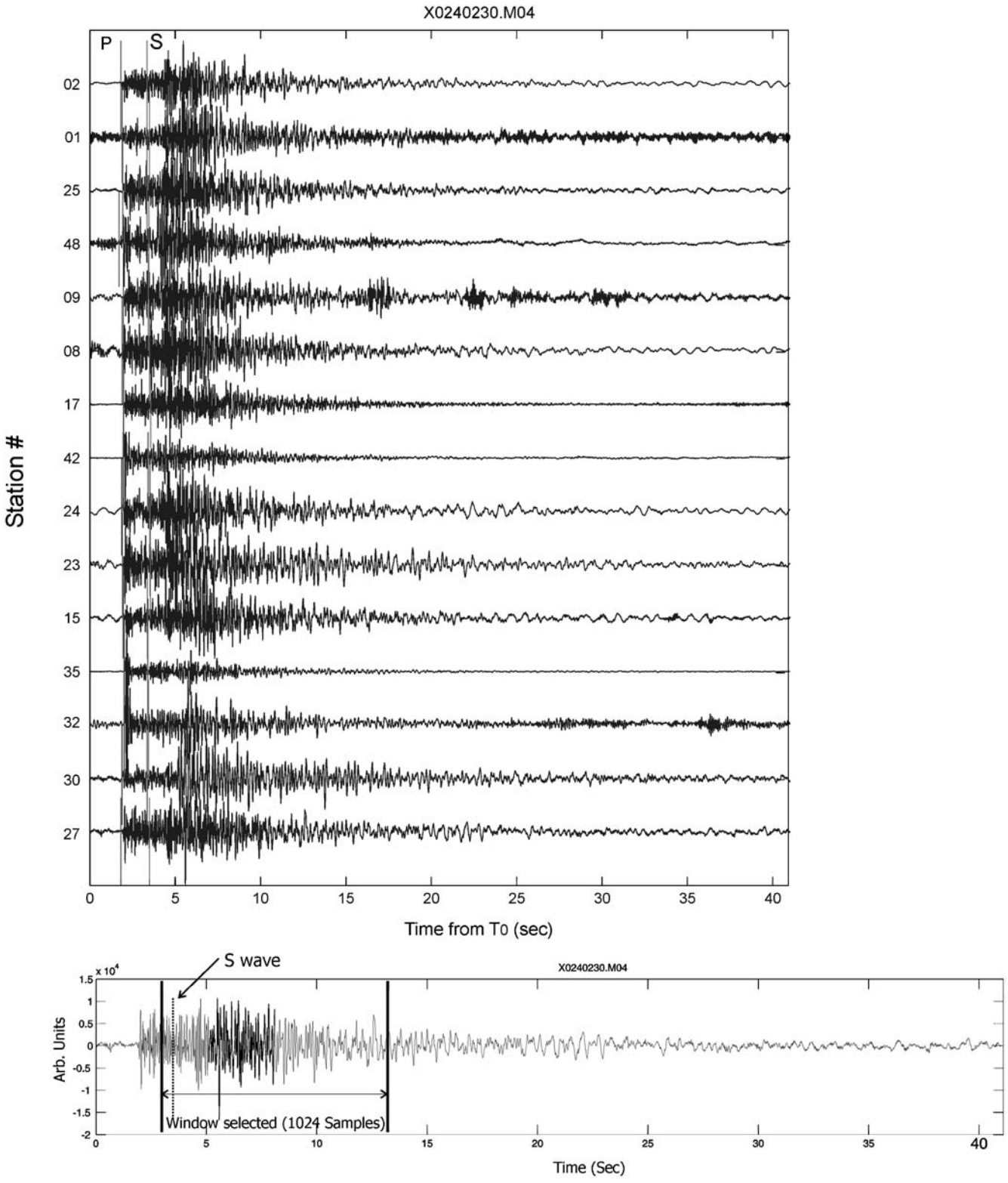

Figure 5. (a) Vertical components seismograms for 15 selected stations seismograms for event 72. The amplitudes of the records have been normalized. Also the *P* and *S* arrivals are marked (gray vertical lines). (b) Section of processing window.

neighborhood function ($h_{c(x)j,t}$) determines the connection between the neighboring neurons. A usual choice is the Gaussian function. Also the neighborhood function usually varies with the iteration number, t, because the effective radius of the neighborhood function is set smaller as the iterations go on.

