Geophysical Journal International

Geophys. J. Int. (2017) **211,** 650–662
Advance Access publication 2017 August 2
GJI Seismology

doi: 10.1093/gji/ggx325

Application of the principal component analysis (PCA) to HVSR data aimed at the seismic characterization of earthquake prone areas

Enrico Paolucci, Enrico Lunedei and Dario Albarello

Dipartimento di Scienze Fisiche, della Terra e dell'Ambiente, University of Siena, Siena, Italy. E-mail: enricopaolucci83@gmail.com

Accepted 2017 July 28. Received 2017 July 27; in original form 2017 May 9

SUMMARY

In this work, we propose a procedure based on principal component analysis on data sets consisting of many horizontal to vertical spectral ratio (HVSR or H/V) curves obtained by single-station ambient vibration acquisitions. This kind of analysis aimed at the seismic characterization of the investigated area by identifying sites characterized by similar HVSR curves. It also allows to extract the typical HVSR patterns of the explored area and to establish their relative importance, providing an estimate of the level of heterogeneity under the seismic point of view. In this way, an automatic explorative seismic characterization of the area becomes possible by only considering ambient vibration data. This also implies that the relevant outcomes can be safely compared with other available information (geological data, borehole measurements, etc.) without any conceptual trade-off. The whole algorithm is remarkably fast: on a common personal computer, the processing time takes few seconds for a data set including 100–200 HVSR measurements. The procedure has been tested in three study areas in the Central-Northern Italy characterized by different geological settings. Outcomes demonstrate that this technique is effective and well correlates with most significant seismostratigraphical heterogeneities present in each of the study areas.

Key words: Earthquake hazards; Seismic noise; Site effects.

1 INTRODUCTION

Several studies carried out in the last decades testify utility of ambient vibration measurements for characterizing local seismic response (e.g. Nakamura 1989; SESAME 2004; D'Amico *et al.* 2008; Albarello *et al.* 2011; Gallipoli *et al.* 2011; Paolucci *et al.* 2015; Farrugia *et al.* 2016). These measurements can be performed in multistation configurations (seismic arrays) or using a single-station acquisitions (e.g. Mucciarelli 1998; Parolai *et al.* 2005; Picozzi *et al.* 2005a; Foti *et al.* 2011). As concerns the latter ones, ambient vibration recording is performed with a three-directional sensor in order to estimate average spectral ratios between the horizontal (H) and the vertical (V) components of the ground motion (horizontal to vertical spectral ratios or HVSRs or H/V). Being aware that the pattern of HVSRs as a function of the frequency (the HVSR curve) is not a direct representation of the local amplification function (e.g. Field & Jacob 1993; Bindi *et al.* 2000; Lunedei & Albarello 2010), it is nevertheless informative about the configuration of the local subsoil (see, e.g. Lachet & Bard 1994; Bard 1999; Haghshenas *et al.* 2008). In particular, despite of the fact that disagreement exists about the physical interpretation of the HVSR curve (e.g. Fäh *et al.* 2001; Lunedei & Albarello 2010; Albarello & Lunedei 2011; Sanchez-Sesma *et al.* 2011), there is a general consensus about the fact that in the presence of a sharp impedance contrast in the subsoil, the HVSR curve shows a marked peak corresponding to the fundamental resonance frequency of the sedimentary cover (e.g. Bonnefoy-Claudet *et al.* 2006). The amplitude of this peak is related (in a non-linear and nowadays still not completely explained way) to the amplitude of the seismic impedance contrast responsible for the observed resonance (e.g. Albarello & Lunedei 2011). Thus, HVSR represents an important tool for detecting seismic resonance phenomena. Furthermore, since the resonance frequency depends on the local V_S profile, HVSR also provides important information about the configuration of mechanical features affecting local seismic response. This does not mean that this kind of analysis is exhaustive for assessing eventual seismic amplification phenomena, but it is of great utility to orientate more detailed (and costly) studies.

Cheapness and quickness of the single-station ambient vibration measurements make this technique a basic element for microzoning studies in Italy. A huge amount of HVSR measurements were performed in the last years after the most damaging earthquakes (e.g. Gallipoli *et al.* 2011; Paolucci *et al.* 2015) and, more in general, in the frame of Seismic Microzonation studies carried out in thousands of municipalities all over the country. In this latter context, single-station acquisitions provide an important support to the first of the three investigation levels proposed by Seismic Microzonation Italian guidelines (Albarello *et al.* 2015; SM Working Group 2015) where extensive surveys are required. In this last context, HVSR technique represents an important tool to

© The Authors 2017. Published by Oxford University Press on behalf of The Royal Astronomical Society.

Downloaded from https://academic.oup.com/gji/article/211/1/650/4060697 by guest on 10 April 2024

localize resonance phenomena and constrain geometry of major seismic impedance contrasts in the explored area (e.g. Albarello *et al.* 2011; Gallipoli *et al.* 2011; Albarello *et al.* 2015). The huge amount of data collected in these studies (tens to hundreds HVSR measurements) makes of interest the development of computer-aided procedures for grouping and classifying similar HVSR curves supporting the identification of areas characterized by an homogeneous seismic behaviour due to the presence of similar subsoil configurations. To this purpose, several approaches have been proposed including hierarchical (e.g. Rodriguez and Midorikawa 2002) and non-hierarchical (e.g. Bragato *et al.* 2007; Ullah *et al.* 2013; Capizzi *et al.* 2015; Panzera *et al.* 2017) cluster analysis. These techniques, however, imply 'ex-ante' choices (e.g. the acceptable level of similarity within the cluster or the overall number of clusters) that make outcomes prone to possible biases induced by the user.

The aim of this work is to investigate the possible use of the principal component analysis (PCA) for the analysis of the multivariate field of HVSR curves. Extensive treatments of PCA are given by Joliffe (2002) for a variety of applications and by Preisendorfer (1988), Davis (2002) and Wilks (2006) in the specific context of geosciences. The basic idea underlying the application of PCA in the present context is that each experimental multivariate pattern available in the area under study (the HVSR curves) can be interpreted as a linear combination of a common set of 'characteristic' patterns (principal components or PCs thereafter) that are mutually uncorrelated. In the frame of the present application, PCs are assumed to correspond to specific subsoil configurations present in the study area. In this respect, the PCA can be also considered a 'Pattern Recognition Algorithm' with unsupervised learning (e.g. Theodoridis & Koutroumbas 2008). The PCA also allows evaluating the importance of each PC in terms of the respective fraction of the observed overall variance of the collected HVSR data by also allowing a global evaluation of the lateral heterogeneity of the explored area. Furthermore, it is also possible to evaluate the respective importance of each PC at any of the considered sites and thus grouping the sites as a function of the most representative PC. In this respect, PCA can also be considered a computer-aided classification algorithm.

In order to explore the reliability, the effectiveness and the eventual drawbacks of the PCA application for the purposes described above, this technique has been applied in different geological contexts where several single-station measurements were collected. PCA basics in the context of HVSR analysis are shortly outlined first. Then, criteria adopted for applying PCA to HVSR patterns is described along with some applications in three different geological contexts in Central-Northern Italy.

2 PCA IN THE CONTEXT OF HVSR STUDIES

In order to perform PCA, we assume that HVSR values corresponding to F frequencies have been measured at S sites. These data are stored in the form of an $(S \times F)$ matrix of observations $[O]$: the generic element O_{sf} of $[O]$ represents the HVSR value relative to the frequency f measured at the sth site. The sth row of $[O]$ (hereafter indicated as $\{O\}_s$) is the HVSR curve determined at the sth site. In order to perform PCA, the centred matrix $[O']$ is considered instead of $[O]$, whose elements are

$$O'_{sf} = O_{sf} - \left(\sum_{n=1}^{F} O_{sn}\right)/F. \tag{1}$$

The key element of the PCA is the $(S \times S)$ variance/covariance matrix $[V_O]$, defined in the form

$$[V_O] = \left([O'][O']^T\right)/(F-1), \tag{2}$$

where the apex T indicates the transpose of the relevant matrix. The diagonal elements of $[V_O]$ represent the variability of the HVSR curve at the sth site, whereas the off-diagonal elements represent the degree of similarity between HVSR curves at the different sites. By the Spectral Decomposition Theorem (e.g. Wilks 2006), one has

$$[V_O] = [E][\Lambda][E]^T, \tag{3}$$

where $[\Lambda]$ is an $(S \times S)$ diagonal matrix whose non-zero elements Λ_{jj} are the eigenvalues of $[V_O]$. These values are arranged in a way that $\Lambda_{jj} \geq \Lambda_{(j+1)(j+1)}$. The columns of the orthogonal matrix $[E]$ are the S eigenvectors of $[V_O]$ such that

$$\sqrt{\sum_{i=1}^{S} (E_{ij})^2} = 1, \quad \forall j = 1, \ldots, S, \tag{4}$$

where E_{ij} are the elements of $[E]$.

