Contents lists available at ScienceDirect

Soil Dynamics and Earthquake Engineering

journal homepage: www.elsevier.com/locate/soildyn

High-resolution local seismic zonation by cluster and correlation analysis

Giovanna Laurenzano [a,*], Marco Garbin [a,b], Stefano Parolai [c,a], Carla Barnaba [a], Marco Romanelli [a], Luca Froner [b]

[a] *Istituto Nazionale di Oceanografia e di Geofisica Sperimentale – OGS, Italy*
[b] *Provincia Autonoma di Trento, Servizio Geologico, Italy*
[c] *Università degli Studi di Trieste, Dipartimento di Matematica e Geoscienze, Italy*

ARTICLE INFO

ABSTRACT

Site response analysis is essential for seismic hazard and risk assessment and for providing useful data for land-use planning. However, current regional site amplification models do not always have the resolution required for sites such as alpine valleys, in which site response is characterized by complex effects. In addition, even in local studies, site response is usually computed at a limited number of sites because it requires the recording of earthquakes through the installation of seismological stations. To spatially extend the site response to a denser grid of points inside the investigated area, we used k-means cluster and correlation analyses and Voronoi tessellation. The method was applied to the evaluation of the site response of the lower Sarca Valley on the northern shore of Lake Garda. Earthquakes were recorded at 19 sites to calculate site response in terms of amplification and duration functions of ground motion. The results show high amplification values (up to 10) at low frequencies (at about 0.7 Hz) in the center of the valley, where the sediments reach a thickness of about 420 m. Moderate amplification values and duration lengthening of several seconds in the range of 1–10 Hz are found instead at the edge of the valley on the sedimentary deposits, while a lack of amplification is observed for the sites located on the bedrock. Both the amplification and duration functions were assigned to the area covered by the single-station noise measurements to obtain a zonation of the study area, resulting in three zones to which Fourier amplification factors can be assigned for specific frequency values. The results obtained can be used directly for hazard and risk scenarios and to improve regional maps with lower resolution at the local level.

1. Introduction

Local seismic effects are evaluated to quantify the influence of site conditions (i.e., geology, topography, morphology, etc.) on seismic ground motions observed during earthquakes (see for instance, Kaklamanos et al. [1] and references therein). In many Alpine valleys, although major investments have been made in buildings and infrastructure, construction planning and land-use regulations do not always adequately consider the nature and extent of seismic site effects, even though these valleys are mostly located in regions of low or moderate earthquake hazard (Faccioli et al. [2]). The lower Sarca Valley on the northern shore of Lake Garda is vulnerable to site response because of the combined 1-D to 3-D effects acting on the thick deposits which compose the sedimentary basin (Faccioli et al. [2], Garbin et al. [3] and Parolai et al. [4]). The valley is densely populated, and evaluation of local seismic effects (site response) becomes necessary for seismic risk mitigation, especially in urban areas (Vessia et al. [5]).

In recent years, the exponential increase in the availability of waveforms, together with technological developments that allow processing large amounts of data, has led to the production of seismic site response maps at regional scales. For example, Bindi et al. [6] produced empirical maps of site amplification for Europe by processing a large number of seismic recordings retrieved at 3200 sites, while Weatherill et al. [7] provided pan-European site amplification model based on VS30 (i.e. the average seismic shear-wave velocity of the uppermost 30 m) maps inferred from topography and geology and on site-to-site ground motion residuals analysis.

Although these kinds of maps provide an important indication of the level of site amplification at large-regional scale, site effects are often determined by local geologic conditions that vary on a shorter spatial scale, as for example in Alpine valleys, in which site response is characterized by complex effects (Faccioli et al. [2,8]).
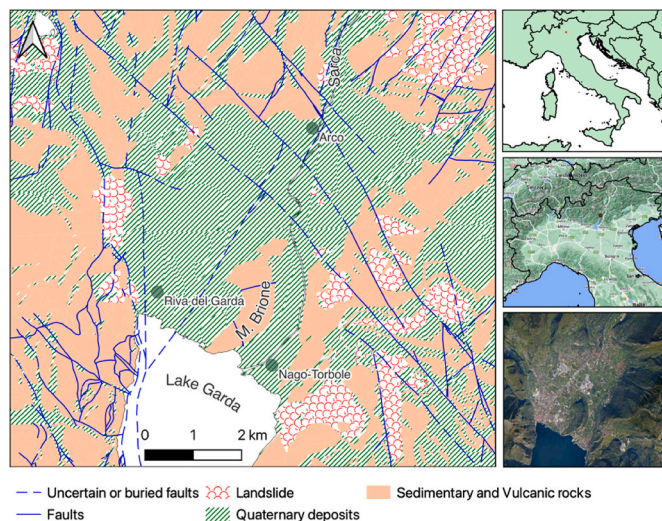

Fig. 1. Overview of the study area with the most important geographical and geological features (based on the Geological Map of Servizio Geologico of Provincia Autonoma di Trento). The approximately NNE-SSW and NW-SE trending faults belong to the Giudicarie and the Schio-Vicenza fault systems, respectively (Castellarin et al.[22]).

Several studies relying on the recordings of temporary networks have enabled estimating site response in a specific area, but very often the distances between recording stations are not sufficient to allow full zonation of the area. However, in the last 20 years, some attempts have been made to solve this problem, in particular by the application of cluster analysis. Bragato et al. [9] proposed a technique using a Bayesian approach to find the connected areas where the similarity between Horizontal to Vertical H/V spectral ratios of Noise (NHV) is maximized. Tselentis and Paraskevopoulos [10] used the self-organizing map technique for cluster analysis based on the similarity of H/V spectral ratios of weak motions. Strollo et al. [11] and Ullah et al. [14] proposed a method to improve the spatial resolution of spectral intensity correction factors and Reference Site Spectral Ratios (RSSR, Borcherdt [12] and Field and Jacob [13]), respectively, by taking advantage of cluster and correlation analysis and using NHV calculated from noise recordings as a proxy. Paolucci et al. [15] used the principal component analysis for identifying sites characterized by similar NHV curves while Capizzi and Martorana [16] recently applied cluster analysis to NHV data for subsurface studies.

In this study, we used a method based on the approach Strollo et al. [11] and Ullah et al. [14] to obtain high-resolution zonation of the seismic site effects of the lower Sarca Valley. We added to cluster and correlation analyses the Voronoi tessellation to spatially extend the site response obtained from earthquake recordings at a limited number (19) of sites to the entire area of the lower Sarca Valley. The method is based on the use of NHV curves of seismic noise obtained at a large number of sites (i.e., 110) as a proxy. We have demonstrated the validity of our approach not only using spectral amplification functions, but also with the original idea of considering the duration lengthening of the seismic shaking. The evaluation of this latter parameter, often overlooked in site response analysis, is essential in contexts such as sedimentary basins or valleys, exhibiting remarkable lengthening of duration from the edges to the center (Beauval et al. [17]), since long duration can lead to non-negligible damage to buildings exposed to ground shaking (Hancock and Bommer [18]). A methodology to evaluate the frequency dependent prolongation of the signal duration of a site was proposed by Parolai and Bard [19], combining the analysis of spectrogram and the RSSR) by the Joint Analysis of Sonogram and Standard Spectral Ratio Methodology (JASSSR).

In the following, we first describe the study area and present the

methods (i.e., the Generalized Inversion Technique - GIT (Andrews [20]) and the JASSSR [19]) that were used to calculate site response at the 19 sites of the temporary network. We then describe in detail the application of cluster and correlation analyses to amplification and duration curves to assign an empirically estimated site response to the 110 sites of noise recordings, and the application of Voronoi tessellation to obtain zonation maps. some

2. Description of the study area

The study area (Fig. 1) is the final part of the Sarca Valley, which is about 5–6 km long and lies roughly between the narrow stretch before the village of Arco to the north and Lake Garda to the south. The valley floor here is about 3–4 km wide and is divided in the south by the ridge of M. Brione, which extends for about 2.5 km from the coast of Lake Garda towards the NNE. The Sarca River flows southward into the narrower eastern part. The valley is surrounded by a rather complex orography, where steep slopes are interrupted by hanging valleys.

The area is densely populated, in particular, close to Lake Garda and along the northern side of the main valley, with the three municipalities of Riva del Garda, Nago-Torbole and Arco and their numerous hamlets. In the valley center there are several industrial and craft enterprises, while the rest of the area, including the slopes, is used for cultivation.

From a tectonic and structural point of view (Fig. 1), the study area belongs to the central sector of the southern Alps, where the seismically active Giudicarie thrust belt, approximately NNE-SSW trending, interferes with the NW-SE faults of the Schio—Vicenza system (Castellarin et al. [21,22]). In particular, the western border of the lower Sarca Valley marks the transition between two different paleogeographic domains: the pelagic domain of the Lombardian basin to the west and the Trento platform to the east. The development of this escarpment was determined by an extensive fault with orientation around NNE-SSW, known as the Ballino-Garda fault, with a length of almost a hundred kilometers (Castellarin et al. [21]).

In the study area, the geological succession exposed on the reliefs spans from the upper Trias (Calcare di Zu) to lower Miocene (Formazione del Monte Brione) (Castellarin et al. [21] and Luciani [23]). Plio-Quaternary deposits, organized in several synthems, connect the slopes to the valley floor, where they reach remarkable thickness even in the order of several hundreds of meters (Felber et al. [24]). As documented by a deep drill hole reaching the bedrock at Riva del Garda, the valley is filled by the Quaternary sedimentary deposits created both by the lake, the Sarca River and smaller creeks nearby. A reflection seismic survey NW - SE oriented, approximately perpendicular to the valley axis, between Riva del Garda and Arco, shows the discontinuity between rock and sediment fill at a depth of about 420 m at its deepest point with respect to the ground surface.

The seismicity of the area is mainly related to the interference between the aforementioned systems (Carulli & Slejko [25] and Viganò et al. [26]). The documented events set in the studied area are the May 22, 1868 MW=4.4, the June 20, 1942, MW=4.1 and the December 13, 1976 MW=4.9 earthquakes, this latter reaching the VII degree of macroseismic effects at Riva del Garda (Locati et al. [27]). However, in the surrounding areas, the IX degree was reached for seismic events that also triggered important landslides; these are the 1046 Middle Adige Valley and the 1117 Verona events (Guidoboni et al. [28], Martin et al. [29] and Ruggia et al. [30]). The instrumental seismicity of the last tens of years shows that epicenters are mainly clustered in the immediate southeastern part of the study area (Viganò et al. [31]).