5. The weight vectors of the neurons are modified in such a way that improves the matching of the BMU neuron with

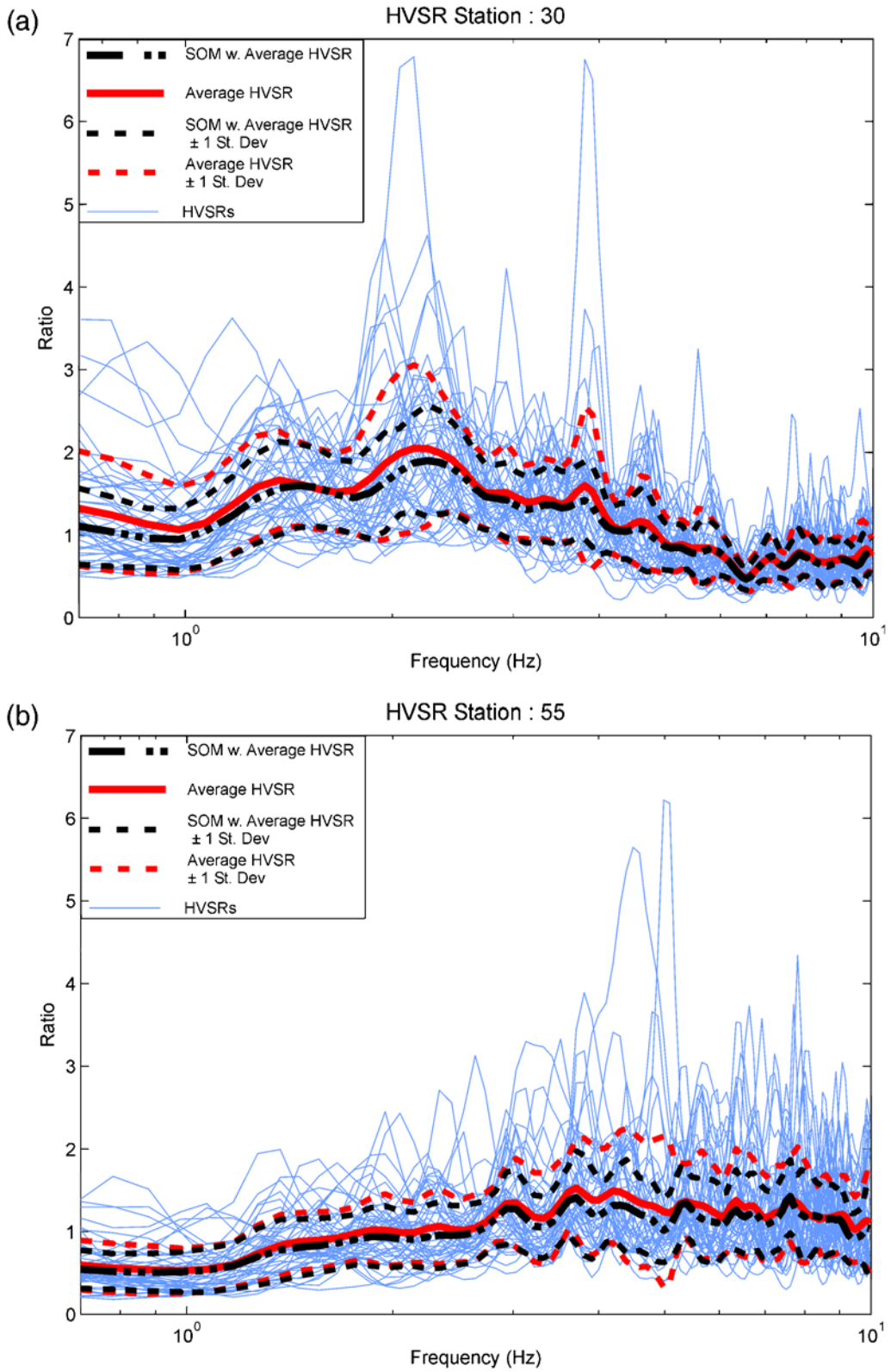

Figure 6. HVSRs for stations (a) 30 and (b) 55 (thin lines in background); the SOM weighted average and the average HVSR as well as the ± 1 standard deviation for the corresponding colors are overplotted. The color version of this figure is available only in the electronic edition.

the application of similar input patterns. This allows the map training and the formation of data clusters. This neuron adaptation process follows the function

$$\mathbf{w}_j(t+1) = \mathbf{w}_j(t) + a(t)h_{c(x)j}(t)[\mathbf{x}_i(t) - \mathbf{w}_j(t)], \quad (3)$$

where t is the iteration number and the learning rate factor, $0 \leq a(t) \leq 1$, that decreases monotonically with the iteration t. The learning factor rate controls the updating of the weight vectors.

The results of the SOM can be visualized by a variety of methods. The unified distance matrix (U-matrix) makes possible the 2D visualization of the SOMs by using properties of the topological relations among neurons and is one of the most frequently used. Each value of the U-matrix is calculated by taking the average distance measure between the corresponding node and its closest adjacent neighbors. Because of the way it is generated, the U-matrix can detect topological relations among neurons and make inferences about the structure of the input data.

Data Analysis

After processing the data as described in the previous section, the band of frequencies of interest (0.5 up to 10 Hz) are exported to be used in the SOM process. Each input vector x_i for event i will be formed using all the frequencies of the corresponding HVSR. The data (all the input vectors) are entered in a Matlab-based implementation of the SOM algorithms (Vesanto *et al.*, 1999), and then the SOM is trained. For every station, the corresponding SOM is created, and each input vector is mapped to a specific node in the output layer.

The next step is to calculate the U-matrix for every station's SOM separately in order to visualize the distances between the neurons of the output layer. In this way, areas/clusters can be located on the U-matrix where similar HVSRs correspond to relatively smaller "distances." As stated previously, an important advantage of the SOM technique is that the similar input data are mapped into neighboring neurons. Because of this property, the spectral ratios that have the more persistent characteristics will be located in the same or nearby neurons on the SOM. It is possible to locate the more densely populated neurons of the SOM by calculating the histogram of the number of input vectors that correspond on each neuron. The most populated neuron is selected, and

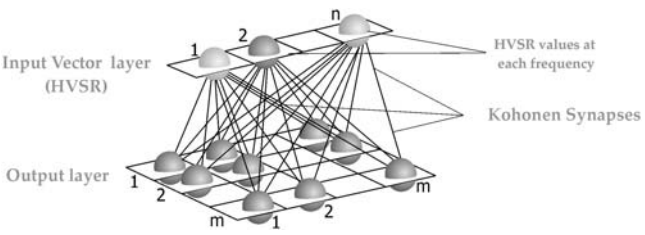

Figure 7. Mapping of the input vector data to the neurons.

the topographic errors of each input vector with this BMU are calculated. Figure 8a presents the calculated U-matrix for station 30 while Figure 8b shows the number of input data points mapped on each neuron of the SOM. The dominant node and its neighboring cluster can be seen in the upper right corner.

Based on the above process and using the calculated topographical errors between the input data and the U-matrix cell with the highest data distribution (most persistent characteristics), it is possible to calculate a weighted mean of the spectra used that will be biased towards the more stable characteristics. The classical form of the weight function is written as

$$w_i = \frac{h_i^{-p}}{\sum_{j=1}^{n} h_i^{-p}},$$

where n is the number of input spectra for each station, h represents the topographic errors of the input from the selected node, w_i is the weight assigned to each input spectra i, and p is an arbitrary positive real number called the power parameter.

The value of p can change depending on how much we want to amplify the HVSRs neighboring the BMU where the majority of input vectors are mapped. In this case a value of $p = 2$ was selected, but this could vary depending on how much we want to emphasize the more persistent characteristics.

The weighted average ($S_{w_{\text{mean}}}$) then, is calculated using the formula

$$\mathbf{S}_{w_{\text{mean}}} = \sum_{i=1}^{n} w_i \mathbf{S_i},$$

where S_i is the ith input spectral ratio.

The SOM weighted average HVSR, when compared to the average HVSR for a site with stable and similar ratios, should not be very different. In the case of less stable average HVSRs, the peaks are generally sharper; wide peaks might be separated into two peaks or become sharper to one value, and also, some minor peaks can be attenuated.