The trace tr$[\Lambda]$ of $[\Lambda]$ is the overall variance of the HVSR values in the whole sets of S sites and F frequencies. The fraction R_j of variance associated with each jth eigenvector is

$$R_j = \Lambda_{jj}/\mathrm{tr}[\Lambda]. \tag{5}$$

The eigenvectors in $[E]$ are arranged in a way such that $R_j \geq R_k$ when $j < k$. One can demonstrate (Appendix) that a new $(S \times F)$ matrix $[U]$ can be defined such that

$$[O'] = [E][U], \tag{6}$$

being the S rows $\{\mathrm{U}\}_j$ of [U]mutually uncorrelated. In this view, the centered HVSR curve at the sth site (the row $\{\mathrm{O}'\}_s$) can be seen as a linear combination of S 'patterns' (the PCs $\{U\}_j$):

$$\{O'\}_s = \sum_{j=1}^{S} E_{sj}\{U\}_j, \tag{7}$$

where E_{sj} are elements of the matrix $[E]$ ('loadings' in the PCA jargon). In the present context, the S PCs identify a set of S HVSR 'patterns' each representative of a fraction of the overall variability of the original data set. It is worth to note that each PC is 'centred' (i.e. the average of its amplitude values with respect the frequency index is zero) and scaled in amplitude as a function of the amplitudes of HVSR curves. In fact, from eq. (A5) one has:

$$U_{jf} = \sum_{k=1}^{S} E^T_{jk} O'_{kf}, \tag{8}$$

which implies that the amplitude of the jth PC at the fth frequency is proportional to the amplitude of the HVSR curves measured at the same frequency at all the S sites 'weighted' by the E^T_{jk} coefficients: the higher is the HVSR value generally measured at the fth frequency, the higher is the amplitude of $\{U\}_j$ at the same frequency. In the following, the amplitudes of the $\{U\}_j$ elements will be indicated as 'apparent' HVSR values. The range D_j of amplitude variation relative to the jth PC will be defined as:

$$D_j = \max\{U\}_j - \min\{U\}_j. \tag{9}$$

Each PC mimics the 'pattern' of an experimental HVSR curve as a function of frequency but no quantitative correspondence is expected to exist between respective amplitudes. It is worth noting that the sign of E_{sj} determines the 'polarity' of the PC at the relevant site:

Downloaded from https://academic.oup.com/gji/article/211/1/650/4060697 by guest on 10 April 2024

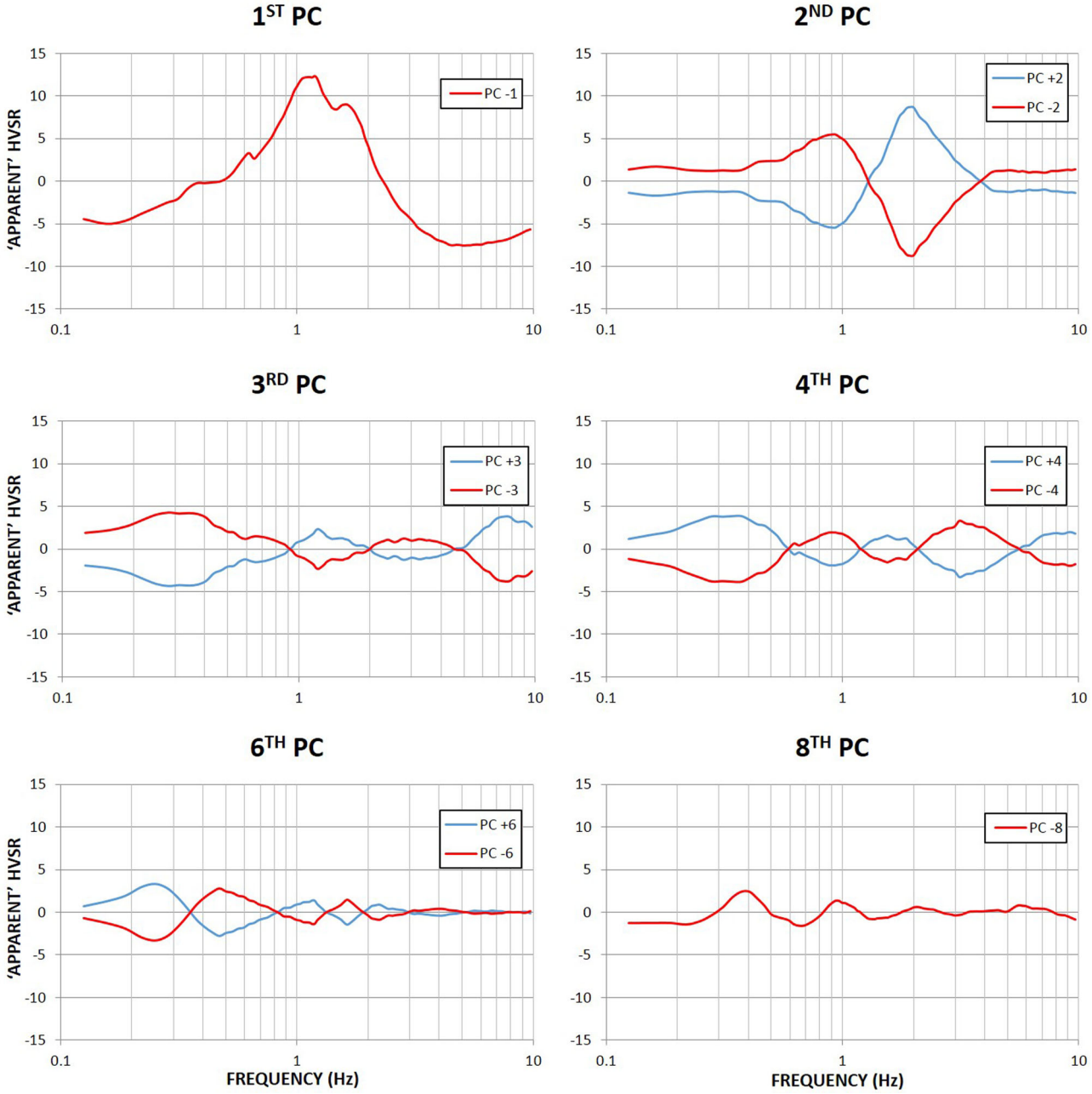

Figure 1. Principal components of HVSR data in the Collesalvetti municipality area 'dominating' at least one site where HVSR values have been measured. Blue and red curves, respectively, represent 'characteristic' HVSR curves with positive and negative 'polarities' depending on the sign of the corresponding 'loading' (see the text for details).

maxima and minima of $\{U\}_j$ reverse when the sign of E_{sj} changes. Since HVSR technique mainly aims at identifying maxima (that will correspond to the resonance frequency of the considered subsoil), each PC could represent two different subsoil configurations depending on the respective E_{sj} sign. Thereafter, a 'weight' W_{sj} will be defined in the form

$$W_{sj} = D_j \left| E_{sj} \right|, \tag{10}$$

which expresses the relative importance of the jth PC at the sth site. It is also is a measure of the relative amplitude of the peaks in the jth PC at the sth site.

From the computational point of view, PCA has been carried out by using the 'princomp' MATLAB® routine, which provides the values of the 'loadings' and the elements of the rows $\{U\}_j$ of $[U]$ ('scores' in the PCA jargon), as well as the vector containing the eigenvalues of the variance/covariance matrix of $[O']$. As specified above, this latter information is needed to calculate the parameters R_j (see eq. 5).

3 CLASSIFICATION CRITERIA

The PCA provides three main outcomes that can be used for the analysis of the HVSR curves obtained in the survey: the set of the 'characteristic' patterns $\{U\}_j$ (i.e. the PCs), the fraction of the overall variance R_1 explained by the first and most important PC

Downloaded from https://academic.oup.com/gji/article/211/1/650/4060697 by guest on 10 April 2024

