

Bindi, D., Zaccarelli, R., Razafindrakoto, H., Yen, M.-H., Cotton, F. (2023): Empirical shaking scenarios for Europe: a feasibility study. - Geophysical Journal International, 232, 2, 990-1005.

<https://doi.org/10.1093/gji/ggac382>

Empirical shaking scenarios for Europe: a feasibility study

D. Bindi¹, R. Zaccarelli¹, H. N. T. Razafindrakoto¹, M-H. Yen¹ and F. Cotton^{1,2}

¹GFZ-German Research Centre for Geosciences, Telegrafenberg, 14473 Potsdam, Germany. E-mail: bindi@gfz-potsdam.de

²University of Potsdam, Institute of Geosciences, 14476 Potsdam, Germany

Accepted 2022 September 28. Received 2022 July 6; in original form 2022 April 26

SUMMARY

We process a large number of seismic recordings in Europe (i.e. about half a million recordings from about 19 500 earthquakes) with the aim of decomposing the Fourier amplitude spectra into source, propagation and site effects. To account for first-order, large-scale regional differences in propagation effects, the spectral decomposition simultaneously solves six different models describing the spectral attenuation within different subregions. Since the decomposition approach is affected by trade-offs that make the solution non-unique, we assume a station installed on rock in Switzerland as reference station and we invert for relative site amplifications. To propagate the reference site condition to the entire data set, we develop a procedure based on a sequence of decompositions considering increasing and overlapping data sets. The applied procedure allows for a consistent evaluation of relative site effects for about 3200 station channels using a single reference station for the whole data set. Comparisons with site amplifications obtained in previous studies at common stations in Italy and Switzerland confirm the site amplification results. The target of this work is to show that the spectral models obtained for attenuation and site effects can be used to generate empirical shaking scenarios in the Fourier domain. Therefore, we conclude our feasibility study by presenting shaking maps generated at different frequencies for hypothetical magnitude 6.5 earthquakes with a Brune-type stress drop of 10 MPa located at different positions across Europe.

Key words:; Fourier analysis; Earthquake ground motions; Seismic attenuation; Site effects.

1 INTRODUCTION

In the last couple of decades, several geophysical disciplines have experienced radical improvements in different directions. On the one hand, looking at the case of seismology, the increasing density of digital networks in several regions of the world, the technological developments for the sensing units and data transmission (Hammond *et al.* 2019), the standardization of formats for archiving and exchanging waveforms and metadata (International Federation of Digital Seismograph networks, FDSN), have led to an exponential increase in data availability and a simplification of data sharing procedures. On the other hand, improvements in computing power together with the possibility of using large amount of high quality data to constrain the geophysical models have driven the development of physics-based simulations capable of capturing the details of ground shaking at frequencies approaching those of interest to engineering seismology (e.g. Graves *et al.* 2011; Roten *et al.* 2016; Whithers *et al.* 2018; Teng & Baker 2019; Whithers *et al.* 2019; Pitarka *et al.* 2022; Jiang *et al.* 2022). Thus, nowadays we are in the position of taking advantage of these two factors, large data availability and detailed numerical simulations. However, one or the other factor cannot fulfill the whole spectrum of ingredients needed for generating seismological models with high predictive power. For example,

despite the large amount of waveforms that each new large earthquake is generating in well-instrumented regions of the world, short distance recordings of large events are still lacking to fully capture the near-source radiation effects (e.g. Paolucci *et al.* 2021). From the modelling perspective, the lack of data near faults limits our improvements in understanding source processes (Ben-Zion 2019), and the level of detail required to simulate ground shaking at frequencies above a few Hz is not yet achievable. Because of these limitations, numerical modelling and empirical components complete each others in several approaches, such as commonly done for calibrating the high-frequency components of hybrid deterministic-stochastic simulations (e.g. Graves & Pitarka 2015; Lee *et al.* 2020; Razafindrakoto *et al.* 2021), or using empirical assessments of site effects (Roten & Olsen 2021; Hu *et al.* 2021) and attenuation (Olsen *et al.* 2018, 2020).

In this study, we present the processing of a large amount of seismological data disseminated through the EIDA—European Integrated Data Archive (Strollo *et al.* 2021) to capture source, propagation and site amplification contributions to ground shaking. The analysis is performed for the amplitude Fourier spectra (Bindi & Kotha 2020) and aims at defining a procedure for computing shaking scenarios for any given source spectra, considering empirical models for propagation and site effects. The latter can be also

considered as the contribution of empirical analysis to complete future numerical simulations for what regards site effects of high-frequency propagation effects. To develop our feasibility study, we first describe the data set used to perform the spectral decomposition, which covers a large portion of Europe. Then, we provide details about how the propagation effects are isolated from source and site contributions, also discussing different approaches proposed to parametrize the spectral attenuation curves in terms of geometrical spreading and anelastic attenuation that could be useful for future simulation studies. An ad-hoc strategy for this study is developed to isolate source and site contributions, taking care to propagate the *a priori* assumed reference site condition over the whole set of amplification functions. Finally, we conclude by showing the empirical shaking scenarios for hypothetical earthquakes occurring at different locations in Europe.

2 DATA AND PROCESSING

We use the *stream2segment* software (Zaccarelli *et al.* 2019) to extract windows containing earthquake recordings (hereinafter referred to as segments) from the continuous data streams stored in EIDA (Strollo *et al.* 2021). The download process is driven by a seismic catalogue created through the International Seismological Center (ISC) web service (<http://www.isc.ac.uk/fdsnws/event/1/query>). To optimize the data base compilation, multiple downloads are performed by splitting the region of interest into several subregions covering the overall area defined by the intervals 33°–71° N in latitude and –12° to 46° E in longitude. We select earthquakes occurred between 1990 January and 2021 May, with depths shallower than 60 km, and with minimum magnitude (as provided by the ISC FDSN compliant web service) varying from 2 in central Europe to 3.5 in the Aegean region.

We extract 4-min long segments, starting one minute before the theoretical *P* onset. Stations located up to 5° from the hypocentral location are queried, requiring only high sampling rate channels (i.e. the velocimetric channels HH, EH and the accelerometric channels HN, HL, HG). Following previous applications (Bindi *et al.* 2019), we use *stream2segment* to process the segments. In particular, we remove the instrumental response and we bandpass filter the segments considering an a-causal Butterworth filter with high-pass corner equal to 0.6, 0.2, 0.1 and 0.06 Hz for magnitudes lower than 3, between 3 and 5, between 5 and 5.7 and larger than 5.7, respectively. The low-pass corner frequency is set to 40 Hz. The selected signal windows correspond to the interval between the 2.5 and 97.5 percentiles of squared velocity cumulated starting from the estimated *P*-wave arrival time, tapered to zero at both ends using a 5 per cent cosine-profile. Spectra are smoothed by applying a triangular window with width equal to 5 per cent of the central frequency. The signal-to-noise ratio (SNR) with respect to pre-event noise is computed over 14 non-overlapping spectral windows, retaining for analysis only those spectra showing SNR larger than 6 over all windows. Further selections are performed for events with magnitude smaller than 4.5 by computing the anomaly score as defined by Zaccarelli *et al.* (2021), setting the threshold score to 0.75. We also analyse event- and station-specific standardized residual distributions for peak ground velocity and acceleration, considering different ground motion prediction equations (Bindi *et al.* 2017). Events and stations in the tails of the standardized distributions are visually inspected to further refine the selection.

Finally, events and stations with less than 3 spectra fulfilling the signal selection, and the few events located at latitudes higher than

57° N are not further considered. Fig. 1 shows the locations of the events and stations considered for spectral analysis, generating the magnitude–distance distribution shown in Fig. 2. In all, we analyse about half a million spectra (computed as the square root of the sum of the two horizontal components squared) relevant to about 19 500 earthquakes recorded by about 3200 station channels.

3 SPECTRAL AMPLITUDE DECOMPOSITION

We apply a spectral amplitude decomposition approach, known as generalized inversion (GIT, Castro *et al.* 1990), to isolate the source, propagation and site contributions to ground shaking in the Fourier domain. The GIT decomposition is based on the linear assumption that the Fourier amplitude spectra (FAS) of the considered recordings are given by the algebraic product (or convolution integral in time domain) between three terms representing the source excitation $S(f)$ at frequency f , the attenuation along source-to-receiver path $P(R, f)$, where R is the hypocentral distance, and the site amplification $Z(f)$, where the latter is mainly controlled by the subsurface geology. Considering a set of earthquakes with sources S_i , with $i = 1, \dots, N$ recorded by a network of stations characterized by site effects Z_j , with $j = 1, \dots, M$, the FAS $O_{ij}(f)$ of the corresponding recordings define, for each frequency f , the following linear system:

$$\log O_{ij}(f) = \log S_i(f) + \log P_{ij}(R, f) + \log Z_j(f) \quad (1)$$

Target of this study is the determination of the propagation and site terms to be used as empirical components in the context of the assessment of shaking scenarios. A detailed analysis of the source spectra will be the subject of a future dedicated study aiming at characterizing the distributions of source parameters for stochastic simulations in Europe. By analysing a redundant data set where the same station recorded several events at different distances and the same event was recorded by several stations, eq. (1) forms an overdetermined linear system that we solve in a least-squares sense (Koenker & Ng 2017). To remove trade-offs among source, propagation and site terms, it is necessary to introduce additional constraints as discussed in the following.

3.1 Attenuation step

We apply a multistep approach to isolate source, propagation and site terms. Following a previous spectral amplitude decomposition performed for the European strong-motion data set (Bindi & Kotha 2020), first we solve system (1) focusing on the propagation term $P(R, f)$. Since the determination of $P(R, f)$ for well-sampled data sets is expected not to be affected by the trade off between the source and site terms (Oth *et al.* 2011), we solve eq. (1) under the assumption that the average site amplification is 1 irrespective of frequency, and we impose a reference distance for the attenuation equal to 1 km (i.e. $P(R = 1, f) = 1$, irrespective of frequency). Moreover, we introduce a large-scale tectonic regionalization to account for possible lateral variations in propagation effects by considering 6 polygons: the central Mediterranean region (region 1), the Alpine region (region 2), Pyrenees (region 3), Iberia (region 4), central Europe and Balkans (region 5), and Aegean region (region 6) as shown in Fig. 1. The aim of the regionalization is to accommodate large-scale, first-order differences in the propagation effects that could bias source and site estimations. The six different attenuation models are simultaneously obtained by assigning each source–receiver ray path to the polygon comprising the longest portion computed by