If the majority of the input vectors are distributed on one node of the SOM or are closely located in neighboring nodes, forming a cluster, most of the calculated HVSRs have some persistent characteristics, and the final result is robust. Because of this, we expect little difference between weighted-average HVSR and average HVSR, but the common features of the ratios are more pronounced. In this way there is one result for each station from which the dominant frequencies and spectral ratios are extracted.

The weighted-average HVSRs for each station are plotted side by side with each other in order to have an initial estimation of the differences and similarities of the sites as well as to see the site response values. Figure 9a shows the SOM weighted-average HVSRs for 16 selected stations. Care should be taken that there is a sufficient number of input

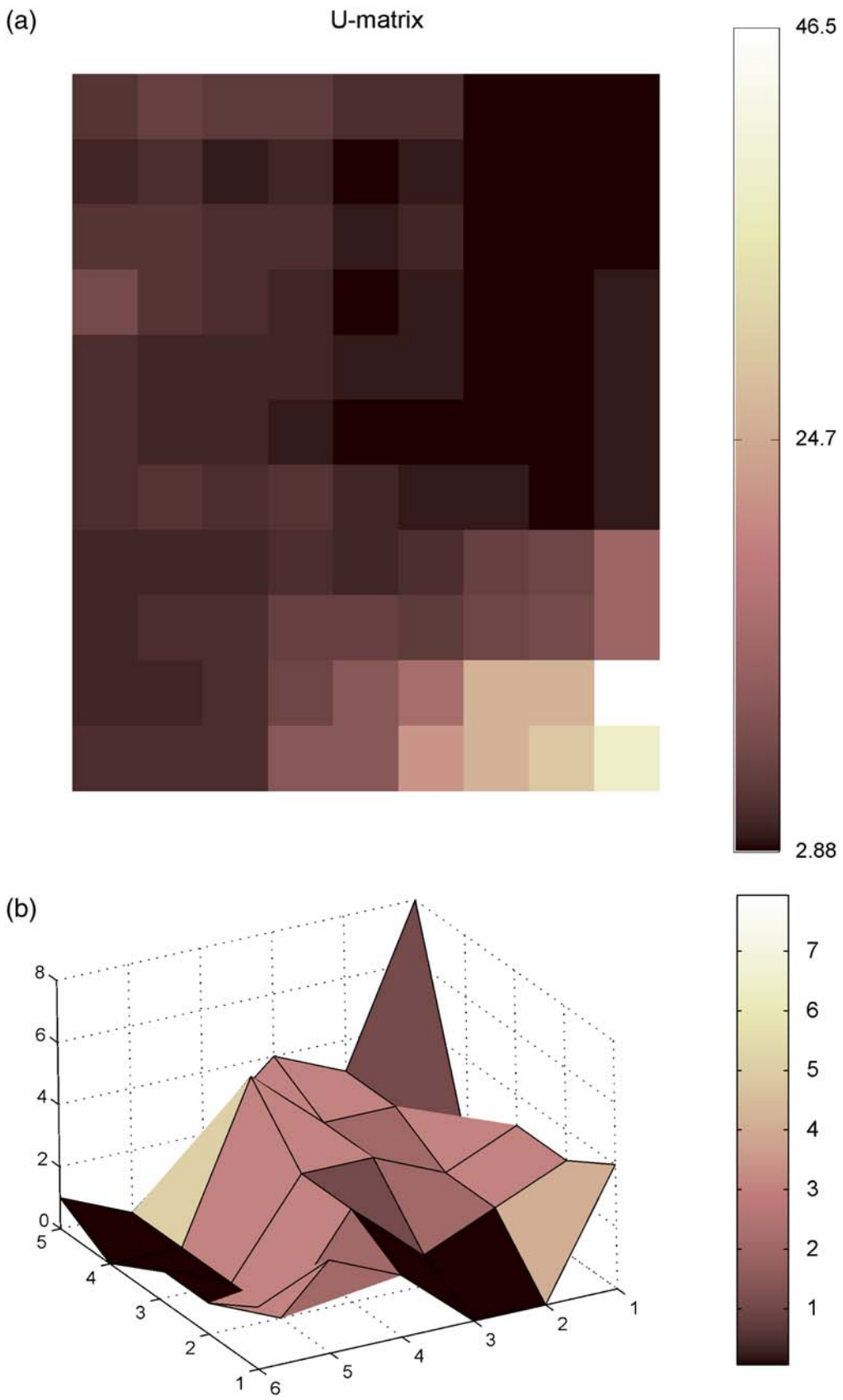

Figure 8. (a) The U-matrix calculated for the events recorded in station 30 (smaller values indicate smaller distance between adjacent neighbors) and (b) the number of input data points on each node of the SOM. The color version of this figure is available only in the electronic edition.