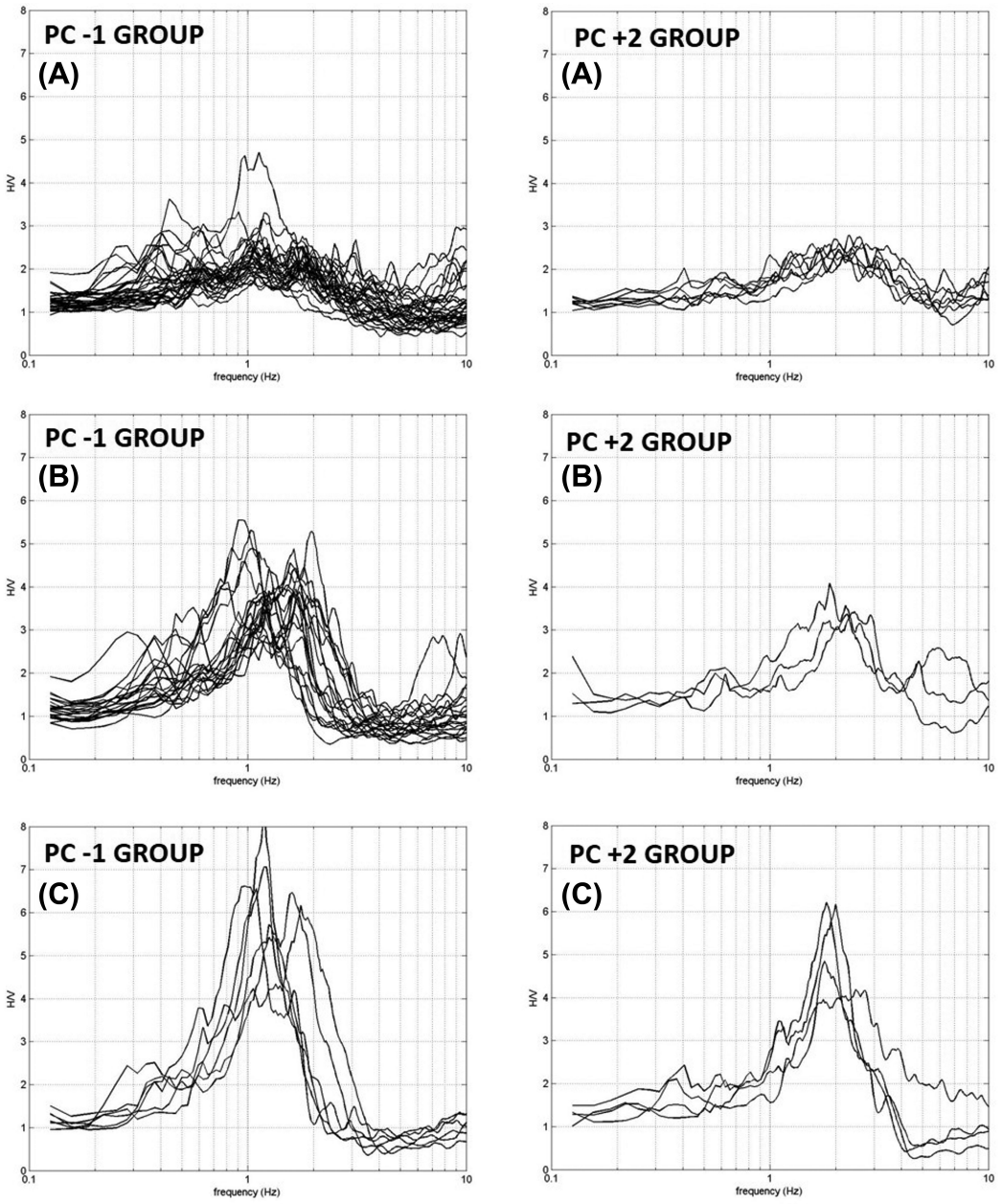

Figure 2. HVSR experimental curves in the Collesalvetti municipality area belonging to the most populated groups, respectively, characterized by PC−1 and PC+2 patterns. For each patterns, sites characterized by different levels of amplitude are separately shown. In particular, sites in A are associated with values of $\hat{W}_s < 2$, sites in B to values $2 \leq \hat{W}_s < 4$ and in C to values $\hat{W}_s \geq 4$.

(eq. 5) and the 'weight' W_{sj} of each characteristic pattern at the specific sth site (eq. 10). It is worth noting that, since two 'polarities' are possible for each PC due to the sign of the relevant coefficient E_{sj}, two 'characteristic' opposite patterns will be considered for each $\{U\}_j$, that will be indicated respectively as $\{U^+\}_j$ and $\{U^-\}_j$.

These three parameters have been considered for classifying the measured HVSR curves and, as a consequence, the corresponding subsoil configurations.

The parameter R_1 is considered to evaluate the 'absolute' level of heterogeneity of the HVSR measurements and, consequently, of the geological configuration responsible for the observed patterns. In particular, when R_1 is high (0.8 to say), the geological configuration can be considered as relatively homogeneous and dominated by the first PC, while low values (0.3 to say), can be interpreted as the effect of marked geological heterogeneities in the explored area. This allows comparing different areas analysed by PCA.

The 'characteristic' patterns of the HVSR curves obtained in the study area will be represented by plotting the $\{U^+\}_j$ and $\{U^-\}_j$ as a function of the frequency index f. In this way, eventual resonance frequencies typical of the study area can be identified by the frequency values corresponding to the maxima of the considered PC.

The 'weights' W_{sj} associated with each PC will be used to classify the S sites as a function on the 'dominant' PC among the ones computed at the sth site. In particular, the dominant PC at the sth site (hereafter indicated as $\{\hat{U}\}_s$) will be chosen as the one corresponding to the maximum $\hat{W}_s$ among the W_{sj} values relative to the sth site.

Downloaded from https://academic.oup.com/gji/article/211/1/650/4060697 by guest on 10 April 2024

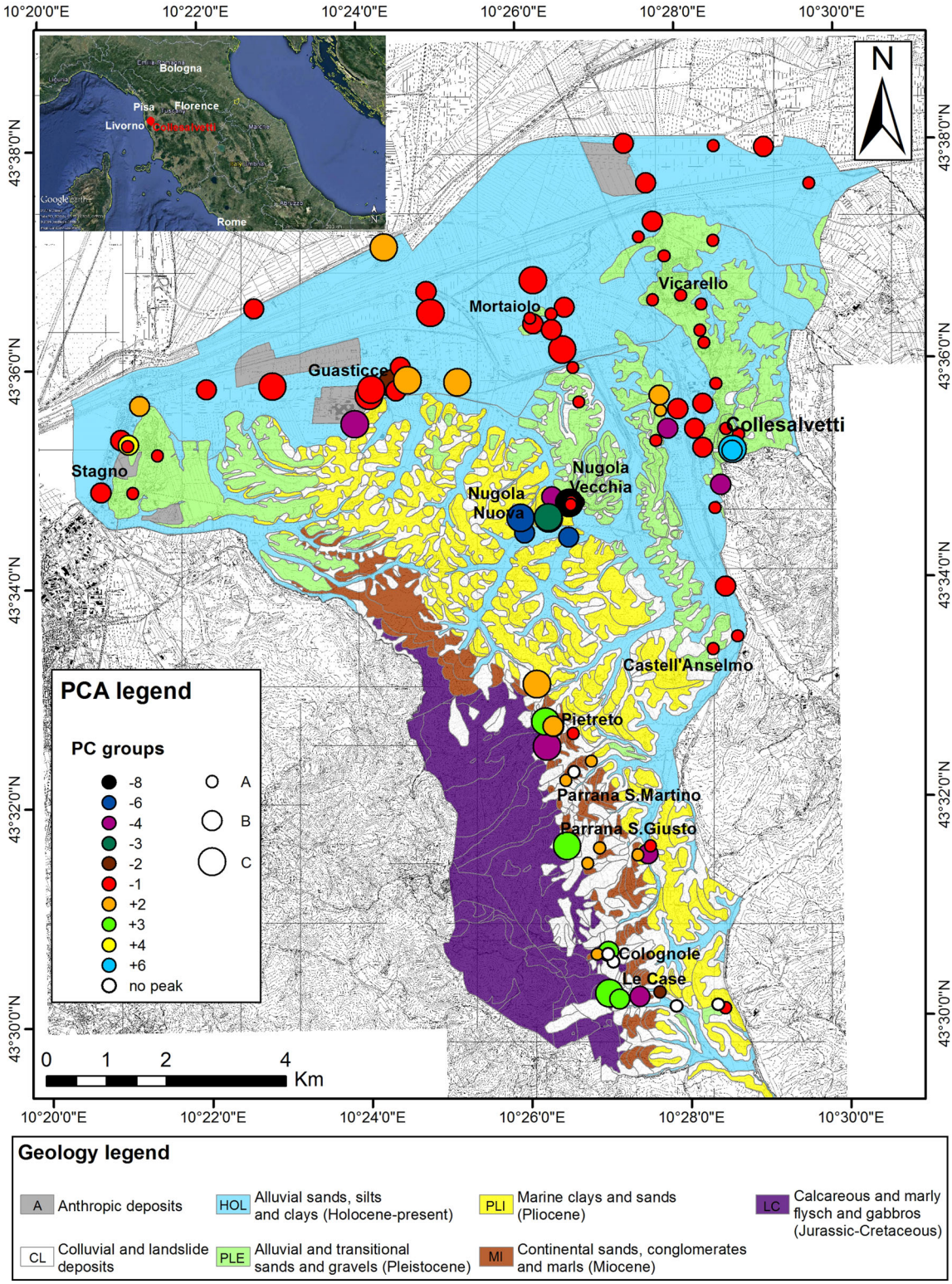

Figure 3. The dots in the map represent the locations of the single-station HVSR measurements, superimposed over the geological map of the Collesalvetti municipality area. The spatial distribution of the PC groups and their corresponding different levels of amplitude (A, B and C; see Fig. 2) are indicated by colour and size of the dots, respectively. Geology data from Elter *et al*. (1964), Lazzarotto *et al*. (1990a,b) and Elter & Marroni (1991).

The amplitude of the maxima of the dominant $\{\hat{U}\}_s$ will be estimated by the corresponding $\hat{W}_s$ value: the highest is $\hat{W}_s$, the larger will be the maxima. This parameter will also allow identifying sites where resonance effects are lacking, that is, where the corresponding HVSR curves are 'flat' due to absence of significant maxima. These situations will correspond to sites where $\hat{W}_s < W_0$, where W_0 is an empirical threshold to be determined. Based on a number of preliminary attempts, the threshold W_0 has been fixed to

Downloaded from https://academic.oup.com/gji/article/211/1/650/4060697 by guest on 10 April 2024