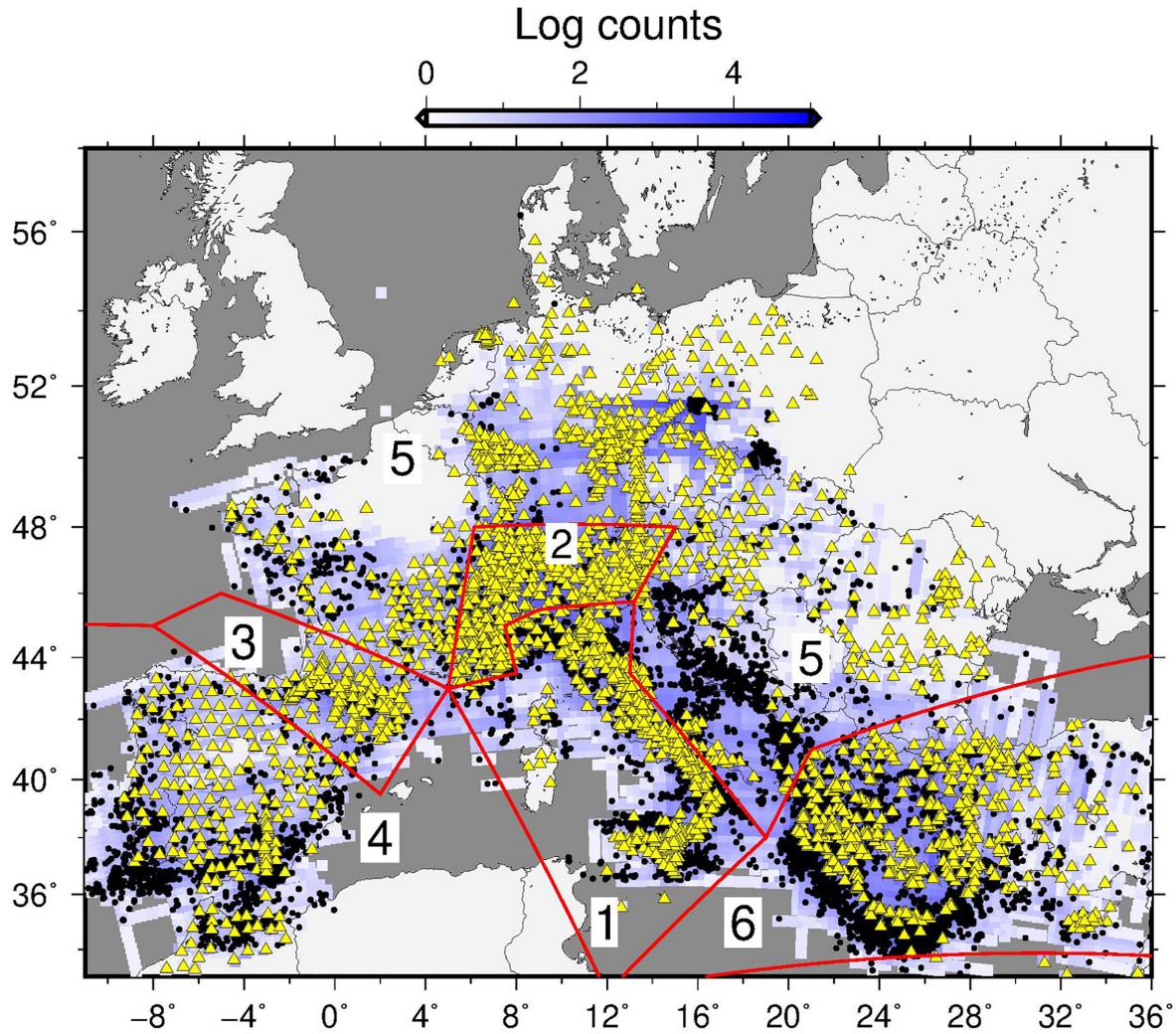


Figure 1. Map showing earthquake (black dots) and station (yellow triangles) locations used for running the spectral decomposition. The colour bar reports the number of source-to-station rays crossing a regular grid with cell size of 0.3° . Red polygons with labels from 1 to 6 indicate areas used to regionalize the spectral attenuation.

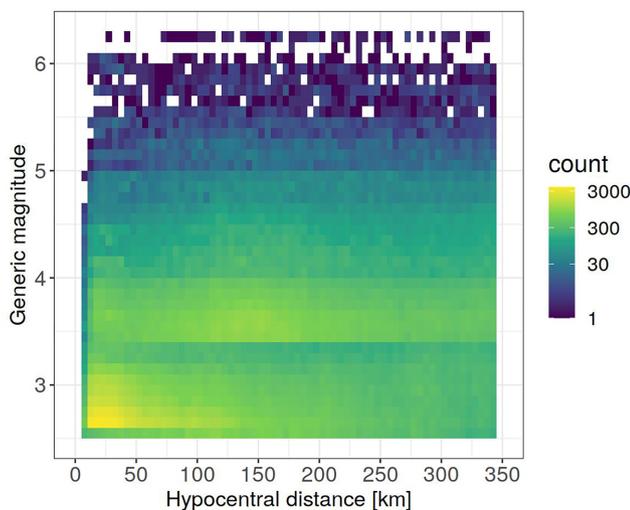


Figure 2. Magnitude versus hypocentral distance density plot computed for the data set used in this study to perform the spectral decomposition (see Fig. 1). The magnitude values are associated to different magnitude scales as obtained from ISC through the FDSN event query, mostly m_b and M_L .

approximating the rays with straight lines connecting the epicentres with the recording sites. For each region, a non-parametric spectral attenuation term is determined by discretizing the hypocentral distance range into intervals defined by nodes R_n , with $n = 1, \dots, L$ and applying a linear interpolation between two consecutive nodes:

$$\log P(R_{ij}, f) = a_n(f) \log P(R_n, f) + a_{n+1}(f) \log P(R_{n+1}, f) \quad (2)$$

where the hypocentral distance R_{ij} between event i and station j is such that $R_n \leq R_{ij} < R_{n+1}$ and the interpolation coefficients a_n are equal to $a_n = (R_{n+1} - R_{ij})/\Delta R$, $a_{n+1} = 1 - a_n$, with $\Delta R = (R_{n+1} - R_n)$. The attenuation obtained for the six regions differs in terms of the attenuation rate with distance and its frequency dependence (Fig. 3). A flattening of the attenuation rate is observed in several regions at distances between 60 and 100 km, probably a consequence of later arrivals due to reflections from crustal discontinuities, such as reflections from Moho (e.g. Burger *et al.* 1987; Chapman & Godbee 2012) whose traveltimes are controlled by several factors such as crustal thickness and hypocentral depth.

To evaluate the impact of introducing the regionalization on the spectral attenuation, the distance dependency of the residuals for three selected frequencies is investigated (Fig. 4). Residuals are computed as differences between the logarithm of the observed

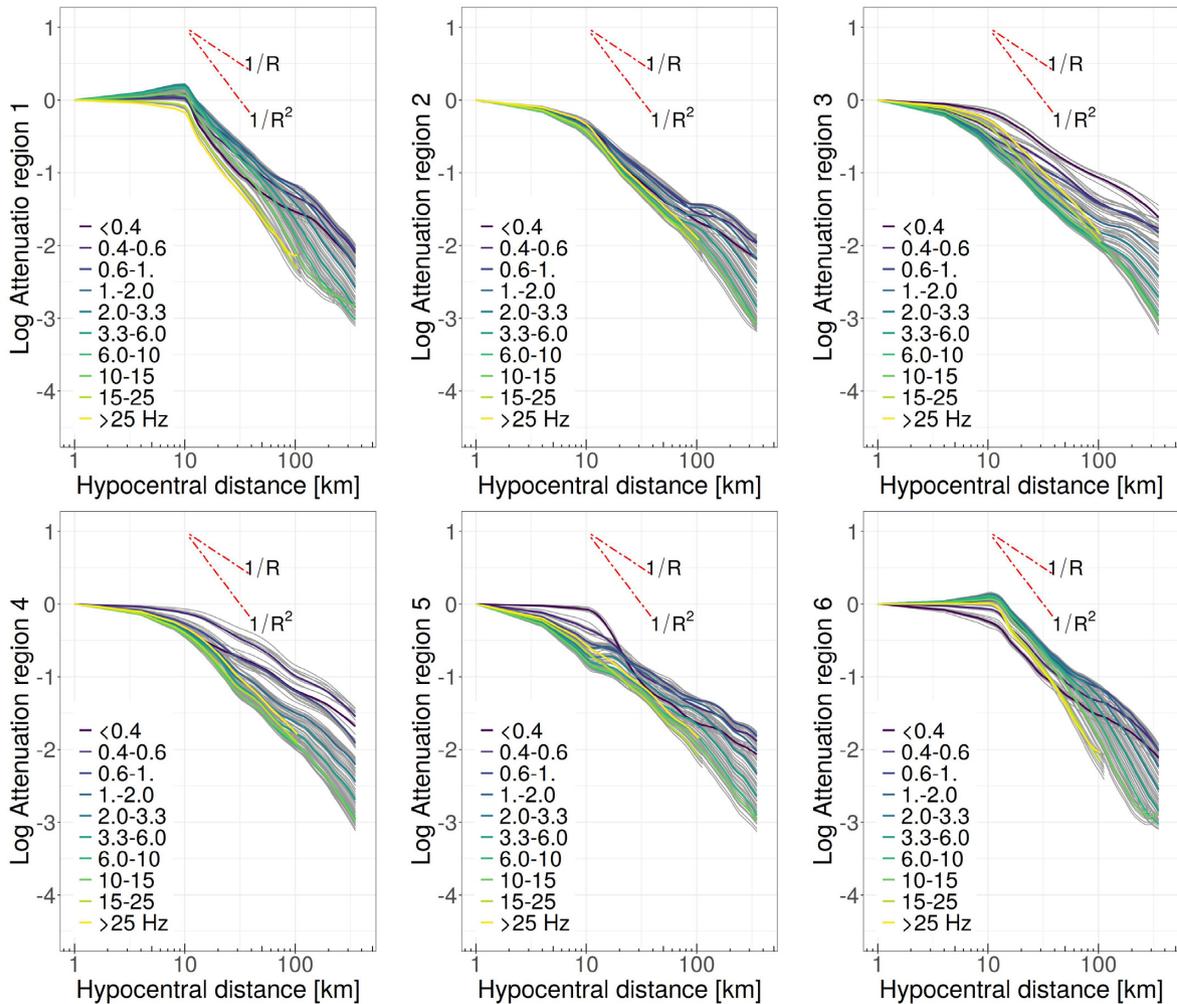


Figure 3. Non-parametric spectral attenuation against hypocentral distance for the six considered regions (Fig. 1). Coloured lines correspond to the attenuation curves (grey lines) averaged over the frequency intervals indicated in the legend. The attenuation rates proportional to the inverse of the distance and to the inverse of the distance squared are shown for reference (red lines).

spectra $\log O_{ij}$ and the values predicted by the constrained least-squares solution of eq. (1). The average and standard deviation of the residuals are computed over discretized distance intervals, grouping data according to regions 1–6. If the spectral decomposition is performed without considering the regionalization (panels a through c), residuals for several regions show a clear distance dependency, in particular at short distances and intermediate frequencies, with only regions 1 and 6 (i.e. those with the highest number of recordings and mostly controlling the solution) showing almost flat average residuals with distance (in particular at low and intermediate distances, panel b). Contrariwise, when a regionalization is introduced in the decomposition (panels d through f), an almost flat trend with distance is obtained for the average residuals within all regions, confirming the benefits of allowing the non-parametric attenuation to be different in different regions. The residual distributions for the six regions obtained with and without regionalization are exemplified for 3.2 Hz in Figs S1 and S2 of the Supporting Information.

