UNIVERSITÀ DEGLI STUDI DI TRIESTE

XXXIII CICLO DEL DOTTORATO DI RICERCA IN

Scienze della terra, fluidodinamica e matematica. Interazioni e metodiche

Improving the quality of seismic monitoring by development of novel ground motion prediction techniques

Settore scientifico-disciplinare: GEO/10 Geofisica della Terra Solida

DOTTORANDO / A
LAURA CATALDI

COORDINATORE
PROF. STEFANO MASET

SUPERVISORE DI TESI
PROF. GIOVANNI COSTA
DR. VALERIO POGGI

ANNO ACCADEMICO 2019/2020

Improving the quality of seismic monitoring by development of novel ground motion prediction techniques

By

Laura Cataldi

University of Trieste, 2020

Acknowledgements

Abstract

L'attività di monitoraggio sismico è basata sull'uso di sismografi ricettivi per la registrazione del moto del suolo causato dai terremoti. Gli strumenti sviluppati nell'ambito del monitoraggio sismico permettono di studiare la sismicità a livello sia regionale che globale e trovano un uso strategico nel contesto della stima della pericolosità sismica. Uno sforzo continuo è necessario per migliorare gli strumenti al servizio del monitoraggio sismico e le conseguenti applicazioni, sia nel campo della stima della pericolosità che in contesti di protezione civile ed ingegneria sismica (come ad esempio nelle mappe di scuotimento del suolo).

Il primo fondamentale passo per migliorare gli strumenti e i modelli sviluppati nell'ambito del monitoraggio sismico è un corretto ed accurato trattamento dei dati. Un'attenta procedura di selezione ed elaborazione è stata seguita sulla base del tipo specifico di dati impiegati, differenziando tra dati strumentali, metadati e valori assegnati da esperti.

Una nuova definizione di intensità strumentale, la quale fornisce una previsione dell'intensità macrosismica sulla base del livello di scuotimento del suolo, viene proposta per il caso dell'Italia. Lo scopo è sostituire le equazioni lineari comunemente impiegate a tale scopo (*Ground Motion to Intensity Conversion Equations*, GMICE), che per loro natura non riescono a trattare in modo completamente corretto il dato dell'intensità macrosismica e la relativa incertezza. Un modello basato sulla tecnica dei classificatori di Bayes Gaussiani (*Gaussian Naïve Bayes*, GNB) è stato sviluppato e calibrato per un set di diversi parametri di moto del suolo. Tale modello fornisce stime di intensità su classi intere, in accordo con la definizione originaria di classi di intensità, ed una relativa probabilità associata. I risultati sono stati testati rispetto ad una formulazione delle più classiche GMICE calibrata sullo stesso gruppo di dati. L'intensità strumentale basata sulla definizione da GNB è risultata fornire prestazioni migliori in termini

di applicazione su dati indipendenti e di capacità di catturare l'incertezza associata al dato. Le stime di intensità basate sull'uso di parametri di massimo moto del suolo (in velocità e in accelerazione) sono risultate le più adatte all'applicazione diretta nei prodotti di monitoraggio sismico e convertite in una scala spezzata adatta all'uso.

Un algoritmo per la modellazione degli spettri di ampiezza di Fourier è stato sviluppato per effettuare un'inversione parametrica da cui ottenere stime di comportamento specifico al sito. Il relativo *software* è stato sviluppato in maniera flessibile per permettere un facile adattamento nella selezione dei modelli, degli stimatori di incertezza, dei coefficienti di peso, del sito di riferimento, della metodologia e del numero di passi usati nell'inversione. Sulla base di osservazioni sismotettoniche è stata individuata un'area nella regione del Nord-Est Italia, da usare come caso di studio per la stima dei parametri spettrali associati ad un gruppo di eventi e stazioni sismiche. Le curve di amplificazione al sito sono state ricostruite dall'analisi congiunta dei prodotti e dei residuali dell'inversione, ed è stato di conseguenza suggerito un gruppo di stazioni adatte ad essere usate come riferimento per il livello regionale di amplificazione su roccia. I valori di momento sismico ottenuti sono compresi tra 1.3×10^{13} e 2.6×10^{15} Nm, quelli di *stress drop* tra 1.5 e 18 MPa. L'attenuazione media è risultata inferiore a quella riportata in letteratura per la regione. Il valore massimo per i fattori di amplificazione al sito ottenuti è di 20, con frequenze di risonanza tipicamente tra 2 e 9 Hz. Il raffronto con la letteratura ha confermato l'affidabilità dei risultati ottenuti, in particolar modo per i termini relativi alle sorgenti sismiche e ai siti. La validazione rispetto a diverse scelte di parametrizzazione ha confermato la stabilità dell'algoritmo di inversione e fornito suggerimenti per migliorare la stima dei parametri legati all'attenuazione lungo il percorso sismico. Uno scenario di esempio è stato analizzato per illustrare il possibile uso degli strumenti sviluppati nel lavoro. Le stime ottenute per le curve di amplificazione sono risultate compatibili con altre osservazioni basate su dati indipendenti e adatte all'uso nei modelli di stima della pericolosità sismica per vincolare meglio la risposta al sito.