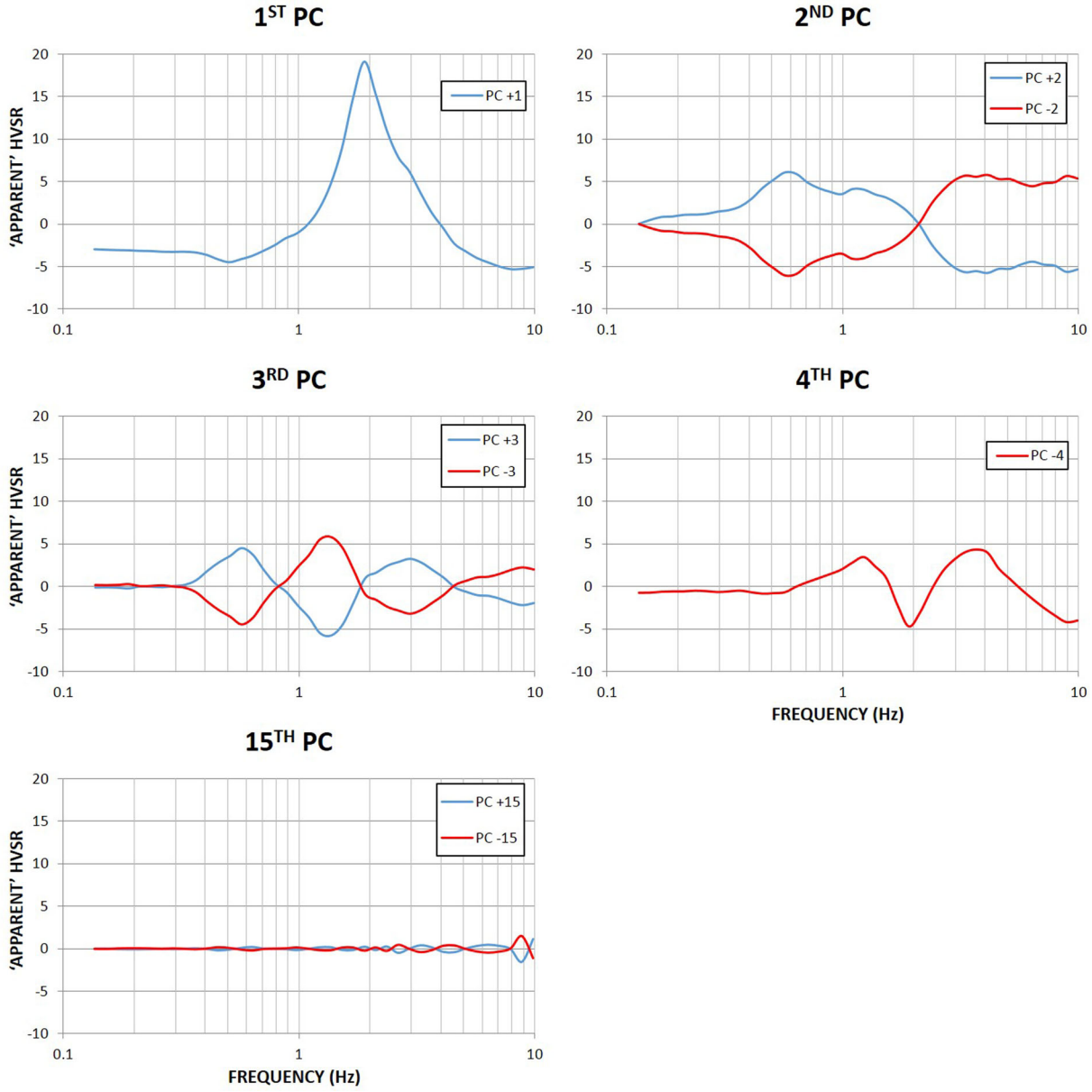

Figure 4. Principal components of HVSR data in the Montecatini Terme municipality area 'dominating' at least one site where HVSR values have been measured. Blue and red curves, respectively, represent 'characteristic' HVSR curves with positive and negative 'polarities' depending on the sign of the corresponding 'loading' (see the text for details).

0.6 since, in general, none of the sites where $\hat{W}_s < 0.6$ shows HVSR peaks above 2, that is the minimum threshold for a significant HVSR maximum by following SESAME criteria (SESAME 2004).

4 CASE STUDIES

4.1 Collesalvetti municipality (Central Italy)

The HVSR data set consists of 100 single-station measurements performed by a three-directional digital tromograph Tromino Micromed (www.tromino.eu) with a sampling frequency of 128 Hz and an acquisition time of 20 min. At each site, the HVSR curve was obtained dividing the time-series into 60 non-overlapping windows with 20 s length and following the procedure described in Picozzi *et al*. (2005b). PCA was applied to HVSR values in the range 0.1–10 Hz.

The most important PC explains about 61 per cent of the overall variance and this suggests a rather heterogeneous subsoil configuration. The shape of PCs dominating at least one site is shown in Fig. 1. The most important one (PC−1 in Fig. 1 corresponding to negative loading) is characterized by a clear main peak slightly above 1 Hz: it dominates about 57 sites out of the 100 considered. The second PC, which explains almost the 20 per cent of the overall variance, includes two patterns (PC+2 and PC−2 in Fig. 1) presenting clear peaks at 1 and 2 Hz, respectively. The remaining dominant PCs (which explain together almost the 12 per cent of the overall variance) show at least two small peaks, some of which located in the range 0.25–0.5 Hz (Fig. 1). All the sites have been grouped as a function of the dominant PC and, within each group, the amplitude

Downloaded from https://academic.oup.com/gji/article/211/1/650/4060697 by guest on 10 April 2024

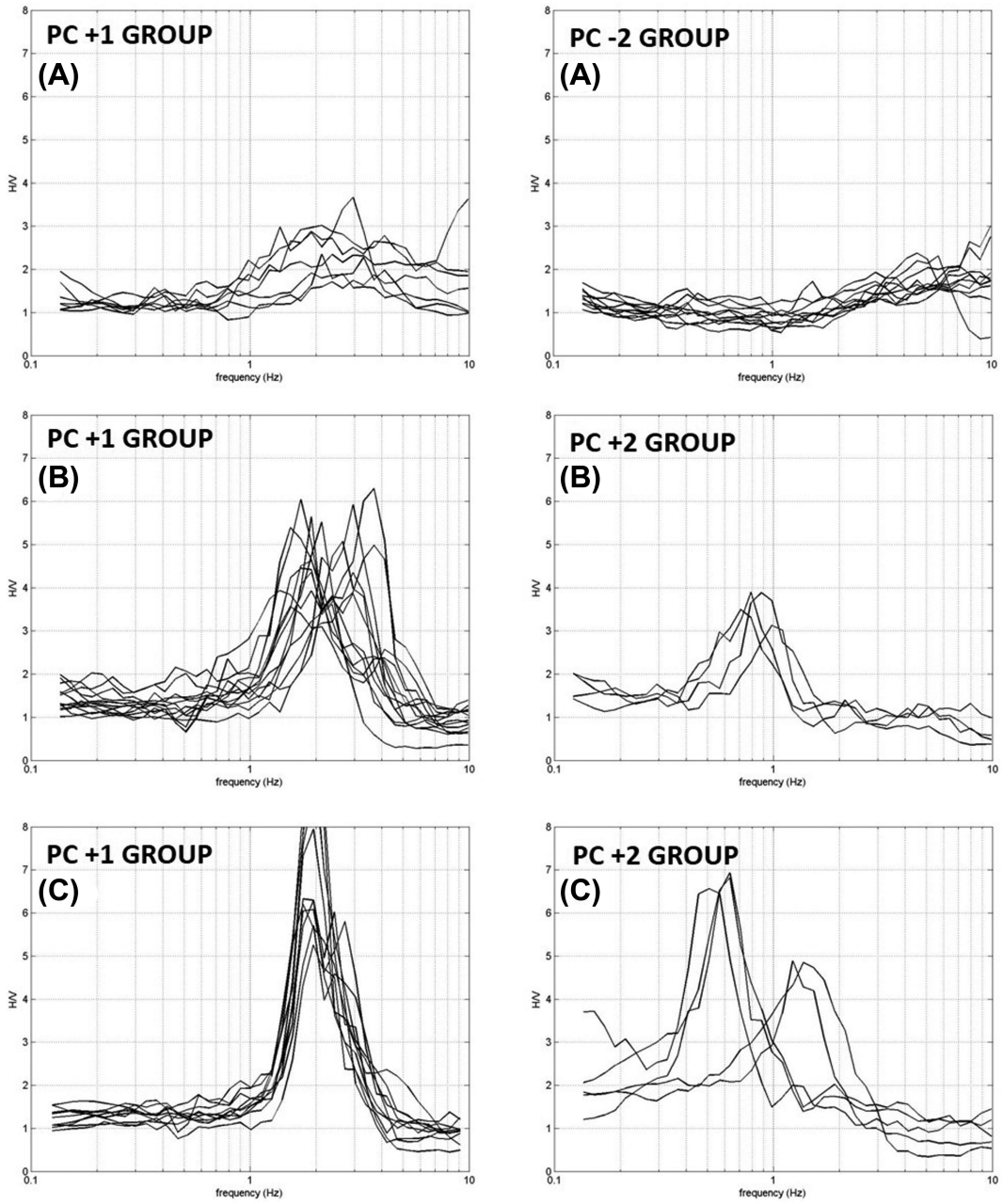

Figure 5. HVSR experimental curves in the Montecatini Terme municipality area belonging to the most populated groups, respectively, characterized by PC+1 and PC±2 patterns. For each patterns, sites characterized by different levels of amplitude are separately shown. In particular, sites in A are associated with values of $\hat{W}_s < 2.5$, sites in B to values $2.5 \leq \hat{W}_s < 4.5$ and in C to values $\hat{W}_s \geq 4.5$.

of the peaks has been evaluated by considering the corresponding $\hat{W}_s$ values in order to put in evidence sites characterized by more intense resonance phenomena. The original experimental HVSR curves corresponding to the most populous groups are shown in Fig. 2, by also considering the respective $\hat{W}_s$ values. These plots show the effectiveness of PCA in identifying representative patterns actually present in the area.