The seismic hazard level of the lower Sarca Valley is moderate, according to the seismic hazard model MPS04 (Stucchi et al. [32]). The peak ground acceleration with 10% probability of exceedance in 50 years spans from 0.100 g to 0.150 g.
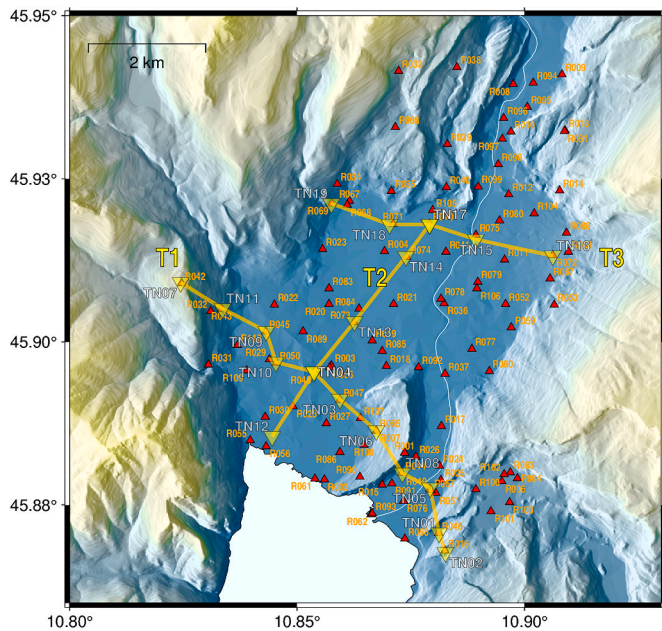
Fig. 2. Map of the location of the seismic noise (red triangles) and earthquake recordings (inverted yellow triangles). The latter are aligned along the three transects T1, T2 and T3, which are indicated by yellow lines.
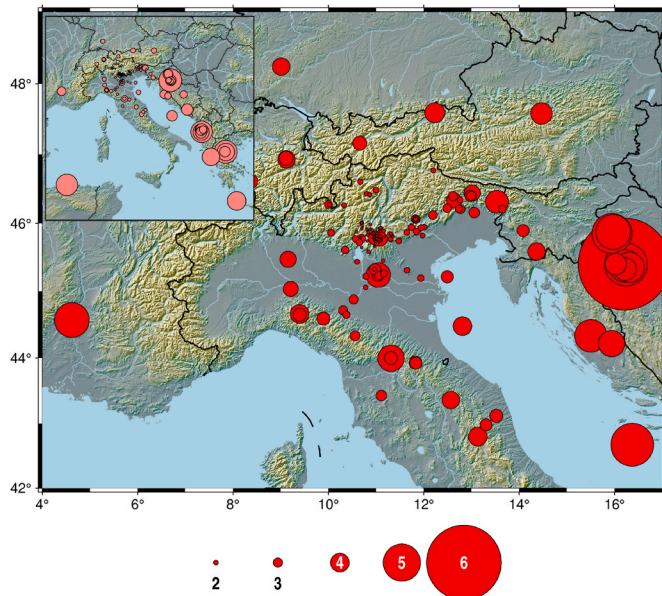

Fig. 3. Map of the epicenters of the earthquakes whose recordings were used in this study. Circle size indicates event's magnitude.

3. Dataset and methodologies for the evaluation of site response

Site response was estimated using earthquake recordings at 19 sites located in the valley (Fig. 2). The applied methodologies include the GIT, which was used to evaluate the spectral amplification curves and the JASSSR, used to evaluate the duration lengthening with respect to a reference site.

Continuous recordings of ground motion were collected over a period of about 2 years (from 2019 to 2021) to create a database of earthquake waveforms. The equipment of the mobile stations consisted of Lennartz LE-3D/1s and Lennartz LE-3D/5 three-component velocimetric sensors, a 24-bit Reftek RT130 data logger recording at a

sampling rate of 100 s.p.s, a GPS antenna and a battery connected to a photovoltaic system where no power supply was available. One station (TN06) was equipped with a 20-s Nanometrics Trillium Compact TC20 sensor and a Nanometrics Centaur digitizer.

As shown in Fig. 2, the stations were aligned along three directions (transects): two transverse (T1 and T3) and one longitudinal (T2) to the lower Sarca Valley with a spacing of about 1 km between stations. Along transect T1, the valley is divided by M. Brione into two parts: the western larger part, where the Riva del Garda historical center stands, and the eastern narrower part, crossed by the actual course of Sarca, with the Nago-Torbole center. Transect T2 crossed the valley longitudinally from Riva del Garda to the south to Arco to the north, while transect T3 crossed the valley close to its narrowing at the northern end of the study area. About 4700 three-component recordings, related to 163 earthquakes with magnitude ML between 0.8 and 6.3 (Fig. 3) were selected for the site response assessment.

To evaluate the spectral amplification with respect to a reference site, we made use of the GIT (Andrews [20]), a well-established, extremely flexible and robust tool, which can be viewed as an extension of the more direct approach RSSR (Borcherdt [12]). GIT was applied to the well-known model, which expresses the shear-wave spectrum $U_{ij}(f)$ at a site i for an event j as:

$$U_{ij}(f) = S_j(f)P(r_{ij}f)H_i(f) \tag{1}$$

where $S_j(f)$ is the j–th event source spectrum, $H_i(f)$ is the i–th site response function, $P(r,f)$ is the source–to–site path function, and r_{ij} is the source–to–site hypocentral distance. We applied GIT following the one–step generalized inversion approach (Oth et al. [33]) implemented in an improved version of the GITANES (GIT ANalysis of Earthquake Spectra) package (Klin et al. [34,35]). To obtain a unique solution for the overdetermined linear system, we added additional constraints relative to the propagation terms and the site terms. The constraints relative to propagation consist of defining a reference distance where the path term is 0 and imposing smoothing variation between contiguous distance intervals (Castro et al. [36]). The other constraint is the specification of a reference station, which we assumed to be free of site effects. Of the considered sites, station TN02, located on the calcarenites of the Calcare di Torbole Formation (Carta Geologica d'Italia, foglio Riva del Garda 080 [21] was chosen as the reference station because of its flat NHV response. An ad-hoc Multichannel Analysis of the Surface Waves at this site confirmed that it belongs to a Class A site (according to the Italian Building Code NTC08) based on the VS30 velocity of 920 m/s (Garbin et al. [3]).

The analysis of the ground motion DUration Lengthening (hereafter DUL) resulted from the observation of large ground motion duration in the earthquakes recorded at the temporary stations in the center of the valley compared to those observed at the rock formations on the valley margins. Fig. 4 shows, as an example, the North-South (NS) components of the recordings of the 08/08/2019 M=3.6 earthquake, which occurred at about 20 km away from the temporary stations. Qualitatively, sites located on rocks (i.e., TN02, TN06, and TN07) have very low amplitude and duration. In contrast, significant amplifications are observed at sites on valley sediments both west and east of M. Brione, with differences in amplitudes and duration that, considering the different width of the valleys and thickness of the sedimentary cover, is likely to be related to the different frequency content of the waveforms.

Over the past 20 years, a number of authors have addressed the issue of estimating and quantifying duration lengthening on sites that are susceptible to site effects. Beauval et al. [17] used a method proposed by Sawada [37] to quantify frequency-dependent lengthening based on signal phase and found consistent signal duration at a test site in the Volvi area. The increase in signal duration is typically associated with the combined influence of basin-edge effects and soil layering (Semblat et al. [38]). Klin et al. [39] explained the unexpectedly long duration of the ground motion observed in the epicentral area of the 2012 Po Plain
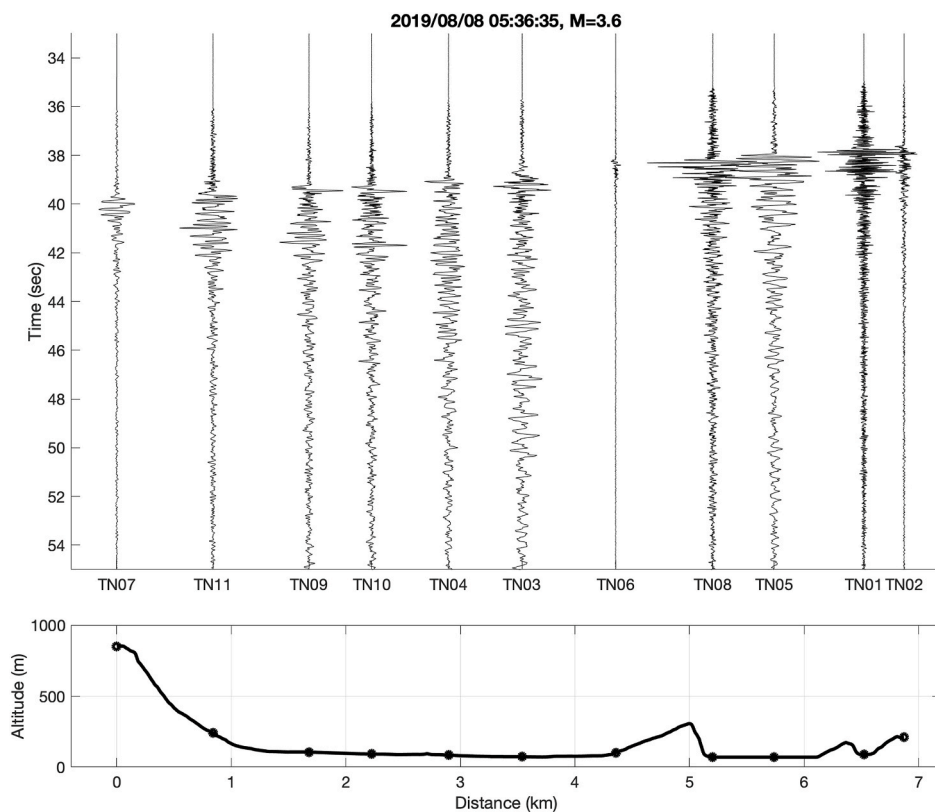
Fig. 4. Seismograms recorded of the Vallarsa (TN) event of 08/08/2019 M = 3.6, at the transect T1 stations (horizontal NS component). The distance to the nearest station is 19.5 km. The elevation profile along the T1 transect is shown in the bottom panel.

seismic sequence with the excitation of surface waves on a buried structural ridge. Abraham et al. [40] found pronounced differences between recordings from the basin and adjacent sites at outcropping bedrock, related to both lithostratigraphic amplification as well as to the 2-D/3-D geometry of the basin. The methodology applied in this work relies on the JASSSR (Parolai and Bard [19]), which combines the analysis of the RSSR technique with a sonogram approach and allows highlighting the differences in frequency lengthening between a signal recorded at a given site and the one recorded at a reference site. The adopted procedure can be summarized as follows:

- time recordings of the same event at the study site and at the reference site (Fig. 5, panel a and b, respectively) are selected and the time–frequency distribution is computed by the S-Transform; the window length corresponds to the time between the arrival of the P-wave and the return to the same noise level as before the event at the study site;
- after denoising (Parolai [41]), the enhanced spectrogram at the study site $C(f,t)$ and at the reference site $B(f,t)$ (Fig. 5, panel c and d, respectively) are obtained by applying a suitable threshold (Parolai and Bard [19]), considering the pre-event noise level of the recordings;
- differential matrix $E(f,t)$ between the two enhanced spectrograms is computed (Fig. 5, panel e);
- duration lengthening (Fig. 5, panel f) is obtained by summing the contribution of $E(f,t)$ that exceeds a certain threshold (i.e., the absolute average of $E(f,t)$).