Abstract

Seismic monitoring employs sensitive seismographs to record the ground motion generated by earthquakes. It provides tools to study regional and global seismicity that are fundamental when applied inside seismic hazard assessment. A continuous effort is necessary to improve monitoring for both hazard assessment and direct applications in civil protection and engineering contexts (e.g., shakemaps).

The first step towards improving monitoring tools and models is to carefully select and handle data. A strict selection and processing procedure tailored to the specific kind of employed data is followed to ensure the quality of the ensuing results.

A novel definition of instrumental intensity for Italy is proposed, to provide a forecast of expected macroseismic intensity based on the ground motion shaking level. It is intended to substitute Ground Motion to Intensity Conversion Equations (GMICEs), which are linear relationships with known issues in treating intensity and its associated uncertainty. A model based on Gaussian Naïve Bayes (GNB) classifiers is developed and calibrated for a set of ground shaking parameters, providing integer-valued intensity forecasts with a known, class-specific associated probability. The results are tested against a more classical GMICE formulation calibrated on the same dataset. Instrumental intensity based on GNB definition is proved to possess better performance on unseen data and better capability of capturing data uncertainty with respect to GMICEs. Forecasts based on peak ground parameters (velocity and acceleration) are selected as most suitable for direct application in seismic monitoring products and converted in a ready-to-use piecewise scale.

An algorithm for Fourier amplitude spectra modelling is developed to perform parametric inversion and provide estimates of site-specific soil behaviour. A

flexible software is developed that supports the customization of employed models, uncertainty estimators, weighting coefficients, reference settings, inversion techniques and number of inversion steps. A case study area in the Northeast Italy region is chosen based on seismo-tectonics considerations, and spectral parameters are estimated for a set of selected events and stations. Site-specific amplification curves are built from a combined analysis of inversion products and residuals, and a set of candidate stations to be used as regional rock amplification reference is suggested. Inverted seismic moments range between 1.3×10^{13} and 2.6×10^{15} Nm; Brune stress drop values range between 1.5 and 18 MPa, comparably with reference values for the region. The average rate of attenuation is lower than the ones reported in earlier studies for the Northeast Italy region. Site-specific amplification factors as high as 20 are found, with typical resonance frequencies in the range $2 - 9$ Hz. Comparison with literature values reinforces the reliability of the results, especially in the case of source and site terms. Validation against different model parametrizations confirms the stability of the inversion algorithm and suggests additional steps to improve the estimate of path attenuation features. A case scenario is built to exemplify the possible use of the developed tools. Estimated amplification curves are found to be compatible with independent empirical observations and suitable for employment in hazard assessment models to better constrain site-specific response.

Table of contents

List of figures

List of tables

1 Introduction

1.1 Motivation and objective

Earthquakes release energy in the form of vibrations that propagate through the layers of the Earth and reach its surface, where we feel them as ground motion. Even if not all seismic events cause major damage, they overall represent one of the most destructive hazards in the world and caused nearly 750000 deaths globally only for the period from 1998 to 2017, according to World Health Organization (WHO). Seismic risk is the result of different contributions, namely:

$$risk = hazard \times vulnerability \times exposure .$$ (1-1)

Vulnerability and exposure provide a description of how many people and infrastructures will be affected by a given seismic event and of its expected impact based on the infrastructure characteristic, respectively. On the other hand, seismic hazard is determined by the frequency and intensity of ground motion in a given area, that is, the seismicity. While vulnerability and exposure can be reduced by applying targeted measures (such as improving the structural stability of buildings or managing the population; UN DHA 1994), seismic hazard is defined by the seismicity of the area and can only be assessed.

Even if hazard itself cannot be mitigated, its precise assessment is a key factor in reducing the total seismic risk, as vulnerability and exposure levels depend on the expected hazard. A comprehensive analysis of seismic hazard assessment strengths and weaknesses is provided in Wang (2010). The quality of hazard assessment is determined by the efficiency of models used to describe it, and consequently also by the availability and quality of seismic data on which models are calibrated. Seismic monitoring uses sensitive devices called seismographs to record the vibrations produced by the earthquakes and analyse them. Products of

seismic monitoring have a twofold usefulness, as they both provide information to help improve the understanding of seismicity and find direct application for civil protection and engineering purposes. For example, different Ground Motion Parameters (GMPs) extracted from seismic recordings can be used to define earthquake related hazard maps. In their most used form, such maps describe the maximum shaking amplitude that an earthquake could produce at a site and the associated shaking severity. Shaking maps produced in real or near-real time after an event are used by the civil protection unit to organize first aid operations in the epicentral area.