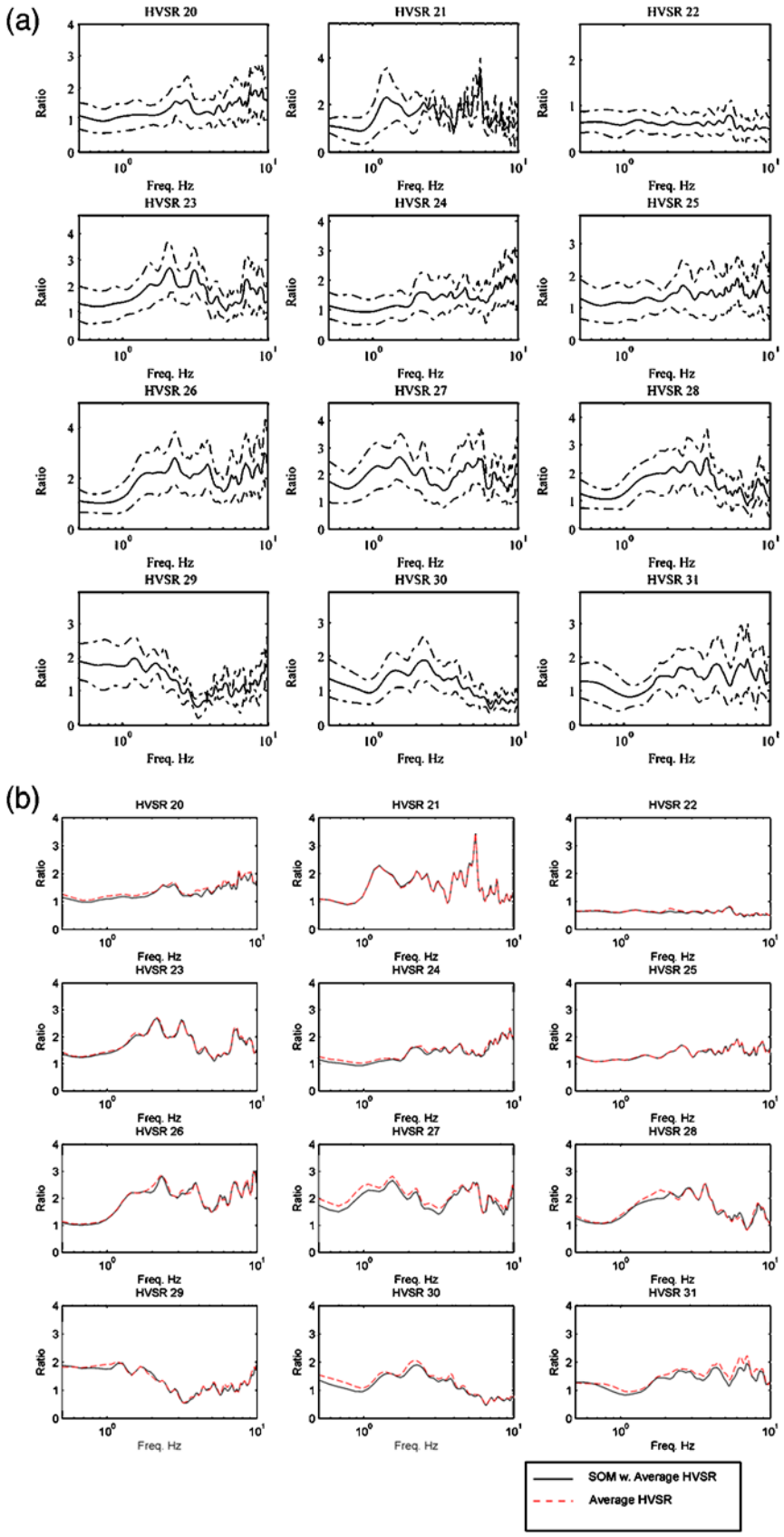

Figure 9. (a) Average HVSRs for 16 selected stations for frequencies between 0.5 and 20 Hz. (b) Comparison of average HVSR with the SOM weighted average for the same stations. The color version of this figure is available only in the electronic edition.

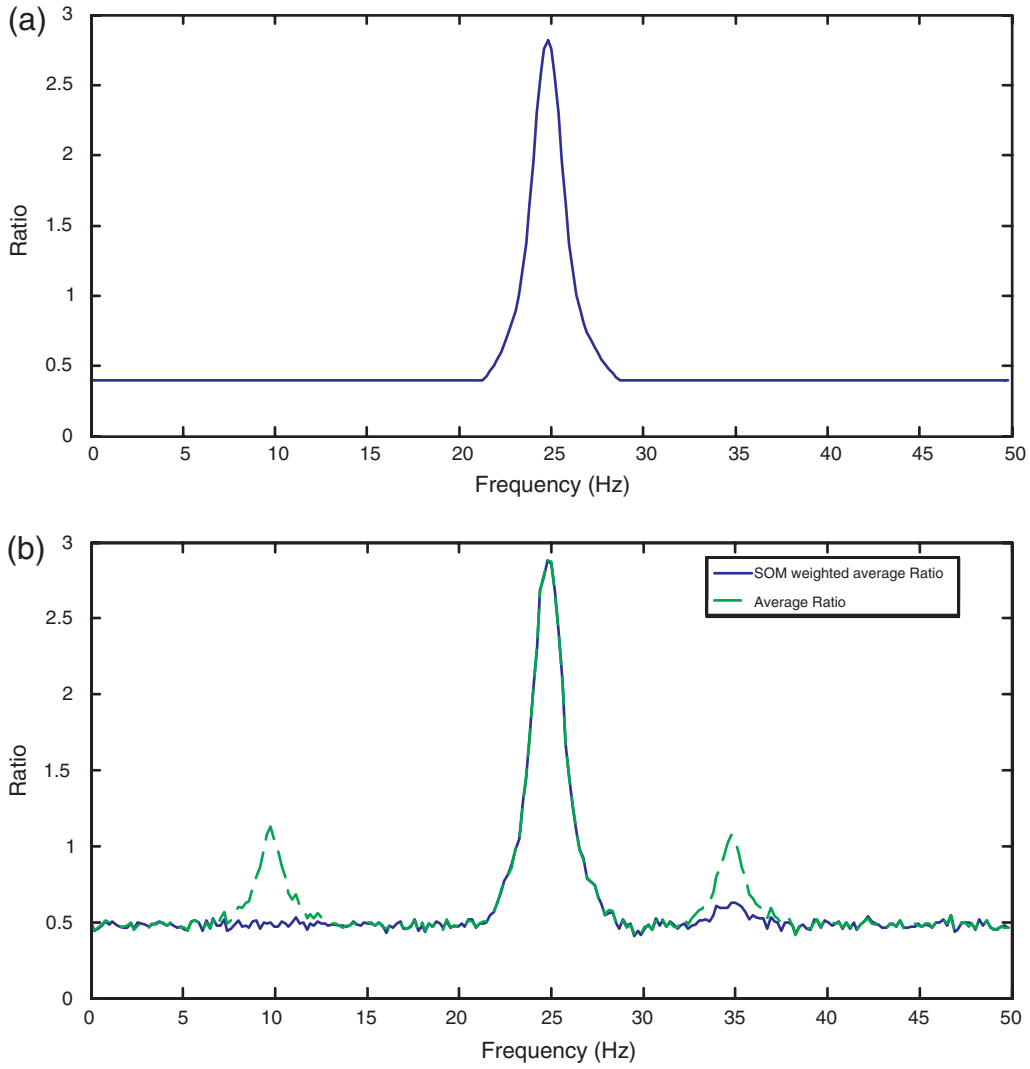

Figure 10. (a) The initial synthetic spectrum. (b) The average spectrum is plotted over the SOM weighted-average spectrum. The color version of this figure is available only in the electronic edition.

spectra. In this way a site-response map is prepared for the region covered by the seismic stations. Figure 9b compares the average HVSR for the same stations with the SOM weighted average. Obviously, this is an additional indication/quality control of the HVSR result and its stability.