In addition to the temporary network of 19 seismological stations, we collected recordings of seismic noise of about 30–45 min in length at 110 sites (Fig. 2). Measurements were made using Lennartz LE-3D/1s and Lennartz LE-3D/5s sensors and Reftek RT130 digitizers. The analysis of the seismic noise data was carried out using the "Nakamura"

method (Nogoshi and Igarashi [42]; Nakamura [43]; Bard [44]; Field and Jacob [45]) and consists in the computation of the spectral ratio between the horizontal and vertical components of the seismic noise, and in the selection of the site resonance frequency f_0. We estimated the NHV curve by averaging the H/V spectral ratios computed on a set of running time windows. Outliers (e.g., near impulsive sources caused by cars) were manually removed. We used a window length of 60 s with a 10% overlap to ensure that at least 20 time windows were available at each station. For the generic time window, the signal was processed as follows: Fast Fourier Transform FFT (including tapering), Konno & Ohmachi [46] smoothing, merging of two horizontal components by computing the geometric mean, H/V spectral ratio for each window, estimation of mean and standard deviation of H/V ratios.

4. Results of the evaluation of site response

Fig. 6 shows the spatial distribution of f_0 extracted from the NHV for the 110 sites. The map helps to identify areas with similar resonance characteristics, in particular a central area northeast of Riva del Garda, characterized by very low f_0 below 1 Hz, and another area east of M. Brione, characterized by f_0 values between 1 and 5 Hz. In contrast, sites on the valley margins or on a slope tend to show high resonance frequencies, even up to 10 Hz, or in some cases a flat curve, indicating thinner or absent sediment cover, as expected coming from the valley floor to the surrounding slopes.

Fig. 7 compares the spectral ratios estimated by GIT for the horizontal and vertical components (black and blue curves) with respect to the reference site TN02 with those obtained by the NHV method (red curve) and by the DUL analysis (gray dashed curve). In general, there is excellent match between the curves for frequencies equal to or lower f_0. Maximum amplification values (close to 10) are reached for the sites located in the center of the wider western valley (i.e., sites TN03, TN04, TN12 TN13 and TN14) and in the southern part of the current Sarca
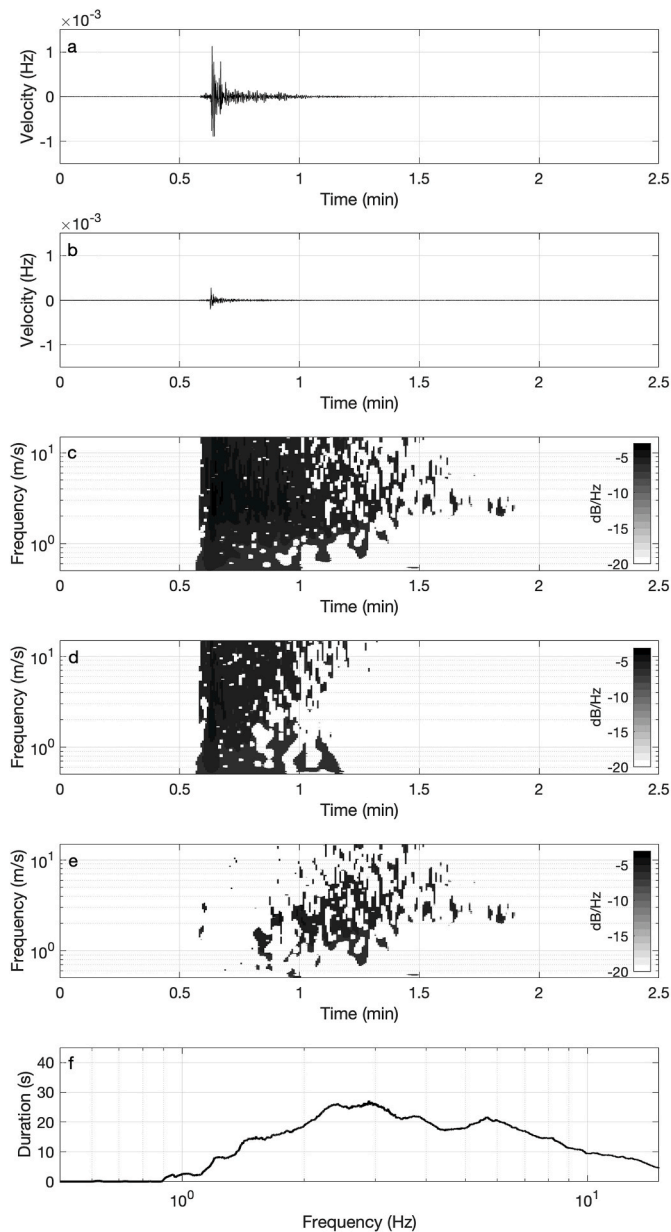
Fig. 5. Seismograms (horizontal EW component) recorded of the Vallarsa (TN) event of 08/08/2019 M = 3.6, at the TN05 site (a) and at the reference site TN02 (b). Spectrograms at the TN05 (c) and TN02 (d) site and differential matrix (e). Lengthening of the duration of the ground motion at site TN05 compared to reference site TN02 (f).

River Valley (i.e., TN05 and TN08).

Comparison of the DUL curves (gray dashed lines) with the corresponding NHV and GIT curves shows that at f_0 or, more generally, at frequencies characterized by a certain level of amplification (i.e., above 2), there is a duration that in some places reaches tens of seconds of lengthening compared to that of the reference site. Maximum values (close to 20 s) are again reached in the middle of the broad western valley (sites TN04, TN12).

5. Zonation of site response

The procedure proposed by Strollo et al. [11] and Ullah et al. [14] was used to improve the spatial resolution of the GIT and DUL curves obtained at the 19 sites of the temporary network over the sites of the single stations noise measurements used for the NHV calculation. In this
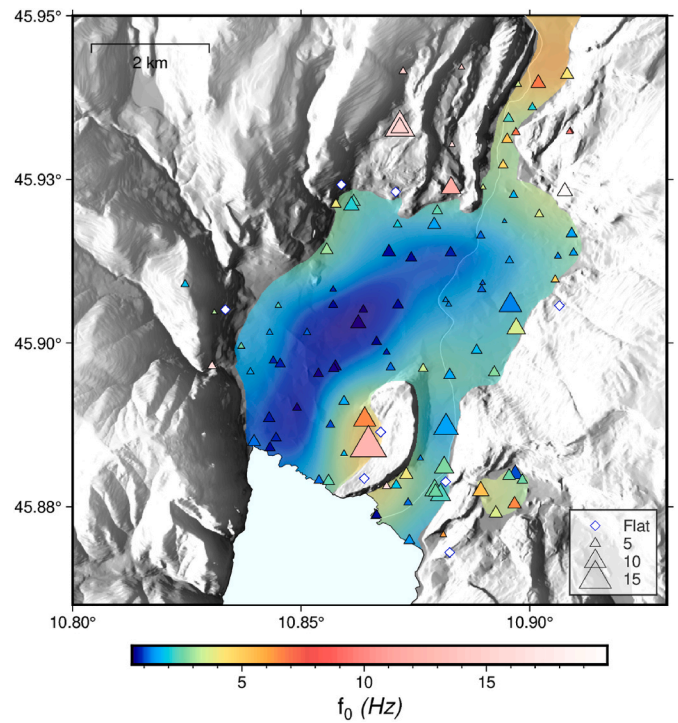

Fig. 6. Map of the resonance frequencies f_0 obtained from the NHV analysis. The color and size of the triangles indicate f_0 and corresponding amplitude of NHV curve at the sites. The values of f_0 were interpolated using the "Natural Neighbor Interpolation" method. White diamonds indicate flat or nearly flat curve.

analysis, a frequency range of 0.5–6 Hz was considered, i.e., the range in which almost all f_0 fall. Above 6 Hz, the presence of secondary peaks, due to differences in the most superficial structure, could in some cases lead to a worse assignment of the curves to the clusters. In addition, by applying a Voronoi tessellation, we were able to create a zonation map of the area. The method is carried out in three steps, which are described in detail in the following sections.

5.1. Validation of the NHV proxy by cluster analysis

The first step is to compare the results of a clustering procedure applied separately to the GIT, the DUL and the NHV curves, at only the locations of the seismological stations. Good overlap of the clusters argues for using the NHV curves as an indicator (proxy) for the other two parameters, i.e., GIT/DUL curves. To do this, we used the k-means clustering algorithm, which divides the elements (in this case, either GIT, DUL, or NHV curves) into non-overlapping clusters, minimizing the dispersion within each cluster and maximizing the dispersion between clusters. Each cluster is assigned a centroid corresponding to the average of the curves belonging to the cluster, and the dissimilarity between a curve (i.e., the NHV/GIT/DUL curve) and the centroid of each cluster is evaluated using the square of the Euclidean distance:

$$DC = \sum_{k=1}^{K} \sum_{n=1}^{N} \sum_{m=1}^{M} \left\| x_{nm,k} - c_k \right\|^2 \tag{2}$$

where M is the number of objects (in our case, the number of frequencies in the NHV/GIT/DUL curves), N is the number of elements (or curves) of the k-th cluster, K is the number of clusters, c_k is the centroid of cluster k, and $x_{nm,k}$ is the value of object m in cluster k of curve n (the NHV/GIT/DUL curve).