The map in Fig. 3 shows the spatial distribution of the groups and their possible correspondence with local geology. It is possible to note that sites dominated by the PC−1 component are mainly located in the northernmost part of the explored area, where the Quaternary sediments (Pleistocenic–Holocenic) crop out with the highest amplitude peaks in correspondence with the most recent ones (Holocenic). This is in line with the hypothesis that major impedance contrast correspond to soft unconsolidated sediments directly overlying the bedrock. At first glance interpretation, resonance frequency slightly above 1 Hz could correspond to a bedrock depth of the order of hundred metres (Albarello *et al.* 2011). The same deposits also correspond to sites dominated by PC+2 and PC−2 patterns showing a single maximum at frequency at 2 and 1 Hz, respectively. These sites could correspond to different depths of the resonant interface (rigid bedrock) respectively shallower and deeper than in the case of PC−1 sites. Sites characterized by PC−1 and PC±2 with higher $\hat{W}_s$ values represent most critical situations, since ground resonance frequencies are close to natural frequencies of most buildings and for the presence of strong impedance contrast

Downloaded from https://academic.oup.com/gji/article/211/1/650/4060697 by guest on 10 April 2024

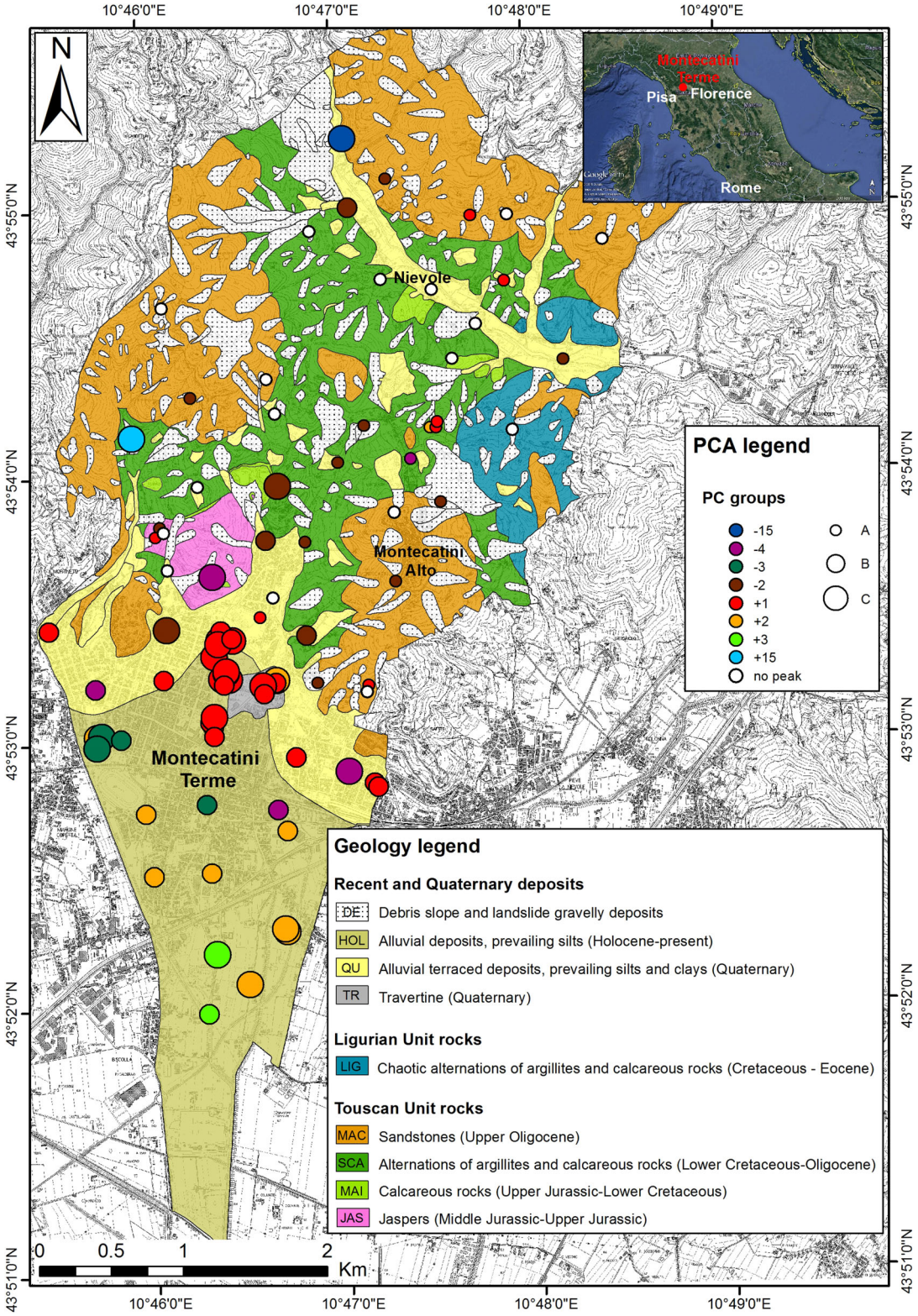

Figure 6. The dots in the map represent the locations of the single-station HVSR measurements, superimposed over the geological map of the Montecatini Terme municipality area. The spatial distribution of the PC groups and their corresponding different levels of amplitude (A, B and C; see Fig. 5) are indicated by colour and size of the dots, respectively. Geology data from Trevisan (1954), Brandi *et al.* (1967) and Puccinelli *et al.* (2000).

in the subsoil, possibly responsible for large amplification of the local seismic ground motion (e.g. Kramer 1996).

Outcrops of Miocenic sediments (continental sands, conglomerate and marls) in the southwestern part of the area are characterized by the presence of PC±2 patterns (with low-amplitude peaks) and by PC+3 and PC−4 patterns characterized by multiple peaks. This suggests the heterogeneity characterizing these deposits, where multiple resonance interfaces exist at different depths (from tens to

Downloaded from https://academic.oup.com/gji/article/211/1/650/4060697 by guest on 10 April 2024

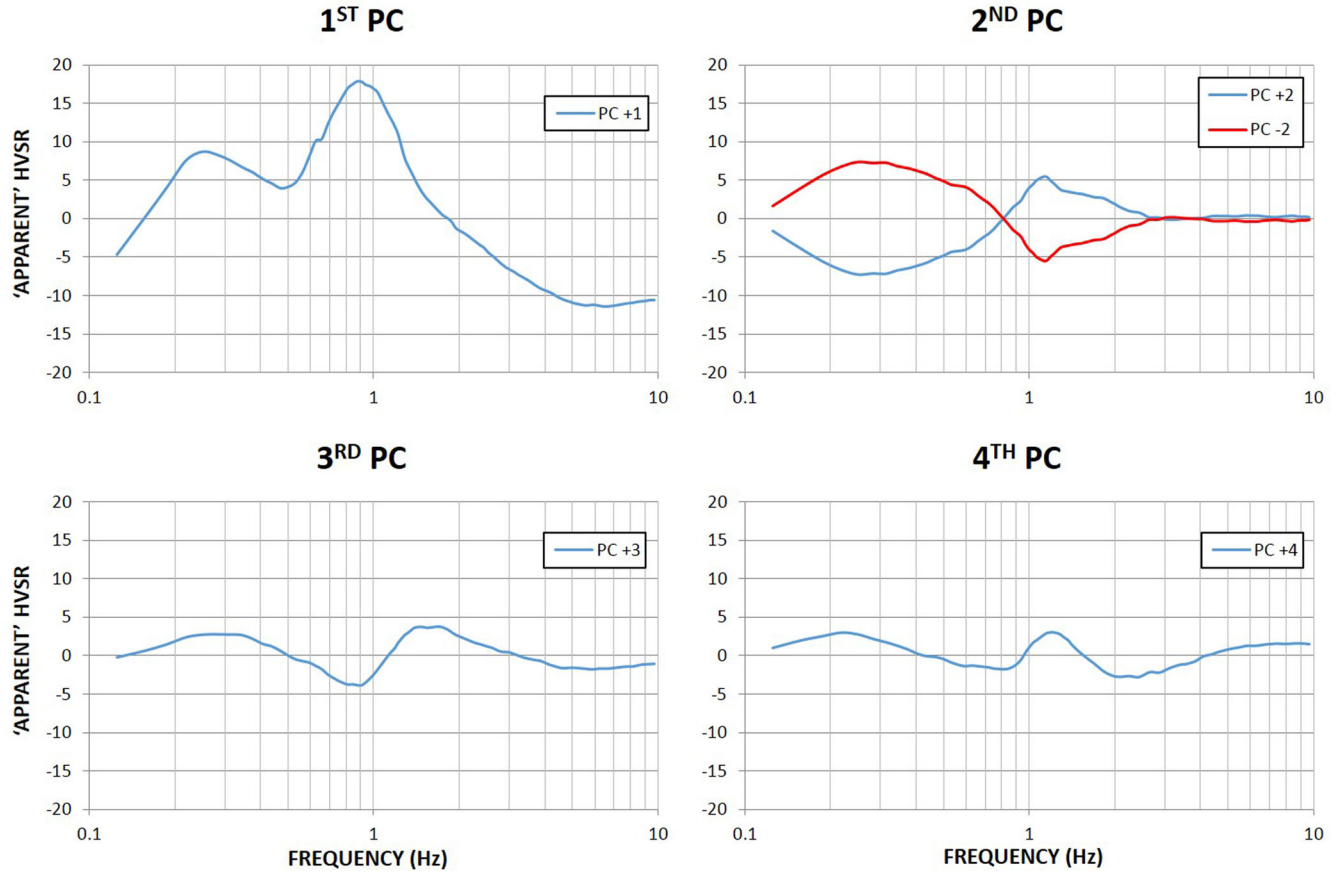

Figure 7. Principal components of HVSR data in the Emilia-Romagna area 'dominating' at least one site where HVSR values have been measured. Blue and red curves, respectively, represent 'characteristic' HVSR curves with positive and negative 'polarities' depending on the sign of the corresponding 'loading' (see the text for details).

hundreds of metres), also due to the presence of colluvial and landslide deposits. Pliocenic deposits (marine clays and sands) in the central part of the area are characterized by the presence of HVSR curves belonging to PC−3, PC−6 and PC−8 groups: all these patterns present a clear peak at 0.3–0.4 Hz, indicating a deep impedance contrast (hundreds of metres).

Five sites, located in the southernmost part of the study area, have been considered to be characterized by the absence of resonance on the basis of $\hat{W}_s$ values ('no peak' in Fig. 3). These measurements are mainly situated where rigid old rocks (Jurassic-Cretaceous) crop out: in these cases, no significant resonance effects are expected due to the lack of soft sedimentary cover above the bedrock.