3.2 Parametrization of the attenuation models

Although the empirical shaking scenarios discussed later in this article use the non-parametric attenuation shown in Fig. 3, a

parametrization of the spectral attenuation may be useful for future applications such those based on stochastic simulations. Therefore, we fit the non-parametric spectral attenuation curves $P(R, f)$ to a parametric model in which the geometrical spreading $G(R, f)$ and the anelastic attenuation $A(R, f)$ contributions are isolated, that is,

$$\log P(R, f) = \log G(R, f) + \log A(R, f) \quad (3)$$

where $\log G$ and $\log A$ are described in terms of parametric spectral models depending on $\log R$ and R , respectively. The separation of the geometrical spreading and anelastic attenuation terms is not unique, as discussed in several publications (Frankel *et al.* 1990; Chapman & Godbee 2012) and the correlation between $\log R$ and R generates a trade-off between the two components. Moreover, alternative model assumptions, such as those on the frequency dependency of $\log G$ and $\log A$, or about the depth dependency of the quality factor Q within the term $\log A$, produce solutions that have different physical interpretations but are almost equivalent in terms of how well they explain the observations (Edwards *et al.* 2008; Mitchell 2010; Zollo *et al.* 2014; Safarshahi & Morozov 2020). After some preliminary attempts (here not shown), we present in the following sections the results obtained by applying two different strategies to parametrize $\log P(R, f)$ for regions 1, 5 and 6. The attenuation for the other regions requires additional work to find a suitable parametrization

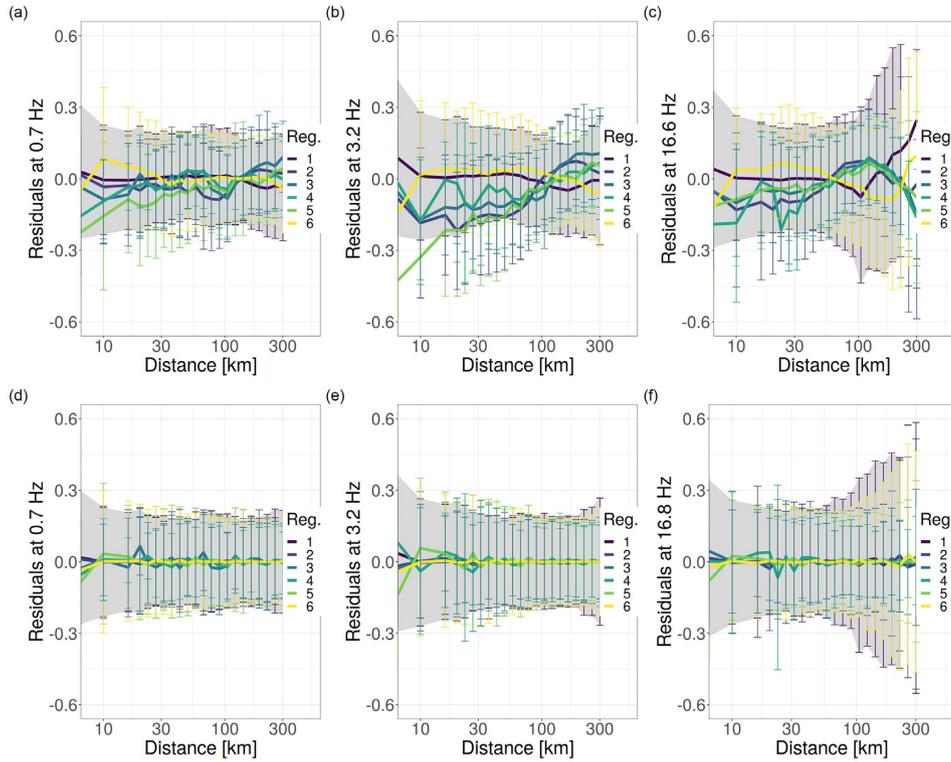


Figure 4. Residuals for the GIT decomposition obtained without considering (panels from a to c) and considering (panels from d to f) the regionalization into six polygons (Fig. 1), for three selected frequencies representing low, intermediate and high frequencies: 0.7 Hz (left), 3.2 Hz (middle) and 16.8 Hz (right). Coloured lines and vertical bars indicate the mean and the standard deviation of the residuals computed over discrete distance intervals, each colour corresponding to a specific region as indicated in the legend. The grey ribbon correspond to the mean \pm one standard deviation of region 1.

over the entire distance and frequency range, and it will be the subject of a future study.

3.2.1 Parametric attenuation for regions 1 and 6

For regions 1 and 6, we assume a frequency-independent geometrical spreading in eq. (3). Since, for distances above about 10 km, $P(R, f)$ has a shape more complex than a constant rate attenuation with distance (Fig. 3), we apply the following strategy to calibrate the parametric model:

(i) *Non-parametric geometrical spreading.* We select $P(f, R)$ for frequencies $1 \leq f \leq 1.5$ Hz and we fit the following model to their spectral average $\log \hat{P}(d)$:

$$\log \hat{P}(d) = \log G(d) + \log A(f = 1, d) \quad (4)$$

where $d = R/R_{ref}$, and $\log A(f = 1, d)$ accounts for the anelastic attenuation contribution evaluated at $f=1$ Hz, that is,

$$\begin{aligned} \log A(f, d) &= -\frac{\pi f^{1-\alpha}}{\ln(10)\beta Q_0} R_{ref}(d - 1) \\ &= -\frac{\pi}{\ln(10)\beta Q_0} R_{ref}(d - 1) \end{aligned} \quad (5)$$

where we set $\beta = 3.2$ km/s, $R_{ref} = 1$ km. In eq. (5), the quality factor $Q(f)$ is assumed to be frequency dependent and modelled as $Q(f) = Q_0 f^\alpha$.

(ii) *Parametric geometrical spreading.* In eq. (4), we consider a piecewise linear model $\log G(d)$ to accounts for changes in the geometrical spreading exponent over different distance ranges. For example, the geometrical spreading model used when imposing two

hinge distances d_1, d_2 corresponding to three slopes s_1, s_2, s_3 is the following:

$$\log G(d) = \begin{cases} s_1 d & \text{for } d \leq d_1 \\ s_1 d_1 + s_2(d - d_1) & \text{for } d_1 < d \leq d_2 \\ s_1 d_1 + s_2(d_2 - d_1) + s_3(d - d_2) & \text{for } d > d_2 \end{cases} \quad (6)$$

When the number of hinge distances is different from 2, eq. (6) is modified accordingly.

(iii) *Anelastic attenuation.* Finally, the propagation term $\log P(f, d)$ is corrected for the geometrical spreading (6) and the resulting $A(f, d)$ is analysed to estimate $Q(f)$ for the different regions.

The apparent geometrical spreading for the two regions are shown in Fig. 5 (panels a and b). We imposed 4 and 3 hinge distances for regions 1 and 6, respectively. A segmented regression (Muggeo 2003) is applied to estimate the optimal location of the hinge distances and the slopes of the piecewise linear model, and the results are summarized in Table 1.

The attenuation curves $\log A(f, d) = \log P(f, d) - \log G(d)$ are used to derive a parametric model for $Q(f)$. The $\log A(f, d)$ curves for the two regions are shown in Fig. 5 (grey lines in panels c and d). Since the shape of $\log A(f, d)$ clearly differs from a simple monotonic decay with distance, we allow $Q(f)$ to vary over different distance intervals, imposing the continuity of the attenuation at the boundaries between two intervals. In the case of one single hinge

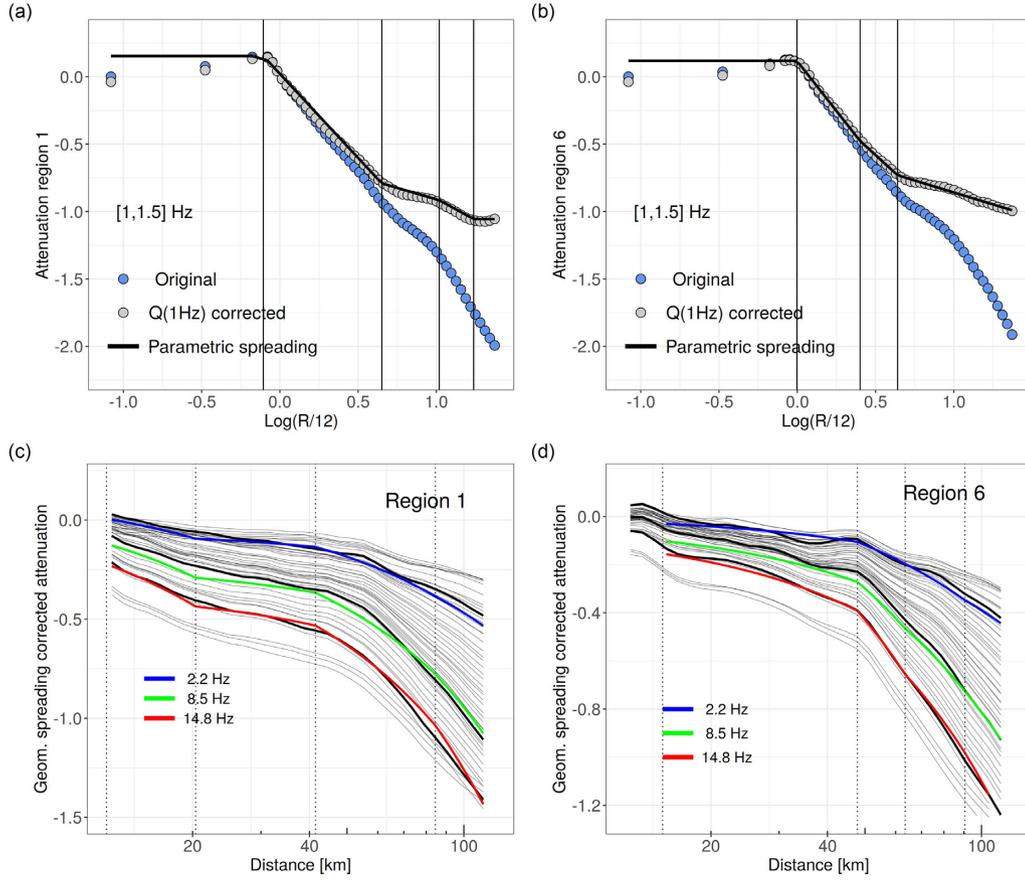


Figure 5. Parametrization of the attenuation for regions 1 (panels a and c) and 6 (panels b and d). Top: average attenuation curves $\log \hat{P}(R/12, f)$ computed considering frequencies in the range 1–15 Hz (blue circles); $\log \hat{P}(R/12, f)$ corrected for anelastic attenuation effects at 1 Hz (grey circles); best-fitting geometrical spreading models (black lines), with location of hinge distances (4 hinges for region 1 and 3 hinges for region 6) indicated by the vertical lines. Bottom: GIT attenuation curves $\log P(R, f)$ corrected for the geometrical spreading term $\log G(R)$ (i.e. $\log[P(R, f)/G(R)]$) for different frequencies between 1 and 20 Hz (thin black lines); curves for 2.2, 8.5 and 14.8 Hz are also shown as thick black lines and compared to the predictions from the final parametric model (coloured lines) given in Tables 1 and 2.

Table 1. Geometrical spreading models for regions 1 and 6 (e.g. see eq. 6). Hinge distances are indicated with h_j , with $j = 1, \dots, 4$; geometrical spreading exponents are indicated with s_i , with $i = 1, \dots, 5$.

Region	Slope	Error	Hinge [km]	Error [km]
Region 1				
s_1	0	Fixed		
s_2	-1.24	0.02	h_1	9.4
s_3	-0.34	0.04	h_2	53.7
s_4	-0.63	0.10	h_3	125.2
s_5	0	Fixed	h_4	207.6
Region 6				
s_1	0	Fixed		
s_2	-1.47	0.02	h_1	11.9
s_3	-1.06	0.06	h_2	30.2
s_4	-0.35	0.02	h_3	52.5

distance Rq_h , the model is the following:

$$\log A(R, f) = \begin{cases} -\frac{\pi f^{1-\alpha_1}(R-R_{ref})}{Q_{01}\beta}, & \text{for } R \leq Rq_h \\ -\frac{\pi f^{1-\alpha_1}(R_h-R_{ref})}{Q_{01}\beta} - \frac{\pi f^{1-\alpha_2}(R-R_{ref})}{Q_{02}\beta}, & \text{for } R > Rq_h \end{cases} \quad (7)$$

where $Q(f) = Q_{01}f^{\alpha_1}$ and $Q(f) = Q_{02}f^{\alpha_2}$ describe the quality factor for distances smaller and larger than Rq_h , respectively. When

more than one hinge distance is considered, eq. (7) is modified accordingly. For regions 1 and 6, we allowed Q to vary over four different distance ranges, as summarized in Table 2. The comparison between the non-parametric attenuation curves (black lines) and the parametric model exemplified for three frequencies (colored lines) in Fig. 5 (panels c and d) confirm that the parametric model (Tables 1 and 2) captures well the spectral attenuation characteristics within regions 1 and 6 over wide distance and frequency ranges.

3.2.2 Parametric attenuation for region 5

For region 5, we consider a parametrization where both the geometrical spreading and the quality factor Q vary with frequency, but without introducing a distance dependence on Q . To develop the parametric model, we follow these steps:

(i) we compute the average attenuation curves $\log \hat{P}(R, f)$ over different frequency ranges between 1 and 16 Hz (see first column in Table 3);

(ii) for each frequency interval, we fit the following model to $\log \hat{P}(R, f)$:

$$\log \hat{P}(R, f) = \log G(R/R_{ref}, f) + \eta(R - R_{ref}) \quad (8)$$

where $\log G(R/R_{ref}, f)$ is a piecewise linear model like in eq. (6) but with hinge distances and slopes varying with frequency; the

Table 2. $Q = Q_0 f^\alpha$ models for regions 1 and 6 (see eq. 7), considering different distance ranges.