To evaluate the optimal number of clusters K, we used both the CH index of Calinski and Harabasz [47] and the Silhouette index (Rousseeuw [48]).
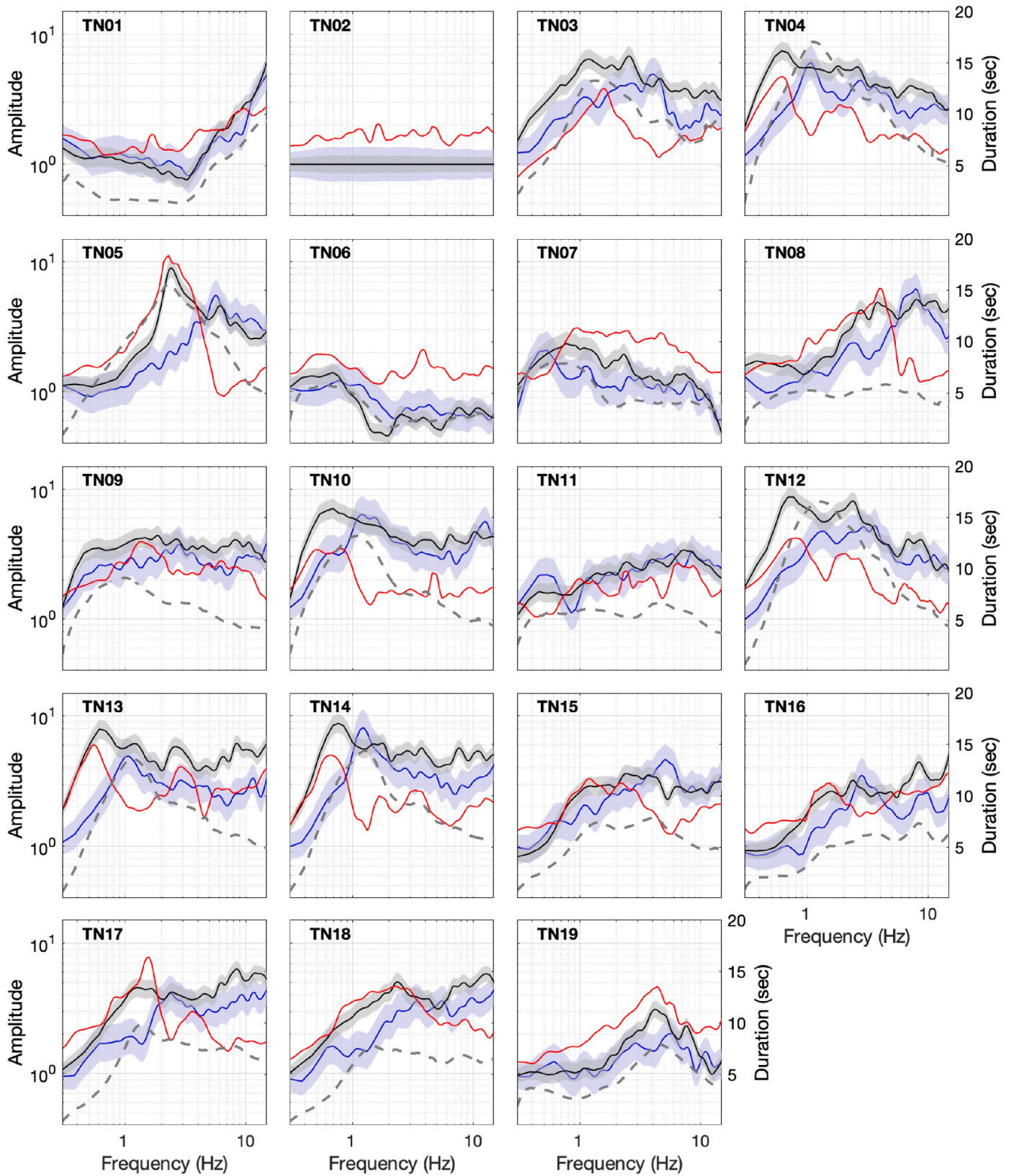
Fig. 7. Site response curves estimated for the 19 sites. Black and blue lines: GIT curves with respect to reference site TN02, geometric mean of horizontal components and vertical components, respectively (the 1st standard deviation is shown as transparent filling in the same colors); red lines: NHV curves, gray dashed lines: DUL curves.
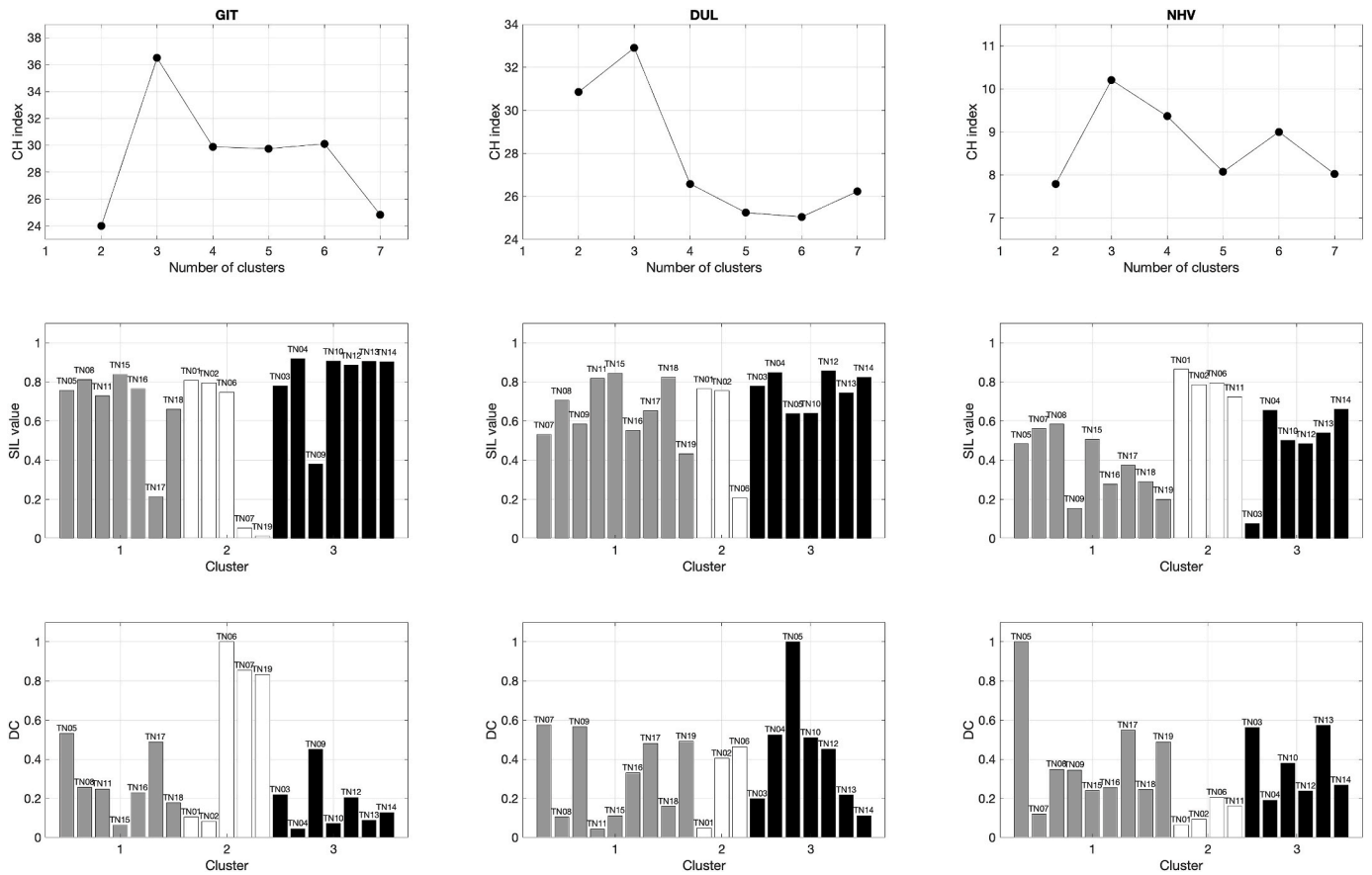
Fig. 8. CH index in terms of the number of clusters (top panels), SIL (middle panels) and DC (bottom panels) calculated for each curve for GIT amplification (left), DUL (middle) and NHV (right) at the 19 recording sites of the temporary seismic network. The colors refer to the three clusters.

The CH index is defined as:

$$CH = \frac{\sum_{k=1}^{K} N \left\| c_k - c \right\|^2}{K-1} \Bigg/ \frac{\sum_{k=1}^{K} \sum_{n=1}^{N} \left\| x_n - c_k \right\|^2}{P - K} \tag{3}$$

where c is the global centroid, x_n is the *n-th* of N curve in the *k-th* cluster and P is the total number of curves. The best number of clusters is the one that maximizes CH.

The Silhouette index (hereafter SIL) is instead defined as:

$$SIL(n) = \frac{b(n) - a(n)}{max\left(a(n), b(n)\right)} \tag{4}$$

where $a(n)$ is the mean distance (dissimilarity) of curve n to all other objects in the same cluster and $b(n)$ is the smallest mean distance of curve n to all other curves in any other cluster. The value of *SIL* varies from −1 to +1, where +1 means that the curve is well-represented in the cluster, while −1 represents poor classification.

The CH index, obtained as a function of the cluster's number separately for the three sets of curves is shown in the upper panels of Fig. 8. As can be seen, CH always shows the same optimal number of clusters, which is always equal to three regardless of the parameter considered (GIT/DUL/NHV). In the same figure, the middle and the bottom panels show the SIL and the distance DC of each curve from the centroid of the cluster to which it belongs.

Figs. 9 and 10 show the results of the cluster analysis. It can be seen that the three identified clusters group the results of nearly the same stations as follows:

– white cluster (i.e., CLU2 in Fig. 10, panels a-c), which includes sites on rocks or valley edges characterized by flat response or low amplification values and very low (1–2 s) duration lengthening (Fig. 9);
– black cluster (i.e. CLU3 in Fig. 10, panels a-c), which includes sites in the central part of the western wider valley and is characterized by high amplification values and duration lengthening of the order of 10–15 s at low frequencies. Note that the maximum amplification values are reached at frequency values below 1 Hz, while the curves are wider when duration lengthening is considered, reaching the maximum between 1 and 2 Hz. This discrepancy could be due to frequency differences associated with 1-D and 2-D/3-D lateral propagation effects, as already observed by Bindi et al. [49] in the Gubbio basin;
– gray cluster (i.e., CLU1 in Fig. 10, panels a-c), which includes the sites in the eastern valley and on the northern edge and is characterized by maximum values at higher frequencies (between 1 and 10 Hz).