Macroseismic intensity is conventionally used to express how strongly an earthquake affects a specific place (e.g., Sieberg 1930; Grünthal 1998). It is a collective measurement, meaning that it depends on the observation of many factors, and is expressed as a scale of ordered categories, usually represented by roman (i.e., integer) numbers. Its definition makes it a less precise quantity with respect to other, directly measurable seismic features, but has the advantage of not needing any other instrument than the human eye to be recorded. For this reason, macroseismic surveys have been a useful tool to study historical events that occurred before the instrumental era (Guidoboni and Stucchi 1993). In fact, macroseismic measures are often the only available data for major earthquakes in an area, because of the long average time span between their occurrences. As major earthquakes are the ones controlling the maximum hazard level, macroseismic data has proven to be very useful in the context of hazard assessment. Additionally, macroseismic intensity is routinely used in other applications contributing to risk mitigation, such as vulnerability assessment of buildings (Lagomarsino and Giovinazzi 2006). Even if instrumental data have now taken the lead in the field of seismic monitoring, thanks to technological and infrastructural development, macroseismic intensity is still used in combination with it.

Immediately after an event, the expected intensity can be estimated based on the values of instrumental ground motion parameters. Ground Motion to Intensity Conversion Equations (GMICEs) are the most extensively used method for

defining such instrumental macroseismic intensity. GMICEs express the forecasted intensity as a linear function of the ground motion parameters, usually either Peak Ground Acceleration (PGA) or Peak Ground Velocity (PGV). The main issue with GMICEs is that such forecasts are expressed as decimal values, being the linear regression continuous, and thus need further elaboration either in the form of rounding or through statistic interpretation. Even so, it is a common habit to directly display the decimal results with the risk of interpreting them as an improvement in the precision of the measurement, while macroseismic classes are defined as natural, ordered categories and the distances between the categories are not known (Agresti 2013). To avoid this problem, some authors (e.g., Kuehn and Scherbaum 2010; Lancieri *et al.* 2015) proposed switching to a radically different method for instrumental intensity definition, the Gaussian Naïve Bayes (GNB) classifiers, which is capable of correctly handling ordinal data throughout the whole procedure. GNB also has the advantage of providing information on the probability associated to the output forecast.

When direct measurements are not available, Ground Motion Prediction Equations (GMPEs) are used to predict the maximum amplitude of acceleration, velocity and spectral periods at a site. GMPEs calculate a rock-referenced ground shaking forecast as a function of magnitude and distance and then scale it for a proxy site amplification factor, usually based on the average seismic shear-wave velocity from the surface to a depth of 30 meters (v_{S30}). GMPEs are efficiently used to estimate ground motions for use in seismic hazard assessment and are the most widely used representation of earthquake ground motion employed in engineering practice. Their precision is dependent on the intrinsic variability of the seismic features used to calibrate them, and extra carefulness must be paid in applying them inside their validity range only. For example, many GMPE formulations avoid describing effects of strong directivity, near-source conditions, or pulse shapes.

Many authors over the years tried to push the resolving power of GMPEs by making their functional form more and more articulate, for example by including different parameters to describe the faulting style and the soil classification.

However, even if increasing data could allow for more complex parametrization, it also increases the diversity of ground-motion observations and thus the associated aleatory variability of the obtained relationships (Kotha *et al.* 2020). Moreover, new predictor variables might not always be available and become an issue during application (Bindi *et al.* 2017). Aristizábal *et al.* (2017) investigated different approaches to integrate site effects together with GMPE into seismic hazard assessment, other than using site proxies inside GMPEs. They found that, least the site-specific knowledge is severely limited, site-specific approaches like the use of instrumental site response amplification functions present the major advantage of allowing a reduction of the within-event variability. Near-surface conditions can have a substantial effect on ground motions, so the availability of precise information on site effects can positively affect the quality of the forecast and, consequently, of the hazard estimate. This is especially true for sites located on soft and sedimentary soil, which are typically amplified relative to sites on bedrock. The extraction of site-specific amplification curves from residuals of Fourier spectral inversion, for instance, has been proven to provide amplification values consistent with predicted 1D behaviour at rock sites and more complex 2D or 3D behaviour at soil sites (Edwards *et al.* 2013; Michel *et al.* 2014).

Available instrumental seismic data is routinely combined with estimations made with GMPEs to generate ground shaking maps; an example of implementation is the USGS ShakeMap software (Wald *et al.* 1999c). In its default implementation, observed data is interpolated with ground shaking forecasts, which are calculated for a rock profile using GMPEs and converted into site-specific values using an average site amplification factor, based on v_{S30} (when available) or on geological maps and slope. The corresponding instrumental intensity is calculated using GMICEs. Uncertainties both on GMPEs and GMICEs merge in the actual uncertainty on intensity, and therefore their correct estimation determines the degree