Region	Frequency range [Hz]	Distance range [km]	Q_0	Error	α	Error
Region 1	[1,20]	[12,20.3]	48	3	0.61	0.03
		[20.3,41.5]	305	32	0.56	0.04
		[41.5,84]	98	2	0.63	0.01
		[84.5,120]	120	2	0.48	0.01
Region 6	[1,20]	[15,47.8]	294	12	0.40	0.02
		[47.8,63.5]	113	8	0.45	0.03
		[63.5,90.6]	105	4	0.60	0.02
		[90.6,120]	147.3	1	0.46	0.03

Table 3. Region 5: coefficients of eq. (8), shown as grey points in Fig. 7.

Frequency range [Hz]	Intercept	$\eta \times 10^{-3}$	s_1	s_2	s_3	s_4	h_1 [km]	h_2 [km]	h_3 [km]	R_{\min}, R_{\max} [km]
[1.0,1.5]	-0.59793	-1.09243	-0.49225	-1.01410	-0.34128	-0.71333	18.7	61.5	120.9	1,300
[1.5,2.0]	-0.59793	-1.27797	-0.55332	-1.07010	-0.30066	-0.78653	19.0	56.6	109.2	1,300
[2.0,2.5]	-0.65627	-1.64593	-0.59066	-1.04130	-0.24733	-0.77981	21.2	53.7	98.5	1,300
[2.5,3.0]	-0.67454	-1.83783	-0.62444	-1.04580	-0.34077	-0.79755	21.2	52.0	98.6	1,300
[3.0,3.5]	-0.77249	-2.03658	-0.66595	-1.03710	-0.36376	-0.80593	24.7	52.9	97.0	1,300
[3.5,4.0]	-0.77249	-2.09904	-0.68519	-0.95013	-0.41570	-0.80592	20.4	53.9	90.8	1,300
[4.0,4.5]	-0.74310	-2.06718	-0.72010	-1.00780	-0.45857	-0.86242	20.1	52.7	97.3	1,300
[4.5,5.0]	-0.74310	-2.16540	-0.75823	-1.04940	-0.45354	-0.88479	23.2	50.3	92.4	1,300
[5.0,5.5]	-0.79761	-2.31109	-0.75968	-1.10140	-0.44542	-0.87460	27.4	49.7	90.6	1,300
[6.0,7.0]	-0.83084	-2.60393	-0.80473	-1.16740	-0.52994	-1.07380	33.8	47.2	101.9	1,200
[7.0,8.0]	-0.83084	-2.95693	-0.78083	-1.33440	-0.47812	-0.98988	34.7	48.7	98.0	1,200
[8.0,9.0]	-0.87418	-3.35816	-0.76016	-1.30720	-0.50918	-0.87940	34.7	49.1	91.9	1,200
[9.0,10.0]	-0.85779	-3.54785	-0.76244	-1.31080	-0.55730	-0.90892	36.8	48.8	99.1	1,200
[10,12.0]	-0.81954	-3.74821	-0.72884	-1.29250	-0.45976	-0.71506	33.6	52.3	73.7	1,200
[12,14.0]	-0.82909	-3.98826	-0.44394	-1.56500	-0.46331	-0.85616	5.3	10.7	16.5	1,200
[14,16.0]	-0.86782	-3.98826	-	-1.62010	-0.54815	-1.13330	-	10.9	19.8	5,200

coefficient η accounts for the anelastic attenuation and varies with frequency;

(iii) finally, the obtained frequency-dependent quantities (i.e. η , hinge distances and geometrical spreading slopes) are fit with simple parametric models to describe their dependency on frequency.

Fig. 6 exemplifies the procedure and shows the results obtained for three selected frequencies. Panels from (a) to (c) present the different terms of eq. (8): the attenuation with distance averaged over three frequency intervals $\log \hat{P}(R, f)$ (blue points); $\log \hat{P}(R, f)$ corrected for the η term (grey points); the best-fitting geometrical spreading $\log G(R/R_{\text{ref}}, f)$ (black lines). The obtained frequency-dependent η , hinge distances h_1, \dots, h_3 , and slopes s_1, \dots, s_4 are shown in Fig. 7 as grey points, along with the final parametric models (black lines) introduced to parametrize their frequency dependency. The comparisons between the non-parametric attenuation at 2, 5.2 and 11.6 Hz (blue points) with the predictions from the parametric model (red lines) are shown in Fig. 6, panels from (d) to (f), obtained considering the final parametrization shown as black lines in Fig. 7. The good match between the non-parametric GIT results and the predictions confirms the suitability of the parametric model.

4 SOURCE-SITE DECOMPOSITION

FAS corrected for the propagation term $P(R, f)$ are used to split the source and site terms. Since GIT can only provide solutions relative

to the constraints applied to remove the trade-offs, we isolate the site and source terms by applying an ad-hoc strategy. We first run an inversion by considering only stations belonging to the network CH managed by Swiss seismological service (SED) and recordings associated to region 2. We select as reference station CH.LLS (station Linth-Limmern, <https://stations.seismo.ethz.ch/en/station-information/current-stations/>), installed on rock with shear wave velocity averaged over the uppermost 30 m equal to $v_{s,30} = 2925 \text{ m s}^{-1}$ (Fäh *et al.* 2009). The site amplification at the reference station LLS is constrained to assume values equal to 1 for frequencies f below 10 Hz and to the function $\exp[-0.015\pi(f-10)]$ above 10 Hz, to account for near-surface attenuation effects at high frequencies (Anderson & Hough 1984).

4.1 Source-site decomposition: iterative procedure

After computing the site amplification for CH network using data in region 2, we perform a sequence of decompositions by progressively enlarging the data set at each step. Fig. 8 exemplifies the procedure by providing snapshots of the stations analysed at three different iterations: at each step, we extend the data set by adding data from a new region or a new network (green triangles); then, we repeat the decomposition over the enlarged data set but constraining the site amplification of stations already processed to the solution obtained in previous steps (red triangles). Following this approach,

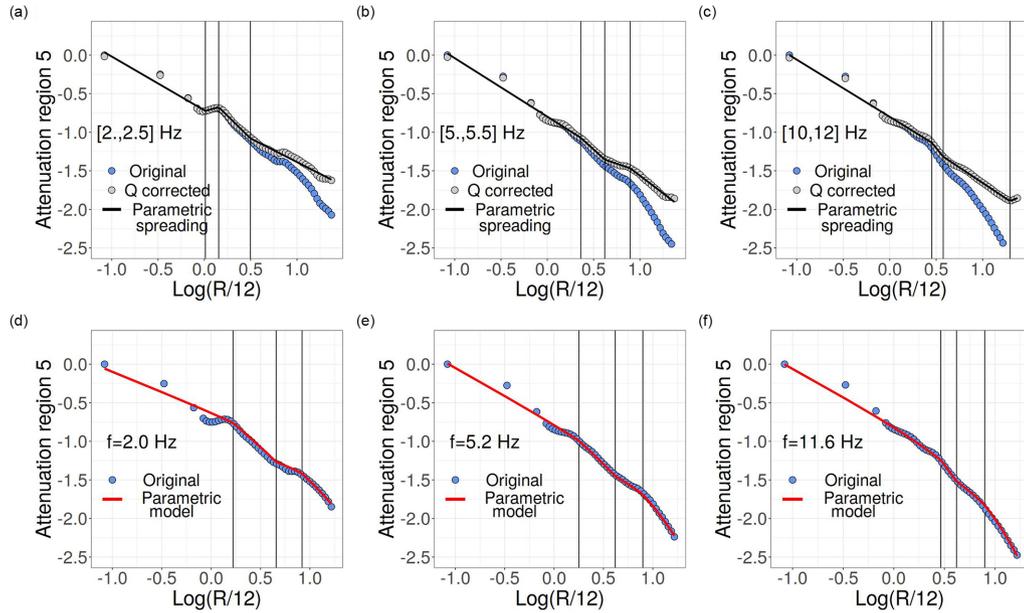


Figure 6. Analysis of the attenuation results in region 5. Panels (a)–(c): average attenuation curves $\log \hat{P}(R/12, f)$ (blue circles) computed considering frequencies in the range (a) 2–2.5 Hz, (b) 5–5.5 Hz and (c) 10–12 Hz; $\log \hat{P}(R/12, f)$ corrected for the anelastic attenuation contribution (η term in eq. 8) is shown as grey circles; the best-fitting geometrical spreading (black curves), with hinge distances indicated by the vertical lines. Panels (d)–(f): comparison between the GIT non-parametric attenuation term $\log P(R/12, f)$ (blue circles) and the prediction from the final parametric model (red), evaluated for (d) $f=2$ Hz, (e) $f=5.2$ Hz and (f) $f=11.6$ Hz. The parametric model is shown with black lines in Fig. 7.

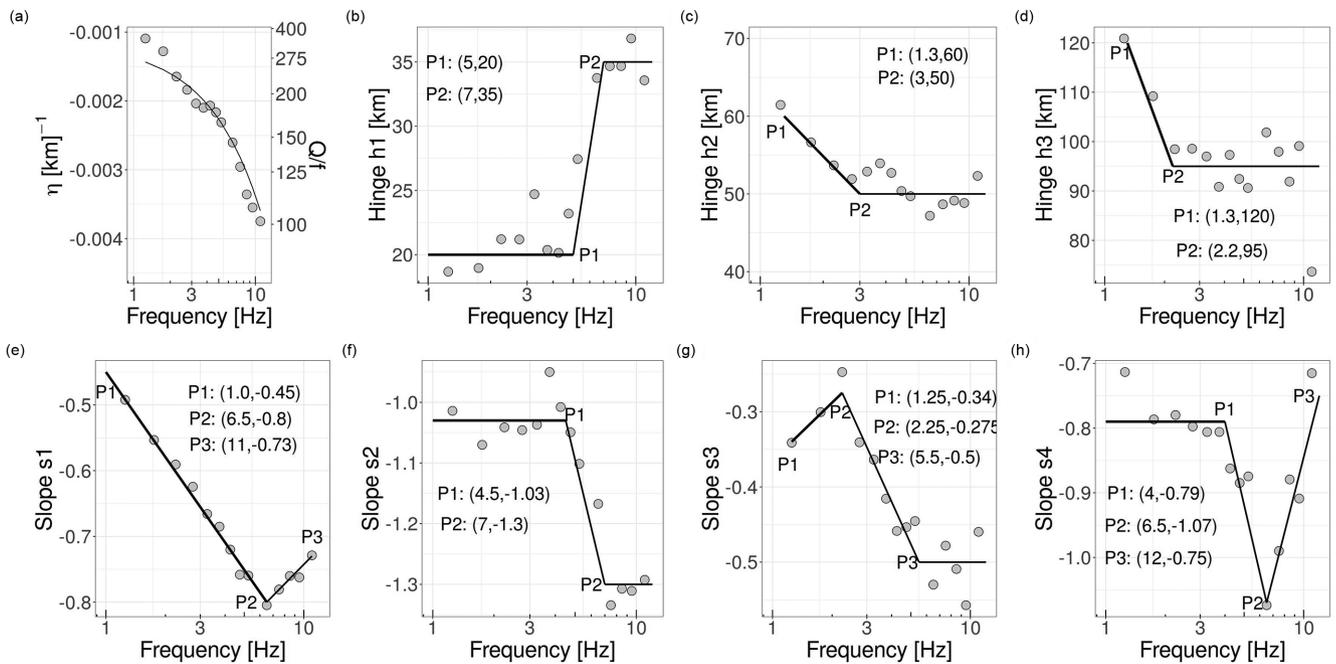


Figure 7. Parametric attenuation model for region 5. (a) Anelastic attenuation coefficient η ; (b) through (d): hinge distances; (e) through (h): geometrical spreading exponents. Grey circles represent the results of the fit performed over different frequency intervals considering a piecewise linear geometrical spreading (values listed in Table 3, see also the examples shown in the top panels of Fig. 6); the final parametric model is represented by the black lines (the break point locations are provided in the panels), where the best-fitting model for η (panel a) is $\eta = -1.157 \times 10^{-3} - 2.215 \times 10^{-4} f [\text{km}]^{-1}$, valid for $1 \leq f \leq 15$ Hz.

we propagate the original reference site condition to the entire data set. Fig. 9 shows the distributions of the obtained site amplifications.