Comparing GIT and NHV (panels a and c in Fig. 10, respectively), the clustering for most sites is identical; the exception is 4 sites, all located on the western edge of the valley, namely sites TN07, TN19, TN09 and TN11. The first two sites (TN07 and TN19) belong to the white group by the analysis of GIT curves (Fig. 10, panel a) and to the gray group when the NHVs are considered (Fig. 10, panel c), while TN09 belongs to the black GIT cluster and gray NHV one. Looking at the values of SIL and DC separately (the two lower panels in Fig. 8), we can see that the GIT curves for the three sites also have very low SIL values and high DC values (indicating a poor representation in the current cluster). For this reason, we decided to combine the three sites in the gray cluster.
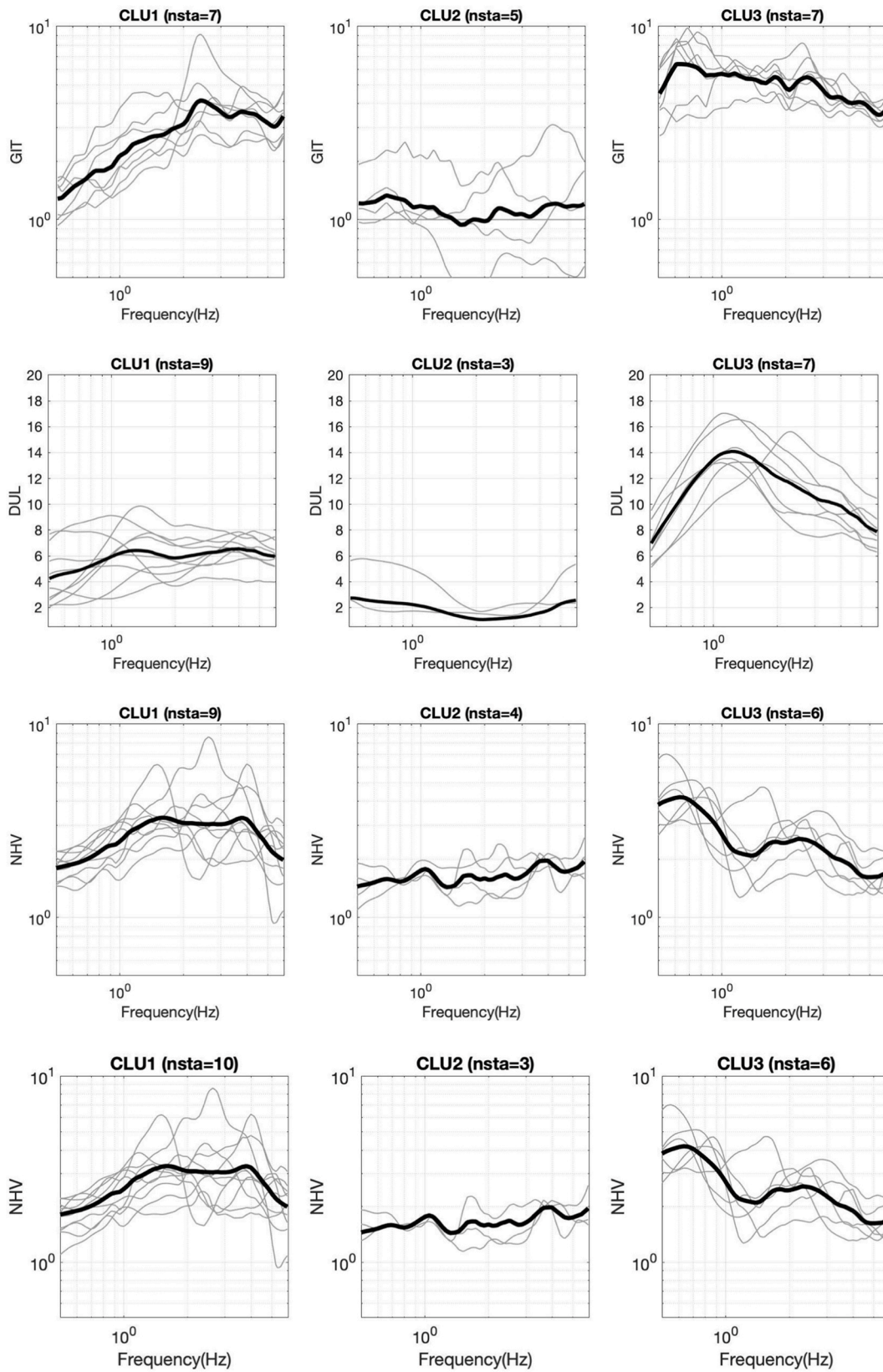
Fig. 9. From top to bottom: GIT, DUL and NHV curves (gray lines) divided into the three clusters and the relative centroid (bold line). CLU1, CLU2 and CLU3 refers to the gray, white and black clusters, respectively. The bottom panel shows the final cluster of NHV curves whose corresponding sites are shown in the right panel of Fig. 10. The individual curves associated with the 19 sites and the centroids deviation are respectively represented by thin gray and thick black lines.
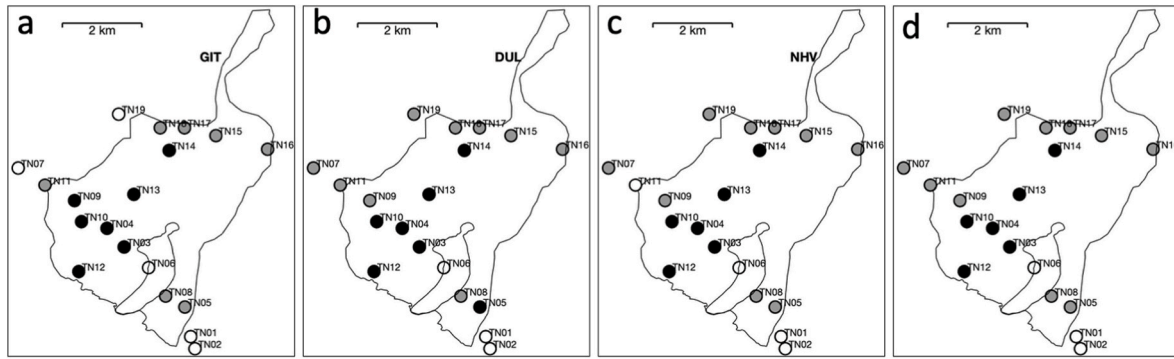
Fig. 10. From left to right: maps showing the spatial distribution of the 19 sites and their respective clusters obtained for GIT (a), DUL (b), NHV (c) curves at the 19 seismological stations. Panel d displays the final configuration (see text). Gray, white and black circles indicate cluster CLU1, CLU2 and CLU3 respectively.

Regarding TN11, the spectral amplification curve could be assigned to both gray GIT (Fig. 10, panel a) and white NHV (Fig. 10, panel c) clusters based on the SIL and DC values. Both NHV and GIT curves are characterized by very low amplitude throughout the frequency range considered. Based on geographic/geological considerations (the site is located at the edge of the valley, on deposits), we also assigned this site to the gray cluster, so that the final configuration is shown in Fig. 10, in panel d.

Finally, we compared the DUL (Fig. 10, panel b) with the just modified (NHV) cluster (Fig. 10, panel d). Only site TN05 differs since it is located in the black DUL and in the gray NHV clusters. The TN05 curve is however characterized by a maximum DC both in the DUL and NHV clusters (Fig. 8), suggesting it is a site straddling the two. Since the DUL curve is characterized by a pronounced peak at about 2.2 Hz (Fig. 7), we decided to place this site in the gray cluster.

The results of the cluster analysis demonstrated that GIT, DUL and NHV curves, although they might look quite different, showed a very similar spatial distribution. The same result was also obtained by Strollo et al. [11] and Ullah et al. [14], which showed that sites having similar RSSR also show similar noise NHV. In this work, we obtained the same result also with GIT and DUL curves.

5.2. Correlation analysis of NHV curves and zonation of the GIT/DUL curves

Once verified that the sites having similar GIT or DUL curves also have similar NHV curves, we can use this similarity in order to assign the spectral amplification/duration lengthening curves to the sites of the single-station seismic noise measurements. The latter were associated with the 19 sites of the seismological stations characterized by the closest NHV curve based on the value given by:

$$CC = \rho \cdot g \tag{5}$$

where ρ is the Pearson cross-correlation coefficient (Davis [50]):

$$\rho_{jk} = \frac{\frac{\sum_{i=1}^{n}\left(x_{ij}-\overline{x}_j\right)(x_{ik}-\overline{x}_k)}{n-1}}{\sqrt{\frac{\sum_{i=1}^{n}\left(x_{ij}-\overline{x}_j\right)^2}{n-1}}\sqrt{\frac{\sum_{i=1}^{n}(x_{ik}-\overline{x}_k)^2}{n-1}}} \tag{6}$$

where j is an index going from 1 to the number of single-station seismic noise measurements sites, k is an index going from 1 to the number of seismological stations, i is an index for frequency, x is the NHV amplitude at frequency i, and $\overline{x}$ is the mean of x. The values of the Pearson coefficient are bounded between −1 and 1 and do not consider the differences in amplitude, but only the similarity in shape. For this reason, the goodness-of-fit coefficient g is also introduced (inspired from Krišteková et al. [51]):

$$g_{jk} = exp\left(-\sqrt{\frac{\sum_{i=1}^{n}\left(x_{ij}-x_{ik}\right)^2}{n}}\right) \tag{7}$$

The goodness-of-fit coefficient ranges from zero to one for increasing levels of agreement between the curves.

In Fig. 11, each panel shows the comparison between the NHV curve at the seismic noise measurement sites (black curve) and the NHV curve of the temporary seismological station (red curve) with the highest value of CC. Of the original number of NHV measurements (i.e., 110), 19 at seismological stations and an additional 8 sites where the NHV curves had high uncertainty indicated by large values of the standard deviations, were excluded from the correlation analysis, resulting in a total number of 83. Based on this association, each of the 83 sites was assigned to one of the three previously identified clusters, that is identified by its centroid.

5.3. Application of the Voronoi tessellation

As a result of the correlation analysis, the GIT amplification functions and the DUL curves were extended to 83 sites in addition to the 19 seismological stations. Nevertheless, in the framework of land-use planning it is important to circumscribe the areas characterized by the values of a certain parameter, such as the spectral amplification.

To this end, we applied a Voronoi tessellation (Okabe et al. [52]) to the study area. Given a set of points, each location in the Euclidean plane is assigned to the closest member of the point set; the set of locations associated with each member forms a region and the obtained set of regions, collectively exhaustive and mutually exclusive except for the boundaries, forms a Voronoi tessellation. Fig. 12 (panel a) shows the application of the Voronoi tessellation on the result of the correlation analysis given the set of NHV measurement sites. Note that the considered area has been enclosed by the contour of the lower Sarca Valley, in order to exclude the outer and more isolated sites, for which there is no spatial contiguity between the points. Moreover, most sites outside the valley belong to other morphological contexts (hills, slope valleys), and their site response may differ significantly from that along the valley, e. g., influenced by topographic effects.