Synthetic Test

In order to assess the results of the proposed methodology, a test with a synthetic data set is performed. Initially a synthetic spectrum ratio (Fig. 10a) was created that had a peak at 25 Hz representing the persistent feature of the ratios. Based on that, 100 synthetic spectral ratios were generated. Of those, 80 consisted of the initial with the addition of white noise. Eighteen of them had in addition to noise, a second peak at 35 Hz with equal ratio value, and finally, two had in addition to noise, a second peak at 35 Hz as before and a new dominant peak at 10 Hz that was 8 times the value of the initial peak. Next, the average synthetic ratio was calculated (Fig. 10b).

The SOM algorithm was applied to the data set. The U-matrix is calculated, and using the hit matrix, the most populated node and its cluster is identified. (Fig. 11a,b) The calculated weights are used in order to assess the SOM weighted-average spectral ratio.

By comparing the two results, it can be seen that, although in the average ratio the two secondary peaks still exist even though attenuated, the SOM weighted average manages to almost completely remove the effect of the peak at 10 Hz and significantly lowers the peak at 35 Hz.

Results

The calculated HVSR amplifications corresponding to all 95 events for all the station sites in Figure 4 are depicted in Figure 12.

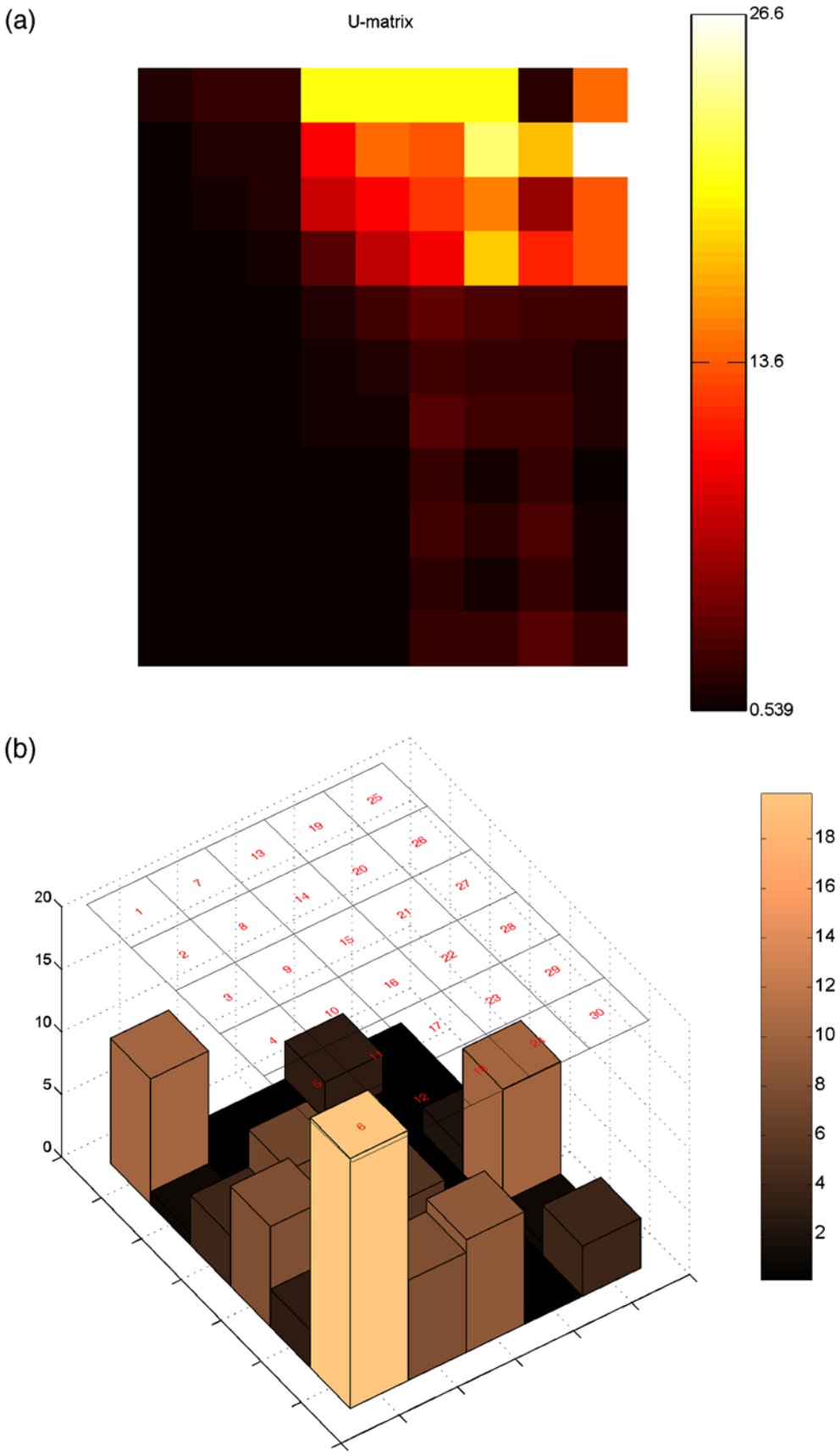

Figure 11. (a) The U-matrix calculated for the synthetic test; (b) the number of input data that is mapped on each node of the SOM. The color version of this figure is available only in the electronic edition.