Fig. 10 compares the site amplifications obtained at four Italian stations with the results of a previous study (Bindi & Kotha 2020)

performed over a different data set. In particular, Bindi & Kotha (2020) decomposed the Fourier spectra disseminated by the Engineering Strong Motion service (Lanzano *et al.* 2019) by constraining to 1 the average amplification of a set of *a priori* selected stations

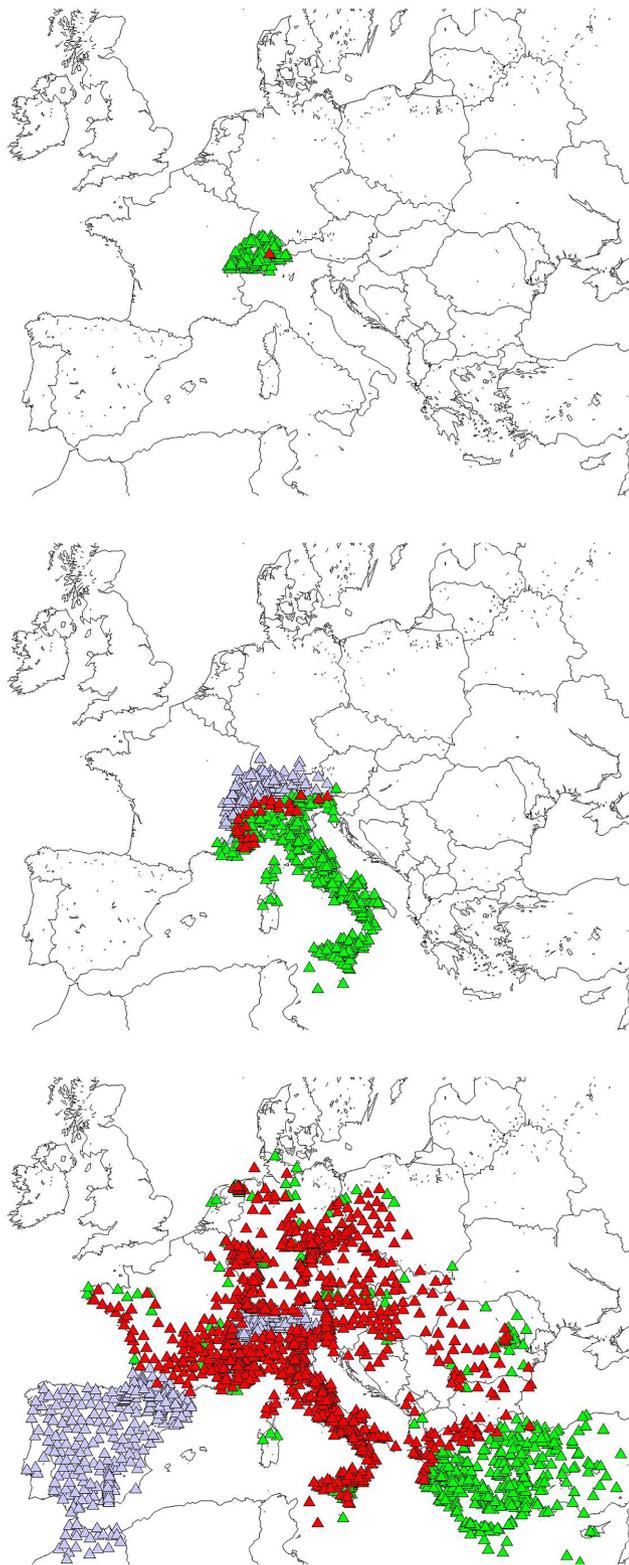


Figure 8. Example of three steps of the iterative procedure followed to separate source and site terms. Red and green triangles indicate stations contributing to the decomposition performed at a given iteration, with site amplifications for the red stations constrained to solutions obtained in previous steps; and the blue triangles indicate the location of stations already processed in the previous steps but not contributing to the specific iteration shown.

installed on rock. Since the site amplifications are relative to different reference site conditions, differences among the amplitudes of the results shown in Fig. 10 are expected. To partially homogenize the reference site conditions between the two studies, the results by Bindi & Kotha (2020) are multiplied by near-surface attenuation term above 10 Hz considering $k_0=0.015$ s as imposed in the GIT decomposition to the reference station *CH.LLS*. There is an overall agreement between the site amplifications provided by the two studies, with amplitude differences within a factor 2. In particular, the spectral shapes and the location of the resonance peaks f_0 agree very well (e.g. $f_0=3$ Hz for station MURB, Monte Urbino; $f_0=7$ Hz for station Norcia, NRCA). A further validation of the retrieved site amplifications is shown in Fig. 11, where the results for six stations in Switzerland (belonging to the *CH* network managed by SED) are compared with the results by Edwards *et al.* (2013). Also for the *CH* stations, the two studies used different data sets, imposed different reference site conditions and, in this case, followed different decomposition strategies since (Edwards *et al.* 2013) developed a parametric decomposition approach. The amplifications used in Fig. 11 for comparison are those obtained from *SED*, considering the results including the k_0 effect. Despite the differences in the amplitude due to the different reference site condition applied, the spectral amplifications obtained in this study shown in panels (a) to (f) of Fig. 11 (red and blue curves for co-located strong motion and broad-band sensors, respectively) are in good agreement with the *SED* estimates (black curves). The strong consistency of the results is highlighted in panels (g)–(i) which exemplify the relative amplification of three stations with respect to the amplification at station ZUR-Zurich shown in panel (f).

5 EMPIRICAL SCENARIOS

The site amplifications and the attenuation models provided by the spectral decomposition are used to compute empirical shaking scenarios. To this purpose, we process the retrieved site terms $Z(f)$ in order to generate amplifications maps. Although site effects are controlled by local geological conditions that can vary over a short spatial scale (tens to hundred meter scale), for our feasibility study we develop scenarios at a more regional scale (from kilometres to tens of kilometres scale). Therefore, we interpolate spatially the spectral amplifications to create maps over a regular grid with spacing 0.15° . We grid the randomly spaced amplification values using adjustable tension continuous curvature splines (Wessel *et al.* 2013), after applying a pre-processing steps to compute the median amplification within cells sampled by several stations. Fig. 12 exemplifies the obtained amplification maps at 1 Hz whereas the maps for 0.6 and 6.35 Hz are shown in Fig. S3 of the Supporting Information. The top panel of 12 shows the amplification at the actual stations locations whereas the bottom one presents the outcome of the spatial interpolation, where large areas not covered by data are masked to avoid excessive unconstrained interpolations.

Before using the amplification maps to compute shaking scenarios, we perform a consistency check by comparing the distribution of the numerical shaking values obtained for a given source spectrum with both observed values and predictions from a parametric ground motion model (GMM). We first fit the whole distribution of the observed spectral values with the following parametric model:

$$\log O(f) = e_1 + e_2(M - 3) + e_3 \log R + e_4 R + \delta S2S + \delta Be + \epsilon \quad (9)$$

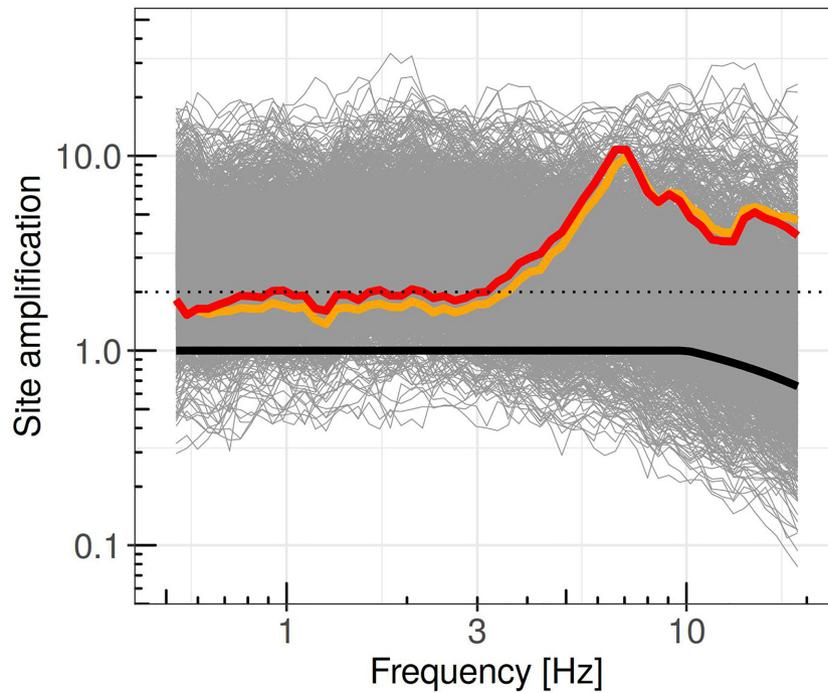


Figure 9. Site amplifications obtained in this study. Results for station NRCA (Norcia, IV network) are shown in colour (red for the strong motion sensor, orange for the broad-band one) whereas the amplification at the reference station LLS (CH network) is shown in black.

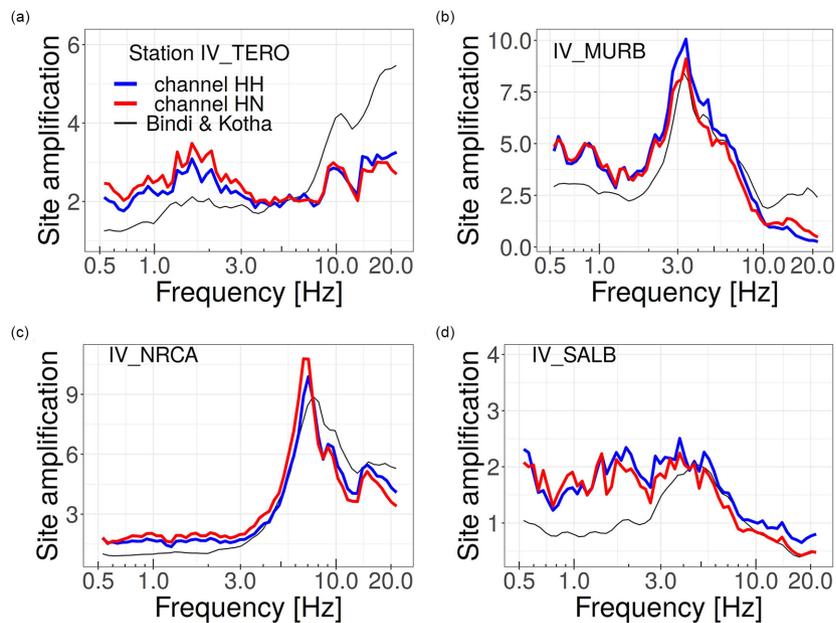


Figure 10. Comparison of the site amplifications obtained in this study for four stations of the IV network with the results of Bindi & Kotha (2020): (a) TERO station (Teramo); (b) MURB station (Monte Urbino); (c) NRCA station (Norcia) and (d) SALB station (San Lorenzo Bellizzi). The red and blue lines indicate the results obtained for the strong motion (HN) and broad-band (HH) channels, respectively; the black lines are the results of Bindi & Kotha (2020) multiplied by $e^{-\pi 0.015(f-10)}$ above 10 Hz.