As can be seen, the sites belonging to the white cluster (flat response or low amplifications and duration) are located on M. Brione. The black cluster (characterized by curves with spectral amplification and duration lengthening in a lower frequency band) shows a spatial distribution both along the axis of the main western valley and along the lakeshore, from Riva del Garda to Nago-Torbole. The gray cluster, characterized by peaks at higher frequencies, extends to the sides of the black cluster and along the entire current valley of the Sarca River to the north. Some isolated points, located on the edges of the valley and east of M. Brione,
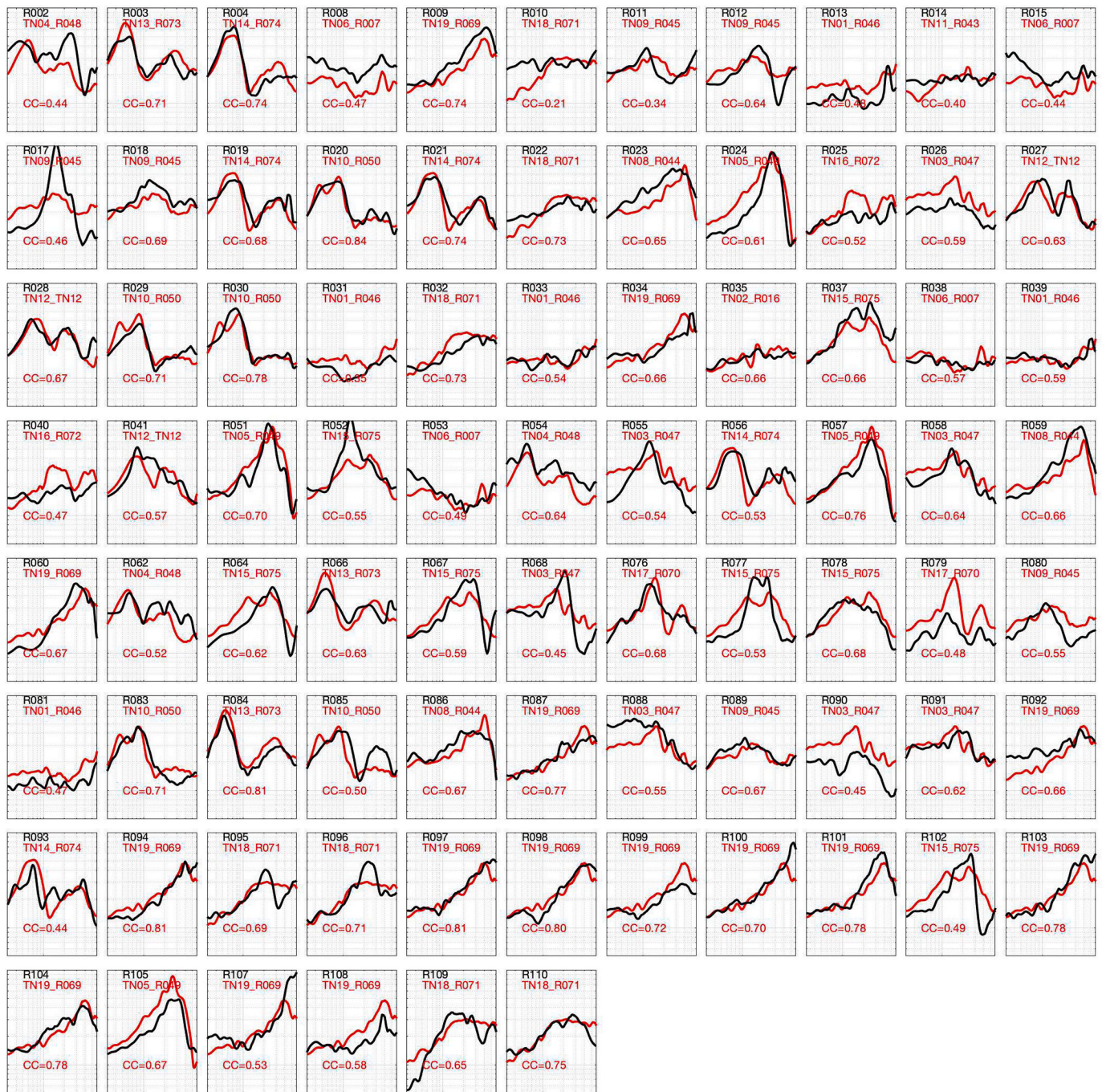
Fig. 11. Results of correlation analysis. For each of the NHV curves obtained at the 83 noise-recording sites (in black), the best-correlated (i.e. with maximum CC) NHV curve obtained at the 19 sites of seismometric network (in red), is given. The values of CC are reported at the bottom of each panel.

were studied individually and assigned to the same cluster as the surrounding points, based on geometrical (distance from the cluster to which they belong) and geological considerations.

The final result is shown in the zonation map in Fig. 12 (panel b). The curves of amplification and duration lengthening for the three identified zones (corresponding to the centroid of their membership cluster) are shown in Fig. 13.

6. Conclusions and applications of the results

The lower Sarca Valley, for which experimental observations indicate that surface ground motion is characterized by site effects (Faccioli et al. [8]; Garbin et al. [3] and Parolai et al. [4]), represents an

emblematic case study where it is essential to evaluate the site response with a high resolution in order to correctly assess its seismic hazard and risk level. Regional maps, in fact, are not suitable due to too low resolution.

In this study, the site response of the lower Sarca Valley was firstly evaluated at a limited number of sites corresponding to the location of earthquake recordings. Then the two quantities representing site response, i.e., spectral amplification curves and lengthening of ground motion duration, were spatially extended to the entire valley, where recordings of earthquakes were not available, by the application of the k-means algorithm for clustering analysis and a correlation analysis, using the NHV recorded both at the seismological stations and those evaluated throughout the valley by single-station noise measurements.
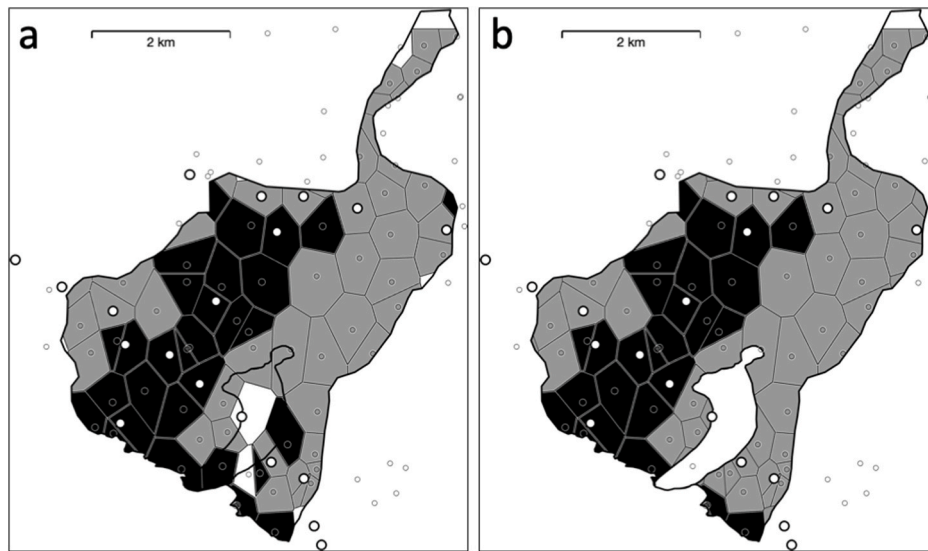
Fig. 12. Results of the clustering correlation analysis: spatial extent of the clusters obtained at the 19 sites (white circles) over the entire study area - the area covered by the noise measurement points (small circles). Panels (a) and (b) show the direct result of the analysis and the final zoning, respectively (see text).
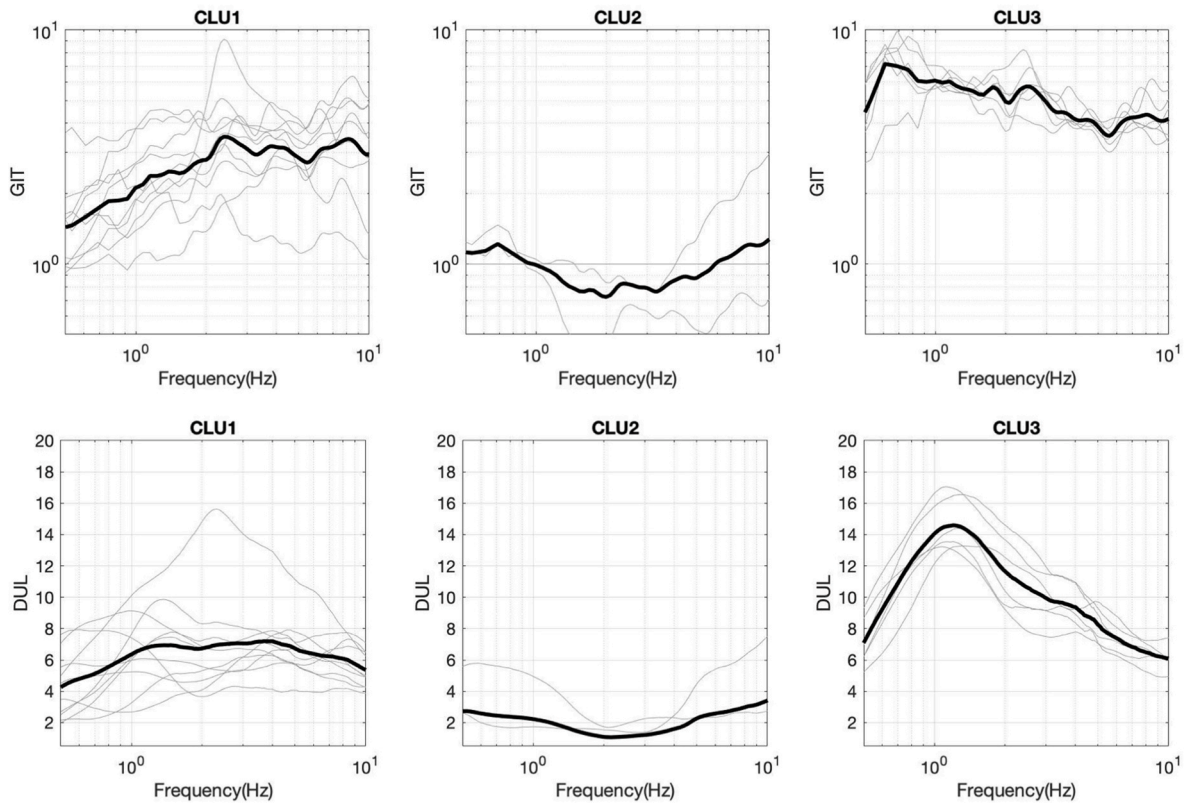

Fig. 13. GIT and DUL curves (upper and lower panels, respectively) associated with the three clusters. The individual curves for the 19 sites and the centroids of each cluster are shown by thin gray and thick black lines, respectively.