4.2 Montecatini Terme municipality (Central Italy)

The HVSR data set consists of 85 measurements performed by three-directional digital tromographs Tromino Micromed and SR04HS Sara (http://www.sara.pg.it/) with a sampling frequency of 128 and 100 Hz, respectively. The ambient vibrations were acquired for 20 min. At each site, the HVSR curve was obtained dividing the time-series into 40 non-overlapping windows with 30 s length and following the procedure described in Picozzi *et al*. (2005b). PCA was applied to HVSR values in the range 0.1–10 Hz.

Fig. 4 shows the 'characteristic' HVSR patterns corresponding to PCs dominating at least one site where HVSR has been measured. The most important PC explains about 50 per cent of the overall variance by denoting the presence of relatively large heterogeneities in the subsoil configuration. The PC+1 pattern is characterized by a clear single peak at about 2 Hz. The second PC, which explains about the 20 per cent of the overall variance, corresponds to two patterns (PC+2 and PC−2). PC+2 shows a wide peak possibly resulting from two close peaks at about 0.6 and 1 Hz, while PC−2 is characterized by a general rise of the apparent HVSR at higher frequencies (3–10 Hz). The remaining dominant PCs explain together almost the 15 per cent of the overall variance. It is possible to note that the 'characteristic' patterns of PC±3 and PC−4 present clear peaks in the range 0.55–4 Hz, while PC±15 are characterized by high-frequency peaks at about 9–10 Hz.

The experimental HVSR curves grouped as a function of the dominating PC are reported in Fig. 5. Only the most populous groups are shown, by also considering the respective $\hat{W}_s$ values.

The map in Fig. 6 shows the spatial distribution of the detected groups and their possible correspondence with local geology. Absence of resonance and PCs with relatively low $\hat{W}_s$ mainly concerns sites located in the northern part of the area where oldest and stiffer geological formations (Ligurian and Tuscan units) crop out. In this area, possibly due to the presence of debris slope and landslide shallow deposits, PCs characterized by high-frequency peaks (PC−2 and PC±15) are revealed. Most interesting resonance effects are revealed where most recent sediments (Holocenic and Quaternary possibly overlying rigid Tuscan Units) crop out: in these sites, PCs with clear peaks in the frequency range 0.5–4 Hz dominate (PC+1, PC+2, PC±3 and PC−4). Resonance frequency around 2 Hz (PC+1) dominates in the central part of the study area

Downloaded from https://academic.oup.com/gji/article/211/1/650/4060697 by guest on 10 April 2024

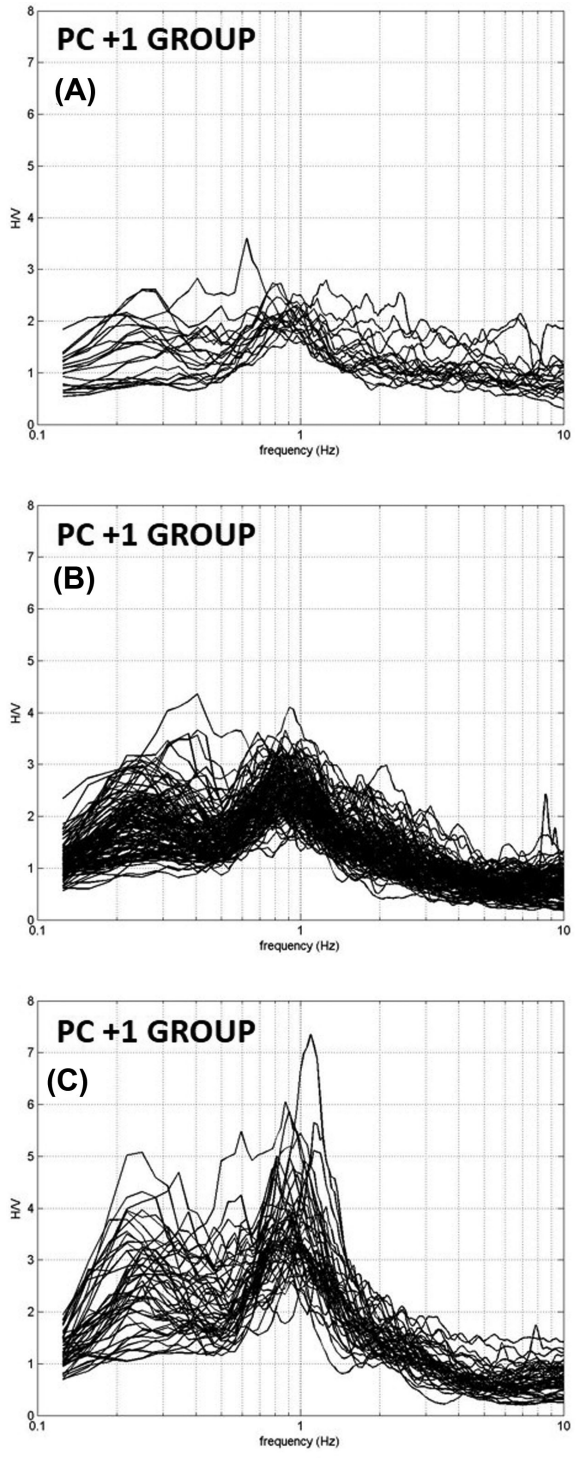

Figure 8. HVSR experimental curves in the Emilia-Romagna area belonging to the most populated group characterized by PC+1 pattern. For this pattern, sites characterized by different levels of amplitude are separately shown. In particular, sites in A are associated with values of $\hat{W}_s < 1.2$, sites in B to values $1.2 \leq \hat{W}_s < 2.5$ and in C to values $\hat{W}_s \geq 2.5$.

(where the main town is located), while other groups are distributed towards the borders and southwards. In particular, PC+2 and PC+3, showing lower frequency peaks (0.5 Hz) dominates sites located southward, while higher frequency peaks (PC−4) dominate border areas. This general pattern is probably related to the buried morphology of the basin characterized by a progressive deepening of the bedrock southward from few tens to about hundred metres.

4.3 Emilia-Romagna area (Northern Italy)

The area under study was hit by the seismic sequence in 2012 (e.g. Scognamiglio *et al.* 2012). The HVSR data set considered for this PCA application is part of the ambient vibration survey carried out in this area by teams from different institutions and universities during the seismic sequence and in the framework of the subsequent Seismic Microzonation project (Martelli & Romani 2013). These measurements are analysed and interpreted together with several seismic array acquisitions by Paolucci *et al.* (2015), in order to estimate the depth of impedance contrasts responsible for seismic resonance phenomena.

The HVSR data set consists of 204 measurements performed by three-directional digital tromograph Tromino Micromed with a sampling frequency of 128 Hz and acquisition time of 15–20 min. HVSR curves were obtained dividing the time-series into 60 non-overlapping windows with 20 s length and following the procedure described in Picozzi *et al*. (2005b). PCA was applied to HVSR values in the range 0.1–10 Hz.

The first PC explains about 80 per cent of the overall variance and this suggest a quite homogeneous subsoil configuration. The patterns of the 'characteristic' HVSR curves corresponding to PCs dominating at least one site (as a whole explaining about 90 per cent of the whole variance) are reported in Fig. 7. PC+1 clearly show two low-frequency peaks corresponding, respectively, to 0.25 and 0.9 Hz. This suggests the presence of at least two main impedance contrasts, respectively, located to depths of several hundreds and about hundred metres (Albarello *et al.* 2011). The second PC, which explains about the 8 per cent of the overall variance, is represented by two patterns, PC+2 and PC−2. The first one is characterized by a peak at almost 1 Hz, while PC−2 shows a broad peak at lower frequencies. As concerns PC+3 and PC+4, their patterns are quite similar to PC+1, with the higher frequency peak shifted to 1.5 Hz and an additional peak located at 8–9 Hz in PC+4 curve.

The original experimental HVSR curves corresponding to the PC+1 group (the most populous one) are shown in Fig. 8, by also considering the respective $\hat{W}_s$ values. The map in Fig. 9 shows that most of HVSR curves (193 out of 204) belong to sites where PC+1 dominates: this is in line with the geological homogeneity of crops out. By examining the spatial distribution of the relevant $\hat{W}_s$ values, it is possible to note that the HVSR curves with high-amplitude peaks are mainly located in Mirandola-Medolla area, consequently with the presence of a sharp impedance contrast (see Paolucci *et al.* 2015). Moreover, it was observed that $\hat{W}_s$ values are always higher than 0.6: this is in keeping with the absence of flat curves in the whole study area.

The general spatial pattern detected by PCA slightly contrasts with differences in the overall HVSR patterns identified by visual inspection by Paolucci *et al.* (2015). In that case, three main subzones where detected showing subtle differences in the HVSR patterns that have not been revealed by PCA.