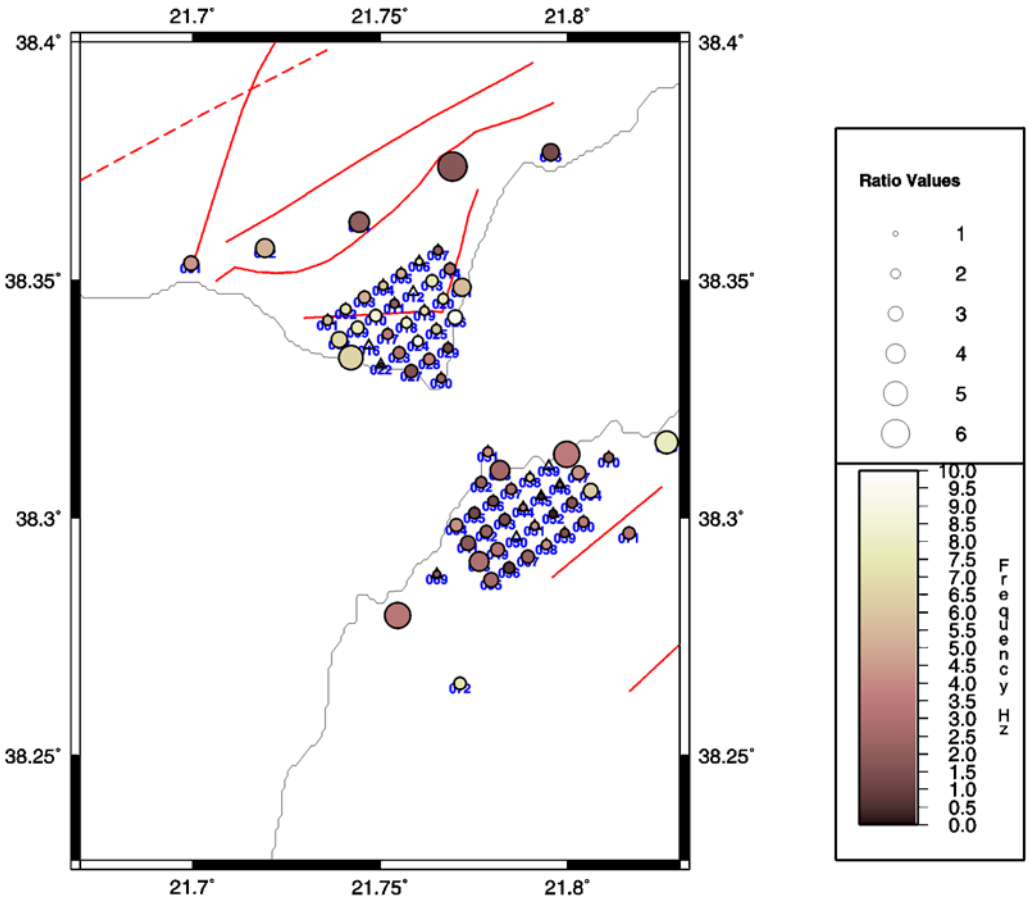

Figure 12. The HVSRs (circle radii) and the predominant frequencies (color scale) for all the instrumented positions. The color version of this figure is available only in the electronic edition.

It can be seen that the dominant frequencies are lower on the Rio side of the strait, where the soft recent sediments are predominant. They range mostly from 1–3 Hz with the average mean amplification factor for soft sediments varying from 0.7 up to 3.5. On the Antirrio site of the Strait, there seems to be a wider range of dominant frequencies that are also generally higher. In general, with the exception of several sites, the stations have predominant frequencies between 1 and 7 Hz with the mean amplification factors ranging mostly from 1 to 3. On the Antirrio side, it seems that the most recent sediments have lower frequencies increasing slightly towards the Pleistocene sediments, but even taking into account stations within the same geological surface formations, the results are more heterogeneous than on the Rio side.

Refering to the results of SOM weighted averaging, the calculated ratios are almost unaffected at the sites where the ratios are stable and a sufficient number of events are recorded (at least 5–20 depending on the case). The sites mostly affected are those with few recorded events, characterized by not-so-stable ratios or contained in some ratios with very large values. This method can be a tool for evaluating the stability and quality of the calculated ratios.

The properties of the SOMs can also be used to investigate the clustering of the resulting HVSRs. For some HVSRs, those corresponding to stations where the peak frequencies are not clearly distinguished, clustering of the HVSR can be more reliable. For this reason, the final HVSRs are used again to calculate a new SOM to define similar behavior clusters using the *k*-means for the clustering. The Davies–Bouldin index (Davies and Bouldin, 1979) is calculated based on the average maximal distance of each cluster to the others and is used as a measure of the cluster separation. Based on this, for this data set, nine HVSR clusters were formed (Fig. 13), and the cluster number of the corresponding stations was plotted on the map of the area. (Fig. 14).

In the city of Antirrio and the immediate vicinity, the relatively flat HVSR clusters 5 and 7 are situated in the Pleistocene deposits or very close to their outcrop. In the Rio area, there are a number of stations that have been classified to cluster 7, but the geological data currently at our disposal are not enough to confirm this zonation.

Conclusion

A new method for quality control and improvement of the results of the HVSR method for weak events has been applied. By using the SOMs, it is possible to have an indica-

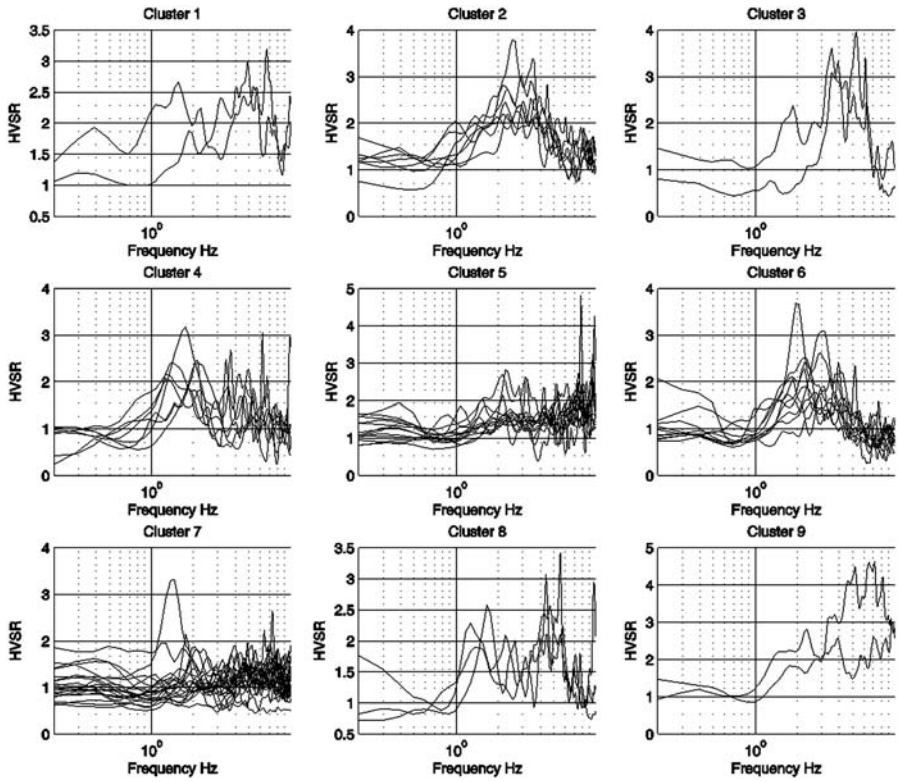

Figure 13. Plot of the HVSRs separated by the 9 clusters.