where the FAS values O are evaluated at frequency f ; M and R are the catalogue magnitude and the hypocentral distance, respectively; $\delta S2S$ and δBe are the inter-station and inter-event residuals entering in model (9) as random effects (Kotha *et al.* 2022) for the station and earthquake grouping levels, respectively; ϵ are the left-over residual and e_k with $k = 1, \dots, 4$ are the coefficients of the

calibrated GMM that generate the median predictions. The consistency check is performed by considering a magnitude 3 earthquake with epicentral coordinates (7.6°E, 47.5°N), using a Brune source shape with stress drop $\Delta\sigma = 1$ MPa. Fig. 13 shows the spectral amplitudes for $f=1.04$ Hz (left) and 6.25 Hz (right) considering events with magnitude in the range 2.9–3.1 (black circles). These

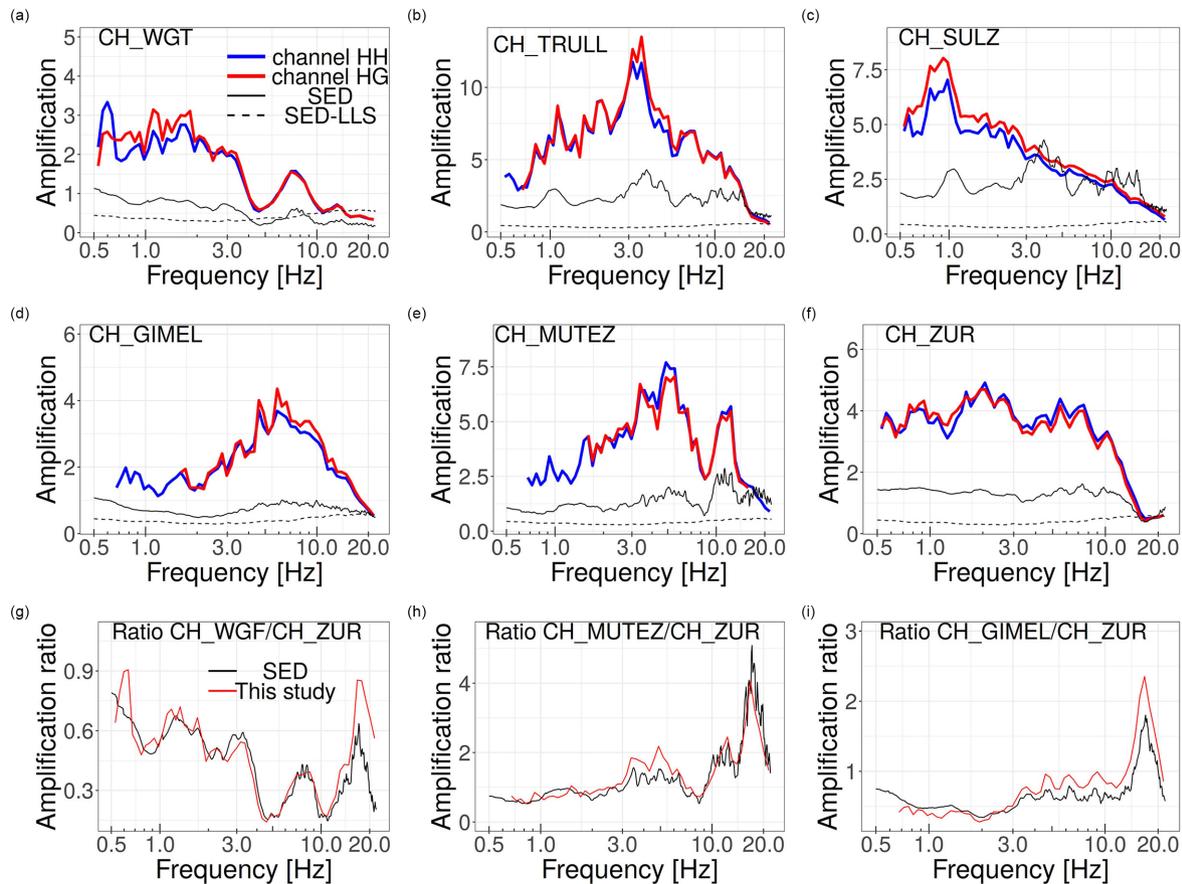


Figure 11. Comparison of the site amplifications obtained in this study for six stations of the CH network with the results of Edwards *et al.* (2013): (a) WGT station (Wägital); (b) TRULL station (Truellikon); (c) SULZ station (SULZ); (d) GIMEL station (St. Georges, Gimel); (e) MUTEZ station (MuttENZ) and (f) ZUR station (Zurich). The red and blue lines indicate the results obtained for the strong motion (HN) and broad-band (HH) channels, respectively; the black lines are the results of Edwards *et al.* (2013). Panels (g)–(i) show the amplification ratios obtained for WGF, MUTEZ and GIMEL stations compared to ZUR, considering the results of Edwards *et al.* (2013) (black) and this study (red).

values are compared with the variability (grey ribbon) associated to predictions from eq. (9), where the GMM variability is assessed by adding and subtracting ϕ to the median prediction obtained for $M = 3$. The intra-event variance ϕ^2 is computed by summing together the variances of the $\delta S2S$ and ϵ distributions. The scenario shaking values computed by combining the source spectra with the GIT attenuation models and the interpolated site amplifications are shown as white circles. The values of $\delta S2S$ of station *CH.LLS* is used to scale the GMM and GIT scenarios to a common reference. Fig. 13 confirms that there is an overall consistency among the scenario shaking scenario and the GMM results; the scenario variability slightly underestimates the GMM within-event variability, probably as consequence of the smoothing applied to compute the amplification maps, and of the limited region considered for generating the scenario.

Finally, the amplification maps are combined with the regional attenuation models to compute the spectral shaking for some selected source scenarios. The results of the procedure implemented for computing empirical shaking scenarios are shown in Fig. 14. We compute the acceleration synthetic scenarios generated by an M_w 6.5 earthquake with different locations (i.e. in central Europe, north-eastern Italy and southern Italy). To exemplify the outcomes from this feasibility study, we generate the source spectrum considering a Brune source model (Brune 1970) with stress drop $\Delta\sigma = 10$

MPa. The obtained shaking is shown in Fig. 14, where the spatial patterns in the maps are controlled mainly by the large scale spatial variability of the site amplifications (Fig. 12) and by the different combinations of the attenuation models derived for regions 1, 2 and 5.

6 DISCUSSION

We have shown that the access through EIDA to a large volume of seismic data recorded in Europe allowed us to compute the ingredients needed for developing empirical shaking scenarios. The development of the models for site amplification and attenuation took advantage from the standardization of the formats used for both archived data (waveforms) and their associated metadata (e.g. station inventories), along with the possibility to retrieve such information through standardized web services (e.g. compliant to FDSN standards). The followed approach guarantees the level of harmonization needed to compute, and compare, shaking scenarios generated by hypothetical seismic sources located in different regions of Europe. In particular, the achieved harmonization allows us to mitigate problems of inconsistency arising when results from different studies are merged, for example inconsistencies due to different reference conditions applied to site amplifications. Since the

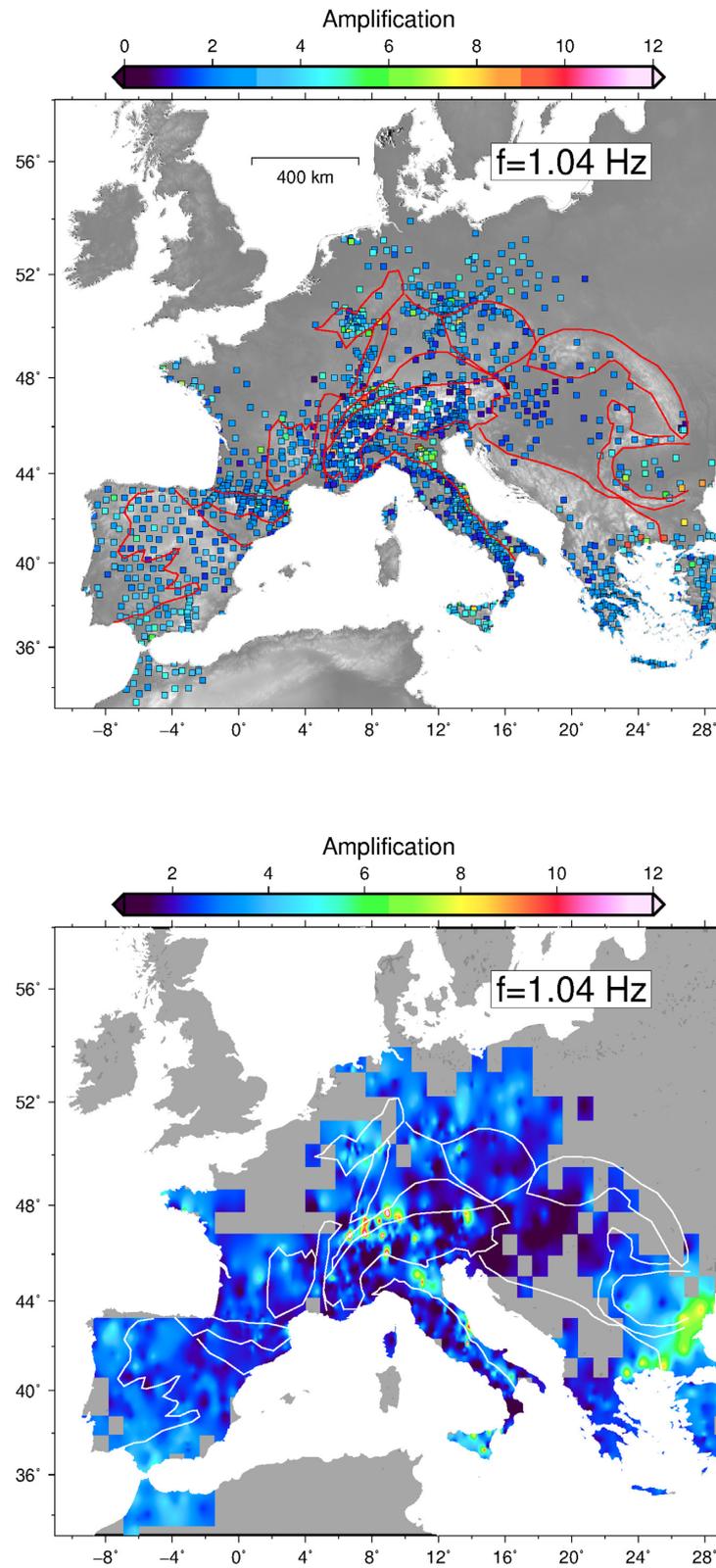


Figure 12. Site amplification values (top) and interpolated map (bottom) at 1.04 Hz.