Moreover, by the application of a Voronoi tessellation, we were finally able to subdivide the area into three zones with different site response characteristics:

- the area of M. Brione, characterized by a flat response;
- the area around the axis of the main valley where the sediments reach a thickness of about 420 m and the lakeshore area, characterized by high amplification values (almost up to 10) and ground

motion lengthening of several seconds at low frequencies (about 0.7 Hz and 1 Hz, respectively);
- the area at the edges of the main valley and the area that includes the current valley of the Sarca River to the north, characterized by lengthening of several seconds correlated with spectral amplification at frequencies higher than 1 Hz (between 2 and 5 Hz, reaching 10 Hz along the strait north of Arco).

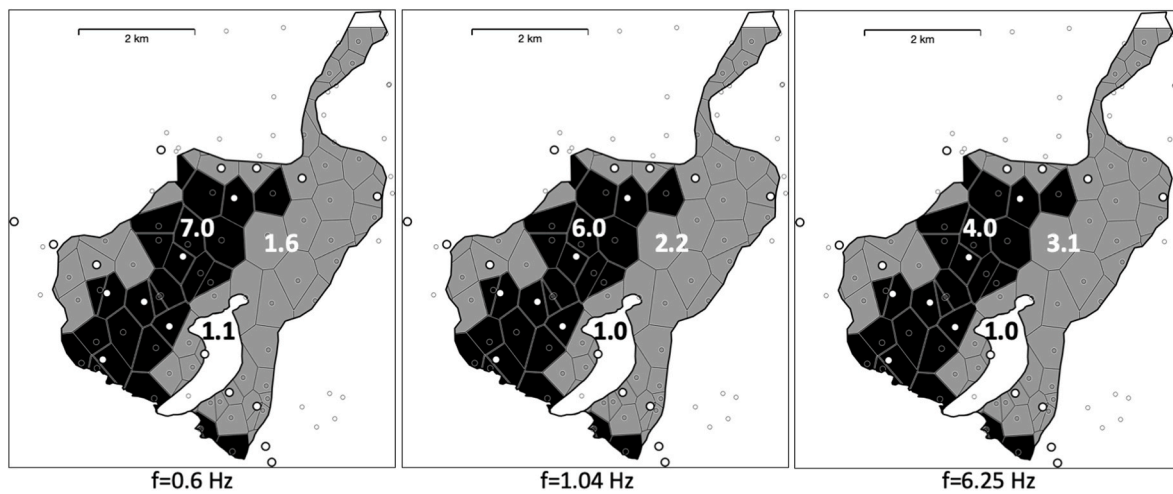Given the GIT amplification and duration lengthening curves

Fig. 14. Maps of amplification values calculated for the three clusters at f=0.6 Hz (left), 1.04 Hz (center) and 6.25 Hz (right).
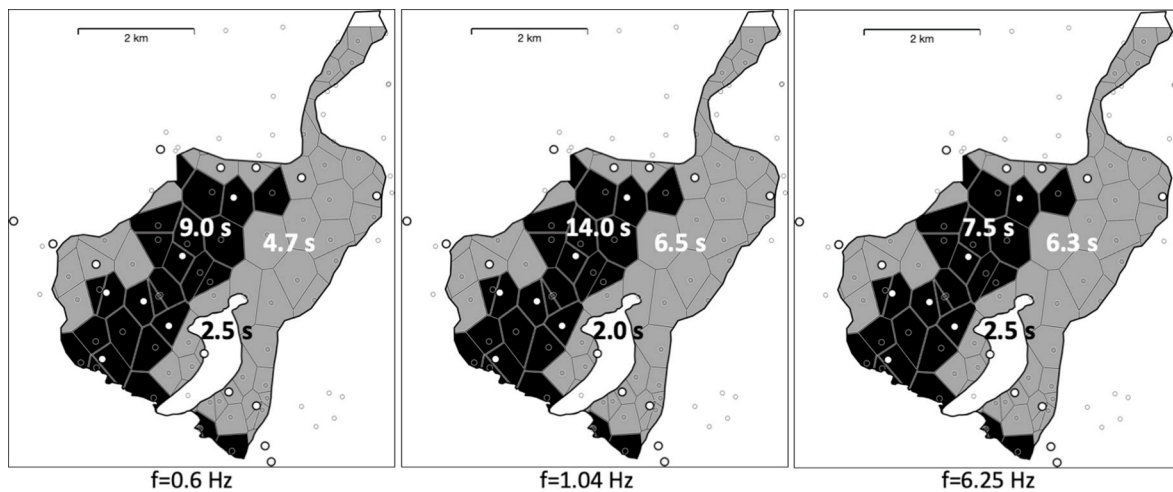

Fig. 15. Maps of duration lengthening values calculated for the three clusters at f=0.6 Hz (left), 1.04 Hz (center) and 6.25 Hz (right).

corresponding to the centroid associated with each zone (bold black curves in Fig. 13), maps of amplification factors and duration lengthening values can be produced. For example, Figs. 14 and 15 show the amplification factors and duration lengthening for each zone at three selected frequencies representing low, medium, and high frequencies, i. e., f=0.6 Hz, f=1.04 Hz, and f=6.25 Hz, the same as those used by Bindi et al. [6] in the maps of site amplification values at regional scale.

Note that the estimated values are obtained after smoothing the spectra with Konno and Ohmachi [46], i.e. the local amplitude values are the result of a weighted average over a fixed (in logarithmic scale) frequency band. The white zone shows no amplification (values around 1 for the three frequency values), the gray zone displays amplification values between 1.6 and 3.1 for increasing frequencies, while the innermost zone of the main valley (black cluster) is characterized by very high values, reaching the maximum of 7.0 at f=0.6 Hz for the amplification and of 14 s at f=1.04 for duration lengthening. As stated before, this difference can be due to the contribution of 2-D/3-D effects in the valley site response (Bindi et al. [49]).

It is worth mentioning that the site response obtained in our study area varies on a shorter spatial scale (tens/hundreds of meters) than the one proposed by the current regional site amplification models at large regional spatial scale (kilometers to tens of kilometers), therefore it actually represents a valid tool to integrate the information given by the latter. Nevertheless, we hope that this can be useful for readers

interested in a first order comparison of the results between regional and local studies (although being aware of the obvious differences that can be expected).

In summary, the approach we proposed has proven successful in determining the spatial variation of seismic response in a complex context such as an alpine valley. While microzonation studies generally refer to an urban context at the level of a single municipality (Moscatelli et al. [53]), this study, which considers three municipalities of the Provincia Autonoma di Trento, has the advantage of providing an harmonized characterization of the site response over a larger area and avoids the problems of reconciling results along common boundaries between different municipalities. Therefore, while still being consistent with the usual requirements of the Italian approach, it also offers the opportunity of avoiding inconsistencies between different municipalities. By defining a hierarchy between areas of an urban center in terms of seismic motion amplification (i.e.: prioritizing actions to improve seismic safety in populated areas; avoiding double resonance phenomena in new designs; locating strategic buildings in the safest areas …), this study also provides important and effective data for those involved in land-use planning. However, we note that this study is only a preparatory (but fundamental) step for the third-level microzonation study, which is a multidisciplinary study that requires more detailed geological and geophysical investigations (Moscatelli et al. [53]). It should also be kept in mind that the proposed method requires the availability of a

sufficient number of earthquake recordings and therefore cannot be considered a fast method, especially in areas of low to moderate seismicity.

Author statement

Giovanna Laurenzano: Project administration, Conceptualization, Supervision, Investigation, Methodology, Data Analysis, Writing.

Marco Garbin: Data Curation, Investigation, Methodology, Writing.

Stefano Parolai: Conceptualization, Supervision, Methodology, Writing.

Carla Barnaba: Geophysical data curation and analysis, Writing.

Marco Romanelli: Data acquisition.

Luca Froner: Data acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

References

[1] Kaklamanos J, Cabas A, Parolai S, Guéguen P. Introduction to the special section on advances in site response estimation. Bull Seismol Soc Am 2021;111(4):1665–76. https://doi.org/10.1785/0120210152.

[2] Faccioli E, Vanini M, Frassine L. Complex site effects in earthquake ground motion, including topography. In: 12th European conference on earthquake engineering. London, United Kingdom: Barbican Center; 2002. 9–13 September.

[3] Garbin M., Laurenzano G., Parolai S. Studio riguardante la risposta sismica locale del tratto terminale della valle del fiume Sarca in prossimità del Lago di Garda. Relazione finale attività - 2019/2022, Relazione 2022/98 Sez. CRS 17, 1 agosto 2022. In Italian.

[4] Parolai S, Laurenzano G, Garbin M. Estimating cross-coupling in site response by seismic noise interferometry: an example from an alpine valley (northeastern Italy). Seismol Res Lett 2022;94(1):140–8. https://doi.org/10.1785/0220220099.

[5] Vessia G, Laurenzano G, Pagliaroli A, Pilz M. Seismic site response estimation for microzonation studies promoting the resilience of urban centers. Eng Geol 2021; 284:106031. https://doi.org/10.1016/j.enggeo.2021.106031.

[6] Bindi D, Zaccarelli R, Razafindrakoto H, Yen MH, Cotton F. Empirical shaking scenarios for Europe: a feasibility study. Geophys J Int 2023;232(2):990–1005. https://doi.org/10.1093/gji/ggac382.

[7] Weatherill G, Crowley H, Roullé A, et al. Modelling site response at regional scale for the 2020 European Seismic Risk Model (ESRM20). Bull Earthq Eng 2022;21: 665–714. https://doi.org/10.1007/s10518-022-01526-5.

[8] Faccioli E, Vanini M. Complex seismic site effects in sediment-filled valleys and implications on design spectra. Earthq Eng Struct Dynam 2003;5(4):223–238..

[9] Bragato PL, Laurenzano G, Barnaba C. Automatic zonation of urban areas based on the similarity of H/V spectral ratios. Bull Seismol Soc Am 2007;97(5):1404–12. https://doi.org/10.1785/0120060245.

[10] Tselentis GA, Paraskevopoulos P. 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 2011;101:579–95. https://doi.org/10.1785/0120090355.

[11] Strollo A, Parolai S, Bindi D, Chiauzzi L, Pagliuca R, Mucciarelli M, et al. Microzonation of Potenza (Southern Italy) in terms of spectral intensity ratio using joint analysis of earthquakes and ambient noise. Bull Earthq Eng 2012;10(2): 493–516. https://doi.org/10.1007/s10518-011-9256-4.