5 CONCLUSIONS

A new approach is here presented to analyse outcomes of extensive ambient vibration surveys to detect main seismostratigraphical configurations at the scale of the order of tens–hundreds square kilometres or larger. This is achieved by the PCA of HVSRs collected

Downloaded from https://academic.oup.com/gji/article/211/1/650/4060697 by guest on 10 April 2024

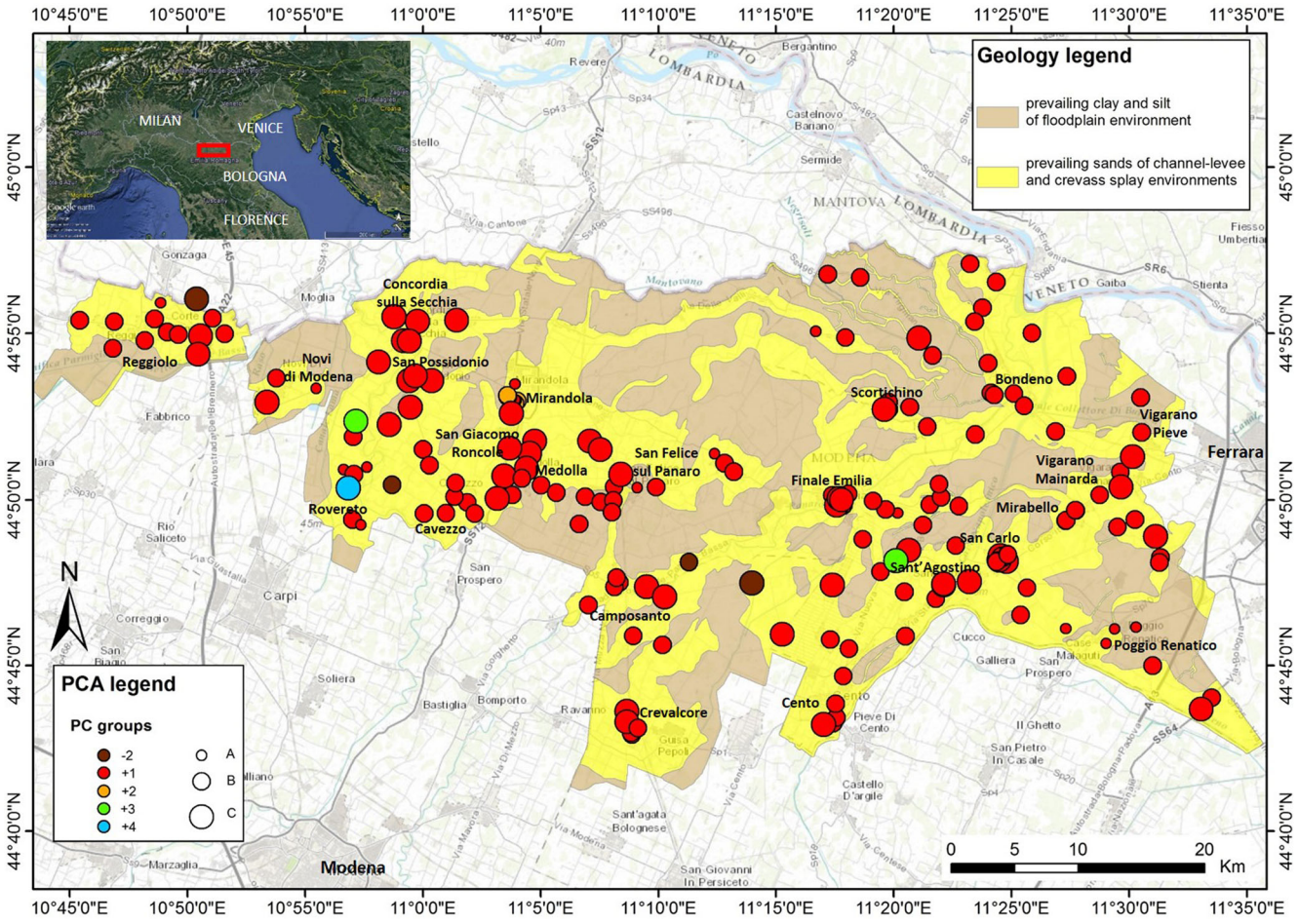

Figure 9. The dots in the map represent the locations of the single-station HVSR measurements, superimposed over the geological map of the investigated area in Emilia Romagna region. The spatial distribution of the PC groups and their corresponding different levels of amplitude (A, B and C; see Fig. 8) are indicated by colour and size of the dots, respectively. Geological data are from Pieri & Groppi (1981), RER & ENI-Agip (1998), Martelli & Romani (2013), Martelli *et al*. (2014) and Paolucci *et al*. (2015).

at a number of sites and is here proposed as a tool exploring main seismic features of an earthquake prone area, whose outcomes may drive more detailed seismic response analyses and seismic microzoning studies.

The proposed approach allows measuring the level of seismostratigraphical heterogeneity in the explored area, identifying a small number of mutually uncorrelated characteristic patterns and grouping sites characterized by similar seismic resonance phenomena. Differently from other approaches, this is possible without any 'ex-ante' assumption about the number and localization of explored patterns. The procedure is numerically fast and allows managing and analysing hundreds of measurements within few seconds on a common personal computer.

Three applications have been presented here and respective outcomes have been compared with geological information available for the explored areas. In all the cases, the level of heterogeneity revealed by PCA agrees with geological information available for the area. Furthermore, since these patterns can be associated with specific geological outcrops, the available geological map can be used to delimitate areas potentially characterized by similar seismic behaviour. In particular, the characteristic patterns deduced from PCA allow the identification of sites where similar seismic resonance phenomena are expected to occur. This also allows a preliminary identification of areas potentially most dangerous for buildings. In fact, when soil resonance frequency is close to resonance frequency of buildings (0.8–5 Hz to say) possible damage enhancement is expected (e.g. Kumar & Narayan 2008). This will allow the identification of situations potentially most critical, where more detailed studies for the quantitative definition of seismic response are needed and where the definition of risk reduction strategies is more urgent.

As concerns the shortcomings of the procedure, examples here described also indicate that only most representative patterns can be revealed by this kind of analysis. In particular, as seen in the case of Emilia Romagna study, PCA was not able to individuate the slight differences among the HVSR patterns revealed by visual inspection. Moreover, it is worth reminding that reliability of PCA outcomes relies on the quality of original HVSR data. Accurate pre-processing (see, e.g. Bonnefoy-Claudet *et al*. 2009; Albarello *et al*. 2011) and careful reliability checking of the curves (e.g. following the quality criteria of SESAME 2004 and Albarello *et al*. 2011) are basic prerequisites: the presence of unreliable experimental HVSR patterns can indeed affect the PCA outcomes by identifying in turn unreliable 'characteristic' HVSR patterns. On the other hand, applying PCA on redundant measurements might allow the identification of true patterns eventually hidden by the noise introduced by biased HVSR curves.

Downloaded from https://academic.oup.com/gji/article/211/1/650/4060697 by guest on 10 April 2024

ACKNOWLEDGEMENTS

We are grateful to G. Peruzzi, P. Pieruccini, M. De Martin Mazzalon, L. Martelli and M. Contemori for their help in performing single-station ambient vibration measurements and for providing geological information of the investigated areas.

REFERENCES

Albarello, D., Cesi, C., Eulilli, V., Guerrini, F., Lunedei, E., Paolucci, E., Pileggi, D. & Puzzilli, L.M., 2011. The contribution of the ambient vibration prospecting in seismic microzoning: an example from the area damaged by the 26th April 2009 L'Aquila (Italy) earthquake, *Boll. Geofis. Teor. Appl.,* **52**(3), 513–538.

Albarello, D. & Lunedei, E., 2011. Structure of an ambient vibration wavefield in the frequency range of engineering interest ([0.5, 20] Hz): insights from numerical modelling, *Near Surf. Geophys.,* **9,** 543–559.

Albarello, D., Socco, L.V., Picozzi, M. & Foti, S., 2015. Seismic hazard and land management policies in Italy: the role of seismic investigation, *First Break,* **33,** 79–85.

Bard, P.Y., 1999. Microtremor measurements: a tool for site effect estimation?, in *The Effects of Surface Geology on Seismic Motion*, pp. 1251–1279, eds Irikura, K., Kudo, K., Okada, H. & Sasatani, T., Balkema, Rotterdam.

Bindi, D., Parolai, S., Spallarossa, D. & Cattaneo, M., 2000. Site effects by H/V ratio: comparison of two different procedures, *J. Earthq. Eng.,* **4,** 97–113.

Bonnefoy-Claudet, S., Cornou, C., Bard, P.-Y., Cotton, F., Moczo, P., Kristek, J. & Fäh, D., 2006. H/V ratios: a tool for site effects evaluation. Results from 1-D noise simulations, *Geophys. J. Int.,* **167,** 827–837.

Bonnefoy-Claudet, S., Baize, S., Bonilla, L.F., Berge-Thierry, C., Pasten, C.R., Campos, J., Volant, P. & Verdugo, R., 2009. Site effect evaluation in the basin of Santiago de Chile using ambient noise measurements, *Geophys. J. Int.,* **176**(3), 925–937.

Bragato, P.L., Laurenzano, G. & Barnaba, C., 2007. Automatic zonation of urban areas based on the similarity of H/V spectral ratios, *Bull. seism. Soc. Am.,* **97**(5), 1404–1412.

Brandi, G.P., Fritz, P., Raggi, G., Squarci, P., Taffi, L., Tongiorgi, E. & Trevisan, L., 1967. Idrogeologia delle Terme di Montecatini, in *Collana Scientifica delle Terme di Montecatini*, Vol. 39, 50 pp., ed. delle Terme di Montecatini.

Capizzi, P., Martorana, R., D'Alessandro, A. & Luzio, D., 2015. Contribution of the cluster analysis of HVSR data for near surface geological reconstruction, in *Congress Proceedings 34 Convegno Nazionale GNGTS*, pp. 50–56, Geofisica Applicata, Atti-Tema 3, INOGS – Istituto Nazionale di Oceanografia e di Geofisica Sperimentale.

D'Amico, V., Picozzi, M., Baliva, F. & Albarello, D., 2008. Ambient noise measurements for preliminary site-effects characterization in the urban area of Florence, *Bull. seism. Soc. Am.,* **98**(3), 1373–1388.

Davis, J.C., 2002. *Statistics and Data Analysis in Geology,* 3rd edn, John Wiley & Sons, 656 pp.

Elter, P., Gratziu, C. & Labesse, B., 1964. Sul significato dell'esistenza di una unità tettonica alloctona costituita da formazioni terziarie nell'Appennino settentrionale, *Boll. Soc. Geol. It.,* **83**(2), 373–394.

Elter, P. & Marroni, M., 1991. Le Unità Liguri dell'Appennino Settentrionale: sintesi dei dati e nuove interpretazioni, *Mem. Descr. Serv. Geol. Ital.,* **XLVI,** 121–138.