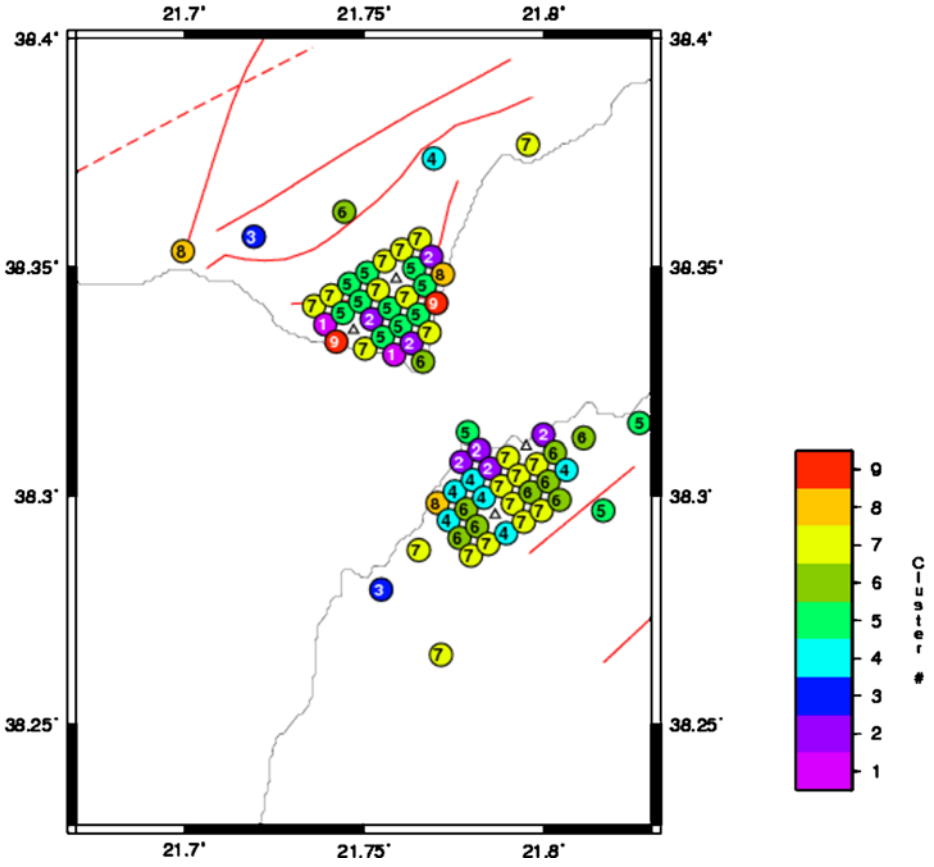

Figure 14. Map of the corresponding clusters. The color version of this figure is available only in the electronic edition.

tion of the similarity between the HVSRs of all the events for every station. Also, using the topographic errors, it is possible to calculate the average HVSR, giving larger emphasis to the spectral ratios with the more persistent characteristics. Also, the SOM algorithm can be used in automating the separation of areas with HVSRs with similar behavior.

The method has been applied to the Rio-Antirrio region for interpreting earthquake data collected on a dense grid of 70 urban sites. The method has allowed us to better evaluate the stability of the calculated ratios and improve our knowledge of the local seismic response. However, the empirical transfer functions estimated in this study have to be modified to include nonlinear effects in order to extend our results to strong ground shaking.

Data and Resources

Seismograms used in this study were collected and provided by Landtech Enterprises SA and are proprietary. They cannot, currently, be released to the public.

Acknowledgments

We would like to acknowledge Stefano Parolai, Pier Luigi Bragato, and Raúl R. Castro for their useful comments.

References

Aki, K. (1988). Local site effects on ground motion, in *Proc. of the Conf. on Earthquake Engineering and Soil Dynamics II-Recent Advances in Ground Motion Evaluation*, 20**,** American Society of Civil Engineering, 103–155.

Bard, P.-Y. (1999). Microtremor measurements: A tool for site effect estimation?, in *Proc. of Second International Symposium on the Effects of Surface Geology on Seismic Motion*, Yokohama, Japan; Balkema, Roterdam, 3**,** 1251–1279.

Bonilla, L. F., J. H. Steidl, J. Gariel, and R. J. Archuleta (2002). Borehole response studies at the garner valley downhole array, Southern California, *Bull. Seismol. Soc. Am.* **92,** 8, 3165–3179.

Bonilla, L. F., J. H. Steidl, G. T. Lindley, A. G. Tumarkin, and R. J. Archuleta (1997). Site amplification in the San Fernando Valley, California: Variability of site-effect estimation using the *S*-wave, coda and H/V methods, *Bull. Seismol. Soc. Am.* **87,** 3, 710–730.

Borcherdt, R. D. (1970). Effects of local geology on ground motion near San Francisco Bay, *Bull. Seismol. Soc. Am.* **60,** 1, 29–61.

Carniel, R., L. Barbui, and P. Malisan (2009). Improvement of HVSR technique by self-organizing map (SOM) analysis, *Soil Dynam. Earthq. Eng.* **29,** 6, 1097–1101.

Carver, D., and S. H. Hartzell (1996). Earthquake site response in Santa Cruz, California, *Bull. Seismol. Soc. Am.* **86,** 1A, 55–65.

Castro, R. R., M. Mucciarelli, F. Pacor, and C. Petrungaro (1997). *S*-wave site-response estimates using horizontal-to-vertical spectral ratios, *Bull. Seismol. Soc. Am.* **87,** 1, 256–260.

Davies, D. L., and W. Bouldin (1979). A cluster separation measure, *IEEE Trans. Pattern Anal. Mach. Intell.*. **1,** 224–227.