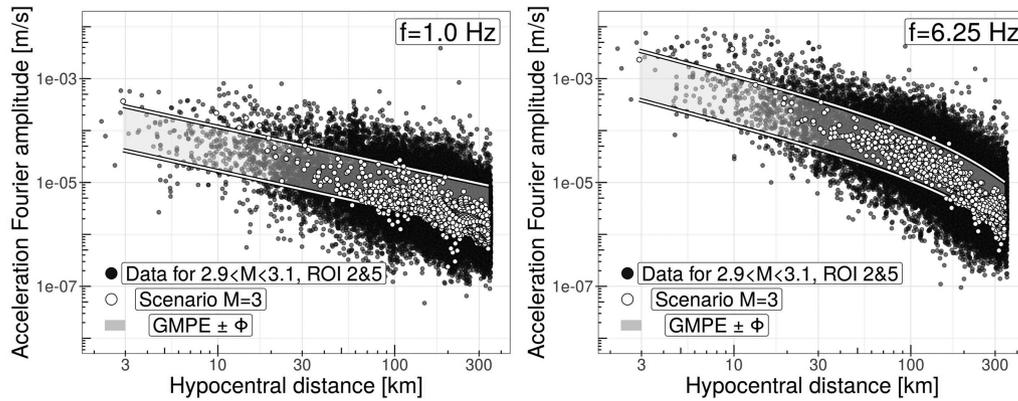


Figure 13. Consistency check between: observed spectral values at 1 Hz (left) and 6.25 Hz (right) selecting events in regions 2 and 5 with magnitude between 2.9 and 3.1 (black circles); median predictions \pm one intra-event standard deviation ϕ considering the GMM in eq. (9); empirical predicted values using the GIT attenuation models and the interpolated site amplifications, for a Brune source model with magnitude 3 and stress drop 1 MPa located at (7.6°E,47.5°N) (white circles).

analysis performed in this study proved to be successful in transforming the wealth of available data into elements for computing empirical shaking scenarios, several aspects touching the different components of the decomposition deserve further developments. About the site term, for developing our concept through the presented feasibility study, we have interpolated the punctual estimates at the station locations over a regular grid with spacing 0.15°. Indeed, for regions where dense networks are operating, a higher spatial resolution can be achieved. Therefore, the interpolation schema can be modified into an adaptive grid with respect to the available data density. Furthermore, data from networks not disseminated through EIDA can be considered to densify the sampling in several regions. For example, strong motion networks used for engineering applications can be added (e.g. the Italian Civil Protection strong motion network RAN). The decomposition approach followed in this study can be applied to easily integrate new data without the need of repeating the whole process. Recordings from an additional set of stations can be processed together with recordings from a set of stations already considered in this study and installed in the same area, provided that some common earthquakes have been recorded by both set of stations. The new decomposition can use the attenuation model already calibrated and, for isolating source and site terms, amplifications at those stations already processed can be constrained by following the approach presented in this study, which allow the propagation of the reference site condition to the new stations. The interpolation of the empirically assessed site amplification can benefit from the usage of additional layers providing information about the spatial variability of parameters correlated to site amplifications. For example, Weatherill et al (2020) have recently shown the feasibility of deriving amplification maps at a regional scale in Japan starting from site amplifications estimated at the station locations and using mappable proxies such as topographic slope and geology. It is also worth noting that, when end-to-end approaches that predict ground shaking directly from waveforms (Jozinović et al. 2020; Münchmeyer et al. 2021) or from knowledge of the event magnitude and location (Lilienkamp et al. 2022) are of interest for early-warning and rapid response actions, deep-learning approaches have proven to be very effective (Florez et al. 2022). In data-rich regions, supervised machine learning approaches have been used to generate ground motion prediction map. For example, Mori et al (2022) took advantage of the availability in Italy of high-density

microzonation data along with high resolution geophysical (e.g. shear wave velocity averaged in the uppermost 30 m) and morphological (e.g. elevation, first- and second-order topographic gradients) to calibrate a machine learning model to predict the ground motion at the high resolution spatial scale of about 50 m, including local site effects with reasonable spatial-correlation structure. Contrariwise to end-to-end approaches, GIT allows to isolate the different physical layers contributing to ground shaking allowing, in turn, to use them as ingredients for numerical simulation or for producing labelled data sets suitable for training machine learning approaches.

Along with the site term, also the resolution of the attenuation model can be improved by either introducing more subdomains where, for example, residuals analysis show the presence of large-scale spatially coherent patterns, or developing alternative approaches to capture the propagation effects (e.g. merging decomposition and tomographic approaches for attenuation). Finally, we have shown the potentiality of the approach by generating shaking scenarios for point sources (Fig. 14) but source models more suitable for moderate to large events can be used to generate the empirical scenarios. In particular, the empirical assessment of propagation and site terms can be embedded within methodologies for simulating the ground shaking at high frequencies using stochastic approaches and accounting for finite source effects (Boore 2009). In terms of propagation models, either the non-parametric GIT results can be used in the stochastic simulations as tables describing the regional attenuation with distance for different frequencies, or using the results of one of the parametrization approaches tested in this study to capture the frequency-dependent changes in the rate of attenuation with distance.

7 CONCLUSIONS

We processed a large volume of European seismic data (i.e. about half a million recordings from about 19 500 events recorded by 3200 station channels) with the aim of determining harmonized descriptions of source, propagation and site effects in the Fourier domain. Contrariwise to end-to-end deep-learning based approaches that proved to be effective for generating prediction maps directly

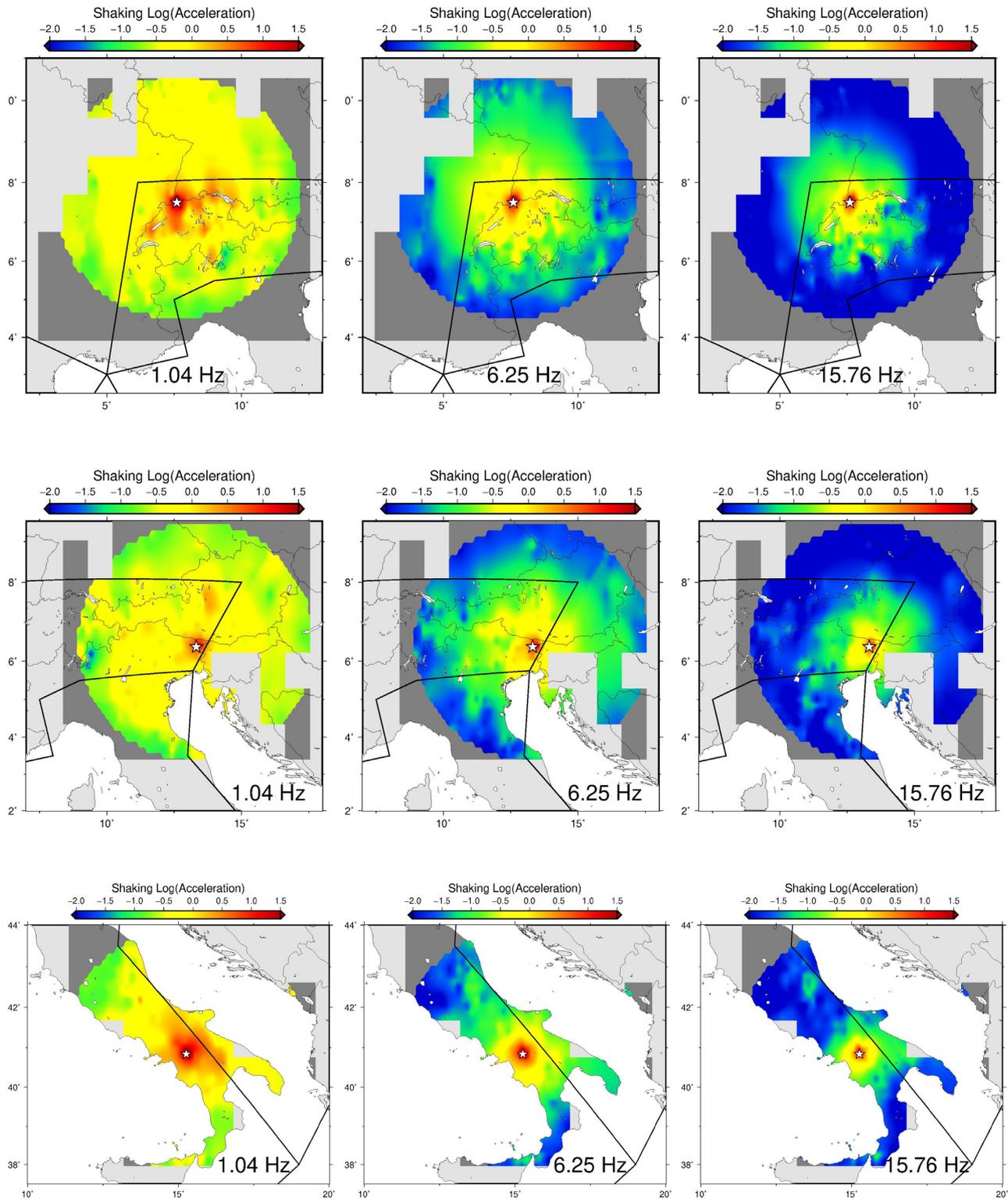


Figure 14. Example of acceleration shaking scenarios at 1.04 Hz (left), 6.25 Hz (middle) and 15.76 Hz (right) for three hypothetical earthquakes occurring in central Europe (top), north eastern Italy (middle) and southern Italy (bottom). The source spectrum is given by a Brune spectrum evaluated for M_w 6.5 and stress drop $\Delta\sigma=10$ MPa.

from seismic waveforms or event metadata, the spectral decomposition provides physical terms that can be used as ingredients, or constraints, for generating shaking maps or performing stochastic simulations. The application of the decomposition results have been exemplified by generating empirical shaking scenarios for

hypothetical earthquakes with different locations in Europe, including site amplifications. The computations can be generalized to generate portfolio of predicted shaking maps relevant to different source scenarios (e.g. for different stress drop values, different source spectral shapes, different finite-fault effects) that, in turn,

can be used either for evaluating the impact of future earthquakes in hazard-oriented studies or to provide hints about the source properties for rapid response actions after the occurrence of an earthquake by comparing observed shake maps (Wald *et al.* 2019) with those in the portfolio. Along with numerous improvements that can be achieved for the different terms of the decomposition as discussed in this study, future efforts will be made in the direction of integrating the proposed approach with physics-based simulations.

DATA AND RESOURCES

We used data and information downloaded from the following sites: EIDA (<https://www.orfeus-eu.org/data/eida/>); ISC (<http://www.isc.ac.uk/>); FDSN (<https://www.fdsn.org/networks/>). A list of the network DOIs is provided in the Supporting Information (Table S1). The derivation of the models was performed using R software (R Core Team 2020) and, in particular, lme4 (Bates *et al.* 2015), dplyr (Wickham *et al.* 2018), ggplot2 (Wickham 2016), Segmented (Muggeo 2003), Matrix (Bates & Maechler 2019) and sparseM (Koenker & Ng 2017) packages. Maps were prepared with GMT software (Wessel *et al.* 2013). Results from this study about attenuation and site terms are currently available upon request. For information about stream2segment package, see Zaccarelli *et al.* (2019). Site-characterization data referring to Swiss stations were drawn from *The Site Characterization Database for Seismic Stations in Switzerland, SED at ETH* (doi: 10.12686/sed-stationcharacterizationdb; <http://stations.seismo.ethz.ch>, last accessed January 7, 2022).

ACKNOWLEDGMENTS

This work has been developed in the framework of METIS (Methods and Tools Innovations for Seismic risk assessment (METIS, grant no. 945121). M-HY has been supported by Urban Engineering Seismology project (URBASIS, grant no. 813137). We thank all the network operators who openly share their data, making this study possible. D. Fäh and C. Cauzzi provided useful information about site characterization of stations belonging to the Swiss network. We thank two anonymous reviewers and the Editor for their constructive comments and suggestions.