Fäh, D., Kind, F. & Giardini, D., 2001. A theoretical investigation of average H/V ratios, *Geophys. J. Int.,* **145,** 535–549.

Farrugia, D., Paolucci, E., D'Amico, S. & Galea, P., 2016. Inversion of surface-wave data for subsurface shear-wave velocity profiles characterised by a thick buried low-velocity layer, *Geophys. J. Int.,* **206**(2), 1221–1231.

Field, E. H. & Jacob, K., 1993. The theoretical response of sedimentary layers to ambient seismic noise, *Geophys. Res. Lett.,* **20**(24), 2925–2928.

Foti, S., Parolai, S., Albarello, D. & Picozzi, M., 2011. Application of surface–wave methods for seismic site characterization, *Surv. Geophys.,* **32**(6), 777–825.

Gallipoli, M.R., Albarello, D., Mucciarelli, M. & Bianca, M., 2011. Ambient noise measurements to support emergency seismic microzonation: the Abruzzo 2009 earthquake experience, *Boll. Geofis. Teor. Appl.,* **52**(3), 539–559.

Haghshenas, E., Bard, P.Y. & Theodulis, N., 2008. Empirical evaluation of microtremor H/V spectral ratio, *Bull. Earthq. Eng.,* **6,** 75–108.

Jolliffe, I.T., 2002. *Principal Component Analysis,* 2nd edn, Springer, p. 487.

Kramer, S.L., 1996. *Geotechnical Earthquake Engineering,* Prentice Hall, 653 pp.

Kumar, S. & Narayan, J.P., 2008. Importance of quantification of local site effects based on wave propagation in seismic microzonation, *J. Earth Syst. Sci.,* **117** (S2), 731–748.

Lachet, C. & Bard, P.Y., 1994. Numerical and theoretical investigations on the possibilities and limitations of Nakamura's technique, *J. geophys. Earth,* **42,** 377–397.

Lazzarotto, A., Mazzanti, R. & Nencini, C., 1990a. Carta Geologica dei comuni di Livorno e di Collesalvetti in scala 1:25:000, *Quad. Museo di Stor. Nat. Livorno,* **11**.

Lazzarotto, A., Mazzanti, R. & Nencini, C., 1990b. Geologia e geomorfologia dei Comuni di Livorno e Collesalvetti, *Quad. Museo di Stor. Nat. Livorno,* **11**(Suppl. 2), 1–85.

Lunedei, E. & Albarello, D., 2010. Theoretical HVSR curves from the full wave field modelling of ambient vibrations in a weakly dissipative layered Earth, *Geophys. J. Int.,* **181,** 1093–1108.

Martelli, L. & Romani, M., 2013. Microzonazione Sismica e analisi della condizione limite per l'emergenza delle aree epicentrali dei terremoti della pianura emiliana di maggio-giugno 2012, relazione illustrativa, Servizio geologico, sismico e dei suoli Regione Emilia Romagna. Available at: http://ambiente.regione.emilia-romagna.it/geologia/archivio_pdf/sismica/MSord70_relazione.pdf/at_download/file/MSord70_relazione.pdf+&cd=3&hl=it&ct=clnk&gl=it.

Martelli, L. *et al.*, 2014. Analysis of the local seismic hazard for the stability tests of the main bank of the Po river (Northern Italy), *Boll. Geofis. Teor. Appl.,* **55** (1), 119–134.

Mucciarelli, M., 1998. Reliability and applicability of Nakamura's technique using microtremors: an experimental approach, *J. Earthq. Eng.,* **2** (4), 625–638.

Nakamura, Y., 1989. A method for dynamic characteristics estimation of subsurface using microtremor on the ground surface, *Q. Rep. Railw. Tech. Res. Inst.,* **30,** 25–33.

Panzera, F., Lombardo, G., Longo, E., Langer, H., Branca, S., Azzaro, R., Cicala, V. & Trimarchi, F., 2017. Exploratory seismic site response surveys in a complex geologic area: a case study from Mt. Etna volcano (southern Italy), *Nat. Hazards,* **86,** 385–399.

Paolucci, E., Albarello, D., D'Amico, S., Lunedei, E., Martelli, L., Mucciarelli, M. & Pileggi, D., 2015. A large scale ambient vibration survey in the area damaged by May-June 2012 seismic sequence in Emilia Romagna, Italy, *Bull. Earthq. Eng.,* **13**(11), 3187–3206.

Parolai, S., Picozzi, M., Richwalski, S.M. & Milkereit, C., 2005. Joint inversion of phase velocity dispersion and H/V ratio curves from seismic noise recordings using a genetic algorithm considering higher modes, *Geophys. Res. Lett.,* **32,** L01303, doi:10.1029/2004GL021115.

Picozzi, M., Parolai, S. & Richwalski, S. M., 2005a. Joint inversion of H/V ratios and dispersion curves from seismic noise: estimating the S-wave velocity of bedrock, *Geophys. Res. Lett.,* **32,** L11308, doi:10.1029/2005GL022878.

Picozzi, M., Parolai, S. & Albarello, D., 2005b. Statistical analysis of noise Horizontal to Vertical Spectral Ratios (HVSR), *Bull. seism. Soc. Am.,* **95** (5), 1779–1786.

Pieri, M. & Groppi, G., 1981. *Subsurface Geological Structure of the Po Plain,* Pubbl.414, PF Geodinamica, C.N.R. p. 23.

Preisendorfer, R.W., 1988. *Principal Component Analysis in Meteorology and Oceanography*, ed. Mobley, C.D., Elsevier, 425 pp.

Puccinelli, A., Verani, M. & Rossini, V., 2000. Nuovi dati sull'assetto idrogeologico dell'area termale di Montecatini Terme (Pistoia) e loro implicazioni nella pianificazione territoriale, *Quad. Geol. Appl.,* **7–3**(2000), 33–48.

Downloaded from https://academic.oup.com/gji/article/211/1/650/4060697 by guest on 10 April 2024

RER & ENI-Agip, 1998. *Riserve Idriche Sotterranee della Regione Emilia-Romagna*, ed. Di Dio, G.M., Regione Emilia-Romagna, Ufficio geologico—ENI-Agip, Divisione Esplorazione and Produzione.S.EL.CA., Firenze, p. 120.

Rodriguez, V.H. & Midorikawa, S., 2002. Applicability of the H/V spectral ratio of microtremors in assessing site effects on seismic motion, *Earthq. Eng. Struct. Dyn.,* **31**(2), 261–279.

Sanchez-Sesma, F.J. *et al*., 2011. A theory for microtremor H/V spectral ratio: application for a layered medium, *Geophys. J. Int.,* **186,** 221–225.

Scognamiglio, L. *et al*., 2012. The 2012 Pianura Padana Emiliana seismic sequence: locations, moment tensors and magnitudes. *Ann. Geophys.,* **55**(4), 549–559.

SESAME, 2004. Guidelines for the implementation of the H/V spectral ratio technique on ambient vibrations: measurements, processing and interpretation. Available at: http://sesame.geopsy.org/SES_Reports.htm or at: ftp://ftp.geo.uib.no/pub/seismo/SOFTWARE/SESAME/USER-GUIDELINES/SESAME-HV-User-Guidelines.pdf, last accessed August 2017.

SM Working Group, 2015. *Guidelines for Seismic Microzonation*. Conference of Regions and Autonomous Provinces of Italy—Civil Protection Department, Rome. Available at: http://www.protezionecivile.gov.it/httpdocs/cms/attach_extra/GuidelinesForSeismicMicrozonation.pdf?.

Theodoridis, S. & Koutroumbas, K., 2008. *Pattern Recognition,* Academic Press.

Trevisan, L., 1954. La nuova sorgente Leopoldina di Montecatini Terme e le condizioni geologiche del sottosuolo, *Boll. Ing. Fir. II,* 8–9.

Ullah, S., Bindi, D., Pittore, M., Pilz, M., Orunbaev, S., Moldobekov, B. & Parolai, S., 2013. Improving the spatial resolution of ground motion variability using earthquake and seismic noise data: the example of Bishkek (Kyrgyzstan), *Bull. Earthq. Eng.,* **11,** 385–399.

Wilks, D.S., 2006. *Statistical Methods in the Atmospheric Sciences,* 2nd edn, Academic Press, p. 630.

APPENDIX

We aim at identifying the matrix $[B]$ representative of linear transform of $[O']$:

$$[U] = [B]\left[O'\right], \tag{A1}$$

such that rows of $[U]$ are mutually uncorrelated. Being $[V_U]$ the variance/covariance matrix of $[U]$, one has:

$$[V_U] = [B][V_O][B]^T, \tag{A2}$$

where $[V_O]$ is the variance/covariance matrix of $[O']$ defined by eq. (2). Following eq. (3), one also has:

$$[V_U] = [B]\left([E][\Lambda][E]^T\right)[B]^T. \tag{A3}$$

As $[E]$ is an orthogonal matrix, that is, $[E]^{\mathrm{T}} = [E]^{-1}$, if $[B] = [E]^{\mathrm{T}}$ one has:

$$[V_U] = [\Lambda], \tag{A4}$$

which implies that the rows of $[U]$ are mutually uncorrelated. Thus, eq. (A1) can be expressed as:

$$[U] = [E]^T\left[O'\right], \tag{A5}$$

which allows computing the matrix $[U]$ of the PCs from the original centered data set $[O']$ and the respective variance/covariance matrix $[V_O]$. From the orthogonality of $[E]$ also descends that:

$$\left[O'\right] = [E][U]. \tag{A6}$$

Downloaded from https://academic.oup.com/gji/article/211/1/650/4060697 by guest on 10 April 2024