[12] Borcherdt RD. Effects of local geology on ground motion near San Francisco Bay. Bull Seismol Soc Am 1970;60:29–61.

[13] Field EH, Jacob KHA. A comparison and test of various site-response estimation techniques, including three that are not reference-site dependent. Bull Seismol Soc Am 1995;85(4):1127–43.

[14] Ullah S, Bindi D, Pittore M, Pilz M, Orunbaev S, Moldobekov B, et al. Improving the spatial resolution of ground motion variability using earth- quake and seismic noise data: the example of Bishkek (Kyrgyzstan). Bull Earthq Eng 2013;11(2):385–99. https://doi.org/10.1007/s10518-012-9401-8.

[15] Paolucci E, Lunedei E, Albarello D. Application of the principal component analysis (PCA) to HVSR data aimed at the seismic characterization of earthquake prone areas. Geophys J Int 2017;211(1):650–62. https://doi.org/10.1093/gji/ggx325.

[16] Capizzi P, Martorana R. Analysis of HVSR data using a modified centroid-based algorithm for near-surface geological reconstruction. Geosciences 2022;12(4):147. https://doi.org/10.3390/geosciences12040147.

[17] Beauval C, Bard PY, Moczo P, Kristek J. Quantification of frequency-dependent lengthening of seismic ground-motion duration due to local geology: applications to the Volvi area (Greece). Bull Seismol Soc Am 2003;93(1):371–85. https://doi.org/10.1785/0120010255.

[18] Hancock J, Bommer JJ. A state-of-knowledge review of the influence of strong-motion duration on structural damage. Earthq Spectra 2006;22(3):827–45. https://doi.org/10.1193/1.2220576.

[19] Parolai S, Bard PY. Evaluation of site effects by means of joint analysis of sonogram and standard spectral ratio (JASSSR). J Seismol 2003;7. https://doi.org/10.1023/B:JOSE.0000005723.52199.0e. 492–479.

[20] Andrews DJ. Objective determination of source parameters and similarity of earthquakes of different size. In: Earthquakes source mechanics. American Geophysical Union; 1986. p. 259–67. https://doi.org/10.1029/GM037p0259.

[21] Castellarin A, Picotti V, Cantelli L, Claps M, Trombetta GL, Selli L, et al. Note illustrative della carta geologica d'Italia alla scala 1:50.000, foglio 080 Riva del Garda. Provincia Autonoma di Trento. L.A.C., Firenze. 2005.

[22] Castellarin A, Vai GB, Cantelli L. The alpine evolution of the Southern Alps around the Giudicarie faults: a Late Cretaceous to Early Eocene transfer zone. Tectonophysics 2006;414:203–23. https://doi.org/10.1016/j.tecto.2005.10.019.

[23] Luciani V. Stratigrafia sequenziale del terziario nella catena del Monte Baldo (Province di Verona e Trento). Mem Sci Geol 1989;41:263–351 [Padova].

[24] Felber M, Veronese L, Cocco S, Frei W, Nardin M, Oppizzi P, et al. Indagini sismiche e geognostiche nelle valli del Trentino meridionale (Val d'Adige, Valsugana, Valle del Sarca, Valle del Chiese), Italia. Studi trentini di Scienze naturali - acta Geologica 2000;75:3–52.

[25] Carulli GB, Slejko D. Seismotectonic characteristics of the Italian central Alps and implications for the seismic hazard. Boll Soc Geol Ital 2009;128(1):201–15.

[26] Viganò A, Scafidi D, Ranalli G, Martin S, Della Vedova B, Spallarossa D. Earthquake relocations, crustal rheology, and active deformation in the central-eastern Alps (N Italy). Tectonophysics 2015;661:81–98. https://doi.org/10.1016/j.tecto.2015.08.017.

[27] Locati M, Camassi R, Rovida A, Ercolani E, Bernardini F, Castelli V, Caracciolo CH, Tertulliani A, Rossi A, Azzaro R, D'Amico S, Antonucci A. Database Macrosismico Italiano (DBMI15), versione 4.0. 2022, Istituto Nazionale di Geofisica e Vulcanologia (INGV). https://doi.org/10.13127/DBMI/DBMI15.4.

[28] Guidoboni E, Ferrari G, Tarabusi G, Sgattoni G, Comastri A, Mariotti D, et al. CFTI5Med, the new release of the catalogue of strong earthquakes in Italy and in the mediterranean area. Sci Data 2019;6:1–15. https://doi.org/10.1038/s41597-019-0091-9.

[29] Martin S, Ivy-Ochs S, Viganò A, Campedel P, Rigo M, Vockenhuber C, Gabrieli F, Mair V, Rossato S. Landslides of the western dolomites: case studies from the Adige and Sarca valleys (NE Italy). Alp Mediterr Quat 2020;33:191–207. https://doi.org/10.26382/AMQ.2020.15.

[30] Ruggia G, Ivy-Ochs S, Aaron J, Steinemann O, Martin S, Rigo M, et al. Reconstructing the Gorte and Spiaz de Navesele Landslides, NE of Lake Garda, Trentino Dolomites (Italy). Geosciences 2021;11(10):404. https://doi.org/10.3390/geosciences11100404.

[31] Viganò A, Bressan G, Ranalli G, Martin S. Focal mechanism inversion in the Giudicarie-Lissini seismotectonic region (Southern Alps, Italy): insights on tectonic stress and strain. Tectonophysics 2008;460:106–15. https://doi.org/10.1016/j.tecto.2008.07.008.

[32] Stucchi M, Meletti C, Montaldo V, Crowley H, Calvi GM, Boschi E. Seismic hazard assessment (2003-2009) for the Italian building code. Bull Seismol Soc Am 2011; 101(4):1885–911. https://doi.org/10.1785/0120100130.

[33] Oth A, Bindi D, Parolai S, di Giacomo D. Spectral analysis of K-NET and KiK-net data in Japan, Part II: on attenuation characteristics, source spectra, and site response of borehole and surface stations. Bull Seismol Soc Am 2011;101(2): 667–87. https://doi.org/10.1785/0120100135.

[34] Klin P, Laurenzano G, Priolo E. GITANES: a MATLAB package for estimation of site spectral amplification with the generalized inversion technique. Seismol Res Lett 2018;89(1):182–90. https://doi.org/10.1785/0220170080.

[35] Klin P, Laurenzano G, Barnaba C, Priolo E, Parolai S. Site amplification at permanent stations in northeastern Italy. Bull Seismol Soc Am 2021;111(4): 1885–904. https://doi.org/10.1785/0120200361.

[36] Castro R., Anderson J.G., Singh S.K. Site response, attenuation and source spectra of S waves along the Guerrero, Mexico, subduction zone. Bull Seismol Soc Am 1990;6A:1481–1503. M80. doi:10.1785/BSSA08006A1481.

[37] Sawada S. Phase characteristics on site amplification of layered ground with irregular interface. In: Irikura K, Kudo K, Okada H, Sasatani T, editors. The effects of surface geology on seismic motion, balkema. Rotterdam: Balkema; 1998. Rotterdam.

[38] Semblat JF, Kham M, Parara E, Bard PY, Pitilakis K, Makra K, et al. Seismic wave amplification: basin geometry vs soil layering. Soil Dynam Earthq Eng 2005;25(7): 529–38. https://doi.org/10.1016/j.soildyn.2004.11.003.

[39] Klin P, Laurenzano G, Romano MA, Priolo E, Martelli L. ER3D: a structural and geophysical 3-D model of central Emilia-Romagna (northern Italy) for numerical simulation of earthquake ground motion. Solid Earth 2019;10(3):931–49. https://doi.org/10.5194/se-10-931-2019.

[40] Abraham JR, Smerzini C, Paolucci R, Lai CG. Numerical study on basin-edge effects in the seismic response of the Gubbio valley, Central Italy. Bull Earthq Eng 2016; 14:1437–59. https://doi.org/10.1007/s10518-016-9890-y.

[41] Parolai S. Denoising of seismograms using the S transform. Bull Seismol Soc Am 2009;99(1):226–234.doi. https://doi.org/10.1785/0120080001.

[42] Nogoshi M, Igarashi T. On the propagation characteristics estimations of subsurface using microtremors on the ground surface. J Seismol Soc Jpn 1970;23: 264–80.

[43] Nakamura Y. A method for dynamic characteristics estimation of subsurface. Quart. Rep. Railway Tech. Res. Inst. (RTRI) 1989;30:25–33.

[44] Bard PY. Microtremor measurements: a tool for site effect estimation? In: Irikura, Kudo, Okada, Sasatani, editors. The effects of surface geology on seismic motion. Rotterdam, Netherland: Balkema; 1999. p. 1251–79. 1999.

[45] Field EH, Jacob KHJ. A comparison and test of various site-response estimation techniques, including three that are not reference-site dependent. Bull Seismol Soc Am 1995;85(4):1127–43.

[46] Konno K, Ohmachi T. Ground-motion characteristics estimated from spectral ratio between horizontal and vertical components of microtremor. Bull Seismol Soc Am 1998;88(1):228–41. https://doi.org/10.1785/BSSA0880010228.

[47] Calinski T, Harabasz J. A dendrite method for cluster analysis. Commun Stat Theor Methods 1974;3:1–27. https://doi.org/10.1080/03610927408827101.

[48] Rousseeuw PJ. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. J Comput Appl Math 1987;20:53–65. https://doi.org/10.1016/0377-0427(87)90125-7.

[49] Bindi D, Parolai S, Cara F, Di Giulio G, Ferretti G, Luzi L, Monachesi G, Pacor F, Rovelli A. Site amplifications observed in the Gubbio basin, Central Italy: hints for lateral propagation effects. Bull Seismol Soc Am 2009;99(2A):741–60. https://doi.org/10.1785/0120080238.

[50] Davis JC. Statistics and data analysis in geology. New York: John Wiley and Sons Inc.; 1986.

[51] Kristeková M, Kristek J, Moczo P. Time-frequency misfit and goodness-of-fit criteria for quantitative comparison of time signals. Geophys J Int August 2009;178 (2):813–25. https://doi.org/10.1111/j.1365-246X.2009.04177.x.

[52] Okabe AB, Boots B, Sugihara K, Chiu SN. Spatial tessellations concepts and applications of Voronoi diagrams. Chichester: John Wiley & Sons; 2000. p. 671.

[53] Moscatelli M, Albarello D, Scarascia Mugnozza G, Dolce M. The Italian approach to seismic microzonation. Bull Earthq Eng 2020;18:5425–40. https://doi.org/10.1007/s10518-020-00856-6.