REFERENCES

- Anderson, J.G. & Hough, S.E., 1984. A model for the shape of the Fourier amplitude spectrum of acceleration at high frequencies, *Bull. seism. Soc. Am.*, **74**, 1969–1993.
- Bates, D., Maechler, M., Bolker, B. & Walker, S., 2015. Fitting linear mixed-effects models using lme4, *J. Stat. Softw.*, **67**(1), 1–48.
- Bates, D. & Maechler, M., 2019. *Matrix: Sparse and Dense Matrix Classes and Methods*, R package version 1.2-18, <https://CRAN.R-project.org/package=Matrix>.
- Ben-Zion, Y., 2019. A critical data gap in earthquake physics, *Seismol. Res. Lett.*, **90**, doi:10.1785/0220190167.
- Bindi, D., Cotton, F., Kotha, S.R., Bosse, C., Stromeyer & Grünthal, D., 2017. Application-driven ground motion prediction equation for seismic hazard assessments non-cratonic moderate-seismicity areas, *J. Seismol.*, **21**, 1201–1218.
- Bindi, D., Zaccarelli, R., Strollo, A. & Di Giacomo, D., 2019. Harmonized local magnitude attenuation function for Europe using the European Integrated Data Archive (EIDA), *Geophys. J. Int.*, **218**, 519–533.
- Bindi, D. & Kotha, S.R., 2020. Spectral decomposition of the Engineering Strong Motion (ESM) flat file: Regional attenuation, source scaling and Arias stress drop, *Bull. Earthq. Eng.*, doi:10.1007/s10518-020-00796-1.
- Boore, D.M., 2009. Comparing stochastic point-source and finite-source ground-motion simulations: SMSIM and EXSIM, *Bull. seism. Soc. Am.*, **99**, 3202–3216.
- Brune, J.N., 1970. Tectonic stress and the spectra of shear waves from earthquakes, *J. geophys. Res.*, **75**, 4997–5009.
- Burger, R.W., Somerville, P.G., Barker, J.S., Herrmann, R.B. & Helmberger, D.V., 1987. The effect of crustal structure on strong ground motion attenuation relations in eastern North America, *Bull. seism. Soc. Am.*, **77**, 420–439.
- Castro, R.R., Anderson, J.G. & Singh, S.K., 1990. Site response, attenuation and source spectra of S waves along the Guerrero, Mexico, subduction zone, *Bull. seism. Soc. Am.*, **80**, 1481–1503.
- Chapman, M.C. & Godbee, R.W., 2012. Modeling geometrical spreading and the relative amplitudes of vertical and horizontal high-frequency ground motions in Eastern North America, *Bull. seism. Soc. Am.*, **102**, doi:10.1785/0120110081.
- Edwards, B., Rietbrock, A., Bommer, J.J. & Baptie, B., 2008. The acquisition of source, path, and site effects from microearthquake recordings using Q tomography: application to the United Kingdom, *Bull. seism. Soc. Am.*, **98**, 1915–1935.
- Edwards, B., Michel, C., Poggi, V. & Fäh, D., 2013. Determination of site amplification from regional seismicity: application to the Swiss National Seismic Networks, *Seismol. Res. Lett.*, **84**, 611–621.
- Fäh, D. *et al.*, 2009. *Determination of Site Information for Seismic Stations in Switzerland. Work Package 4: pegasos Refinement Project*, Swiss Seismological Service ETH, Zürich, Report SED/PRP/R/004/20090831.
- Florez, M.A., Caporale, M., Buabthong, P., Ross, Z.E., Asimaki, D. & Meier, M.-A., 2022. Synthesis of broadband earthquake ground motions using artificial intelligence, *Bull. seism. Soc. Am.*, doi:10.1785/0120210264.
- Frankel, A., McGarr, A., Bicknell, J., Mori, J., Seeber, L. & Cranswick, E., 1990. Attenuation of high-frequency shear waves in the crust: measurements from New York State, South Africa and Southern California, *J. geophys. Res.*, **95**, 17441–17457.
- Graves, R. W. *et al.*, 2011. CyberShake: a physics-based seismic hazard model for Southern California, *Pure appl. Geophys.*, **168**, 367–381.
- Graves, R.W. & Pitarka, A., 2015. Refinements to the Graves and Pitarka (2010) broadband ground-motion 465 simulation method, *Seismol. Res. Lett.*, **86**, 75–80.
- Hammond, J.O.S., England, R., Rawlinson, N., Curtis, A., Sigloch, K., Harmon, N. & Baptie, B., 2019. The future of passive seismic acquisitions, *Astron. Geophys.*, **60**, 37–42.
- Hu, Z., Roten, D., Olsen, K.B. & Day, S.M., 2021. Modeling of empirical transfer functions with 3D velocity structure, *Bull. Seism. Soc. Am.*, **111**, 2042–2056.
- Jiang, J., Erickson, B.A., Lambert, V.R., Ampuero, J.-P. *et al.*, 2022. Community-driven code comparisons for three-dimensional dynamic modeling of sequences of earthquakes and aseismic slip, *J. geophys. Res.*, **127**, doi:10.1029/2021JB023519.
- Jozinović, D., Lomax, A., Štajduhar, I. & Michelini, A., 2020. Rapid prediction of earthquake ground shaking intensity using raw waveform data and a convolutional neural network, *Geophys. J. Int.*, **222**(2), 1379–1389.
- Koenker, R. & Ng, P., 2017. *SparseM: Sparse Linear Algebra, R package version 1.77*, <https://CRAN.R-project.org/package=SparseM>.
- Kotha, S.R., Bindi, D. & Cotton, F., 2022. A regionally adaptable ground-motion model for fourier amplitude spectra of shallow crustal earthquakes in Europe, *Bull. Earthq. Eng.*, **20**, doi:10.1007/s10518-021-01255-1.
- Lanzano, G. *et al.*, 2019. The pan-European engineering strong motion (ESM) flatfile: Compilation criteria and data statistics, *Bull. Earthq. Eng.*, **17**, 561–582.
- Lee, R.L., Bradley, B.A., Stafford, P.J., Graves, R.W. & Rodriguez-Marek, A., 2020. Hybrid broadband ground motion simulation validation of small magnitude earthquakes in Canterbury, New Zealand, *Earthquake Spectra*, **36**, 673–699, doi:10.1177/8755293019891718.
- Lilienkamp, H., von Spetch, S., Weatherill, G., Caire, G. & Cotton, F., 2022. Ground-motion modeling as an image processing task: introducing a neural network based, fully data-driven, and nonergodic approach, *Bull. Seism. Soc. Am.*, doi:10.1785/0120220008.

- Mitchell, B., 2010. Prologue and invitation to participate in a forum on the frequency dependence of seismic Q, *Pure appl. Geophys.*, **167**, 1129, doi:10.1007/s0002-4-010-0180-3.
- Mori, F., Mendicelli, A., Falcone, G., Acunzo, G., Spacagna, R.L., Naso, G. & Moscatelli, M., 2022. Ground motion prediction maps using seismic-microzonation data and machine learning, *Nat. Hazards Earth Syst. Sci.*, **22**, 947–966.
- Muggeo, V.M.R., 2003. Estimating regression models with unknown break-points, *Stat. Med.*, **22**, 3055–3071.
- Münchmeyer, J., Bindi, D., Leser, U. & Tilmann, F., 2021. The transformer earthquake alerting model: a new versatile approach to earthquake early warning, *Geophys. J. Int.*, **225**, 646–656.
- Olsen, K.B., Begnaud, M., Phillips, S. & Jacobsen, B.H., 2018. Constraints of crustal Heterogeneity and Q(f) from regional (< 4 Hz) wave propagation for the 2009 North Korea nuclear test, *Bull. seism. Soc. Am.*, **108**, 1369–1383.
- Olsen, K.B., Begnaud, M., Phillips, S. & Jacobsen, B.H., 2020. Corrections to ‘Constraints of crustal Heterogeneity and Q(f) from regional (< 4 Hz) wave propagation for the 2009 North Korea nuclear test’, *Bull. seism. Soc. Am.*, **110**, 2598, doi:10.1785/0120200173.
- Oth, A., Bindi, D., Parolai, S. & Di Giacomo, D., 2011. 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. seism. Soc. Am.*, **101**, 667–687.
- Paolucci, R., Smerzini, C. & Vanini, M., 2021. BB-SPEEDset: a validated dataset of broadband near-source earthquake ground motions from 3D physics-based numerical simulations, *Bull. seism. Soc. Am.*, **111**, 2527–2545, doi:10.1785/0120210089.
- Pitarka, A., Akinci, A., De Gori, P. & Buttinelli, M., 2022. Deterministic 3D ground-motion simulations (0–5 Hz) and surface topography effects of the 30 October 2016 Mw 6.5 Norcia, Italy, earthquake, *Bull. seism. Soc. Am.*, **112**, 261–286.
- R Core Team, 2020. *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, <https://www.R-project.org/>.
- Razafindrakoto, H.N.T., Cotton, F., Bindi, D., Pilz, M., Graves, R.W. & Bora, S., 2021. Regional calibration of hybrid ground-motion simulations in moderate seismicity areas: application to the Upper Rhine Graben, *Bull. seism. Soc. Am.*, **111**, 1422–1444, doi:10.1785/0120200287.
- Roten, D., Cui, Y., Olsen, K.B., Day, S.M., Withers, K., Savran, W.H., Wang, P. & Mu, D., 2016. High-frequency nonlinear earthquake simulations on petascale heterogeneous supercomputers, *SC 16: Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*, pp. 957–968, IEEE, doi:10.1109/SC.2016.81.
- Roten, D. & Olsen, K.B., 2021. Estimation of site amplification from geotechnical array data using neural networks, *Bull. seism. Soc. Am.*, **111**, 1785–1794.
- Safarshahi, M. & Morozov, I.B., 2020. Robust empirical time-frequency relations for seismic spectral amplitudes, Part 2: model uncertainty and optimal parameterization, *Bull. seism. Soc. Am.*, **111**, 193–205.
- Strollo, A. *et al.*, 2021. EIDA: the European integrated data archive and service infrastructure within ORFEUS, *Seismol. Res. Lett.*, **92**(3), 1788–1795.
- Teng, G. & Baker, J., 2019. Evaluation of SCEC CyberShake Ground motions for engineering practice, *Earthq. Spectra*, **35**, 1311–1328.
- Wald, D.J., Worden, C.B., Thompson, E.M. & Hearne, M., 2019. Earthquakes, ShakeMap, *Encyclopedia of Solid Earth Geophysics. Encyclopedia of Earth Sciences Series*, ed. Gupta, H., Springer, Cham, doi:10.1007/978-3-030-10475-7_182-1.
- Weatherill, G., Kotha, S.R. & Cotton, F., 2020. Re-thinking site amplification in regional seismic risk assessment, *Earthq. Spectra*, **36**, 274–297.
- Wessel, P., Smith, W.H.F., Scharroo, R., Luis, J.F. & Wobbe, F., 2013. Generic mapping tools: Improved version released, *Eos Trans. Am. geophys. Un.*, **94**, 409–410.
- Wickham, H., 2016. *ggplot2: Elegant Graphics for Data Analysis*, Springer-Verlag New York, ISBN:978-3-319-24277-4.
- Wickham, H., François, R., Henry, L. & Müller, K., 2018. *dplyr: A Grammar of Data Manipulation. R package version 0.7.6*, <https://CRAN.R-project.org/package=dplyr>.
- Withers, K.B., Olsen, K.B., Day, S.M. & Shi, Z., 2018. Validation of deterministic broadband ground motion and variability from simulations of buried thrust earthquakes, *Bull. seism. Soc. Am.*, doi:10.1785/012018005.
- Withers, K., Olsen, K., Day, S. & Zheqiang, S., 2019. Ground motion and intraevent variability from 3D deterministic broadband (0–7.5 Hz) simulations along a nonplanar strike-slip fault, *Bull. seism. Soc. Am.*, **109**, 229–250.
- Zaccarelli, R., Bindi, D., Strollo, A., Quinteros, J. & Cotton, F., 2019. Stream2segment: an open source tool for downloading, processing and visualizing massive event-based seismic waveform datasets, *Seismol. Res. Lett.*, **90**, 2028–2018.
- Zaccarelli, R., Bindi, D. & Strollo, A., 2021. Anomaly Detection in Seismic Data-Metadata Using Simple Machine-Learning Models, *Seismol. Res. Lett.*, **92**, 2627–2639.
- Zollo, A., Orefice, A. & Convertito, V., 2014. Source parameter scaling and radiation efficiency of microearthquakes along the Irpinia fault zone in southern Apennines, Italy, *J Geophys Res Solid Earth*, **119**, 3256–3275.

SUPPORTING INFORMATION

Supplementary data are available at *GJI* online.

Figure S1. Residuals of the GIT decomposition at 3.2 Hz, obtained without considering the regionalization (each panel shows the residuals for a given region indicated with numbers from 1 to 6, see Fig. 1).

Figure S2. Residuals of the GIT decomposition at 3.2 Hz, obtained considering the regionalization shown in Fig. 1 (each panel shows the residuals for a given region indicated with numbers from 1 to 6).

Figure S3. Site amplification values (top) and interpolated map (bottom) at 0.6 Hz (left) and 6.25 Hz (right).

Table S1. List of DOIs of seismic network considered in this study. The DOI of networks not registered with the International Federation of Digital Seismograph Networks (FDSN) is indicated by NA.

Please note: Oxford University Press is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the paper.