Di Giulio, G., R. M. Azzara, G. Cultrera, M. S. Giammarinaro, P. Vallone, and A. Rovelli (2005). Effect of local geology on ground motion in the city of Palermo, Italy, as inferred from aftershocks of the 6 September 2002 M_w 5.9 earthquake, *Bull. Seismol. Soc. Am.* **95,** 6, 2328–2341.

Doutsos, T., and G. Poulimenos (1992). Geometry and kinematics of active faults and their seismotectonic significance in the western Corinth-Patras rift (Greece), *J. Struct. Geol.* **14,** 6, 689–699.

Doutsos, T., N. Kontopoulos, and G. Ferentinos (1985). Das westlichc Ende des Korinth-Grabens, *N. Jb. Geol. Paläont. Mh.* **1,** 652–666.

Doutsos, T., N. Kontopoulos, and G. Poulimenos (1988). The Corinth-Patras rift as the initial stage of continental fragmentation behind an active island arc (Greece), *Basin Res.* **1,** 3, 177–190.

Field, E. H., and K. H. Jacob (1995). A comparison and test of various site-response estimation techniques, including three that are not reference-site dependent, *Bull. Seismol. Soc. Am.* **85,** 4, 1127–1143.

Haghshenas, E., P. Bard, N. Theodulidis, and SESAME WP04 Team (2008). Empirical evaluation of microtremor H/V spectral ratio, *Bull. Earthq. Eng.* **6,** 1, 75–108.

Hartzell, S., A. Leeds, A. Frankel, and J. Michael (1996). Site response for urban Los Angeles using aftershocks of the Northridge earthquake, *Bull. Seismol. Soc. Am.* **86,** 1B, S168–S192.

Improta, L., G. D. Giulio, and A. Rovelli (2005). Variations of local seismic response in Benevento (Southern Italy) using earthquakes and ambient noise recordings, *J. Seismol.* **9,** 2, 191–210.

Jongmans, D., and M. Campillo (1993). The response of the Ubaye Valley (France) for incident *SH* and *SV* waves: Comparison between measurements and modeling, *Bull. Seismol. Soc. Am.* **83,** 3, 907–924.

Kanai, K. (1951). Relation between the nature of surface layer and the amplitudes of earthquake motions I, *Bull Earthq Res Inst, Tokyo Univ.* **30,** 31–37.

Klose, C. D. (2006). Self-organizing maps for geoscientific data analysis: Geological interpretation of multidimensional geophysical data, *Comput. Geosci.* **10,** 3, 265–277.

Kohonen, T. (1997). *Self-Organizing Maps*, Series in Information Sciences, Springer-Verlag, New York.

Lachet, C., D. Hatzfeld, P. Bard, N. Theodulidis, C. Papaioannou, and A. Savvaidis (1996). Site effects and microzonation in the city of Thessaloniki (Greece) comparison of different approaches, *Bull. Seismol. Soc. Am.* **86,** 6, 1692–1703.

Lermo, J., and F. J. Chávez-García (1993). Site effect evaluation using spectral ratios with only one station, *Bull. Seismol. Soc. Am.* **83,** 5, 1574–1594.

Mandal, P., R. Chadha, C. Satyamurty, I. Raju, and N. Kumar (2005). Estimation of site response in Kachchh, Gujarat, India, region using H/V spectral ratios of aftershocks of the 2001 M_w 7.7 Bhuj earthquake, *Pure Appl. Geophys.* **162,** 12, 2479–2504.

Milne, J. (1898). *Seismology*, Kegan Paul, London, 320 pp..

Mucciarelli, M., M. R. Gallipoli, and M. Arcieri (2003). The stability of the horizontal-to-vertical spectral ratio of triggered noise and earthquake recordings, *Bull. Seismol. Soc. Am.* **93,** 3, 1407–1412.

Nakamura, Y. (1989). Method for dynamic characteristics estimation of subsurface using microtremor on the ground surface, *Q. Rep. Railway Tech. Res. Inst. (Japan)* **30,** 1, 25–33.

Nogoshi, M., and T. Igarashi (1971). On the amplitude characteristics of microtremor (Part 2), *Zisin* **24,** 1, 26–40 (in Japanese with English abstract).

Penn, B. S. (2005). Using self-organizing maps to visualize high-dimensional data, *Comput. Geosci.* **31,** 5, 531–544.

Steidl, J. H., A. G. Tumarkin, and R. J. Archuleta (1996). What is a reference site?, *Bull. Seismol. Soc. Am.* **86,** 6, 1733–1748.

Toshinawa, T., J. J. Taber, and J. B. Berrill (1997). Distribution of ground-motion intensity inferred from questionnaire survey, earthquake recordings and microtremor measurements—A case study in Christchurch, New Zealand, during the 1994 Arthurs Pass earthquake, *Bull. Seismol. Soc. Am.* **87,** 2, 356–369.

Tselentis, G.-A, and K. C. Makropoulos (1986). Rates of crustal deformation in the Gulf of Corinth (central Greece) as determined from seismicity, *Tectonophysics* **124,** 1–2, 55–57, 61–66.

Tselentis, G.-A. Serpetsidaki, N. Martakis, E. Sokos, P. Paraskevopoulos, and S. Kapotas (2007). Local high-resolution passive seismic tomography and Kohonen neural networks—Application at the Rio-Antirio Strait, central Greece, *Geophys.* **72,** 4, B93–B106.

Tselentis, G.-A., P. Paraskevopoulos, and N. Martakis (2010). Intrinsic Q_p seismic attenuation from rise time of microearthquakes. A local scale

application at Rio-Antirrio, Western Greece, *Geophys. Prospect.*, **58,** 845–859.

Vesanto, J., E. Alhoniemi, J. Himberg, K. Kiviluoto, and J. Parviainen (1999). Self-organizing map in Matlab: the SOM toolbox, *Simulation News Europe* **25,** 54.

Zaslavsky, Y., A. Shapira, M. Gorstein, M. Kalmanovich, V. Giller, N. Perelman, I. Livshits, D. Giller, and I. Dan (2005). Site response from ambient vibrations in the towns of Lod and Ramle (Israel) and earthquake hazard assessment, *Bull. Earthq. Eng.* **3,** 3, 355–381.

Patras Seismological Laboratory
University of Patras
26 500 Rio, Greece
tselenti@upatras.gr
paris@upatras.gr

Manuscript received 13 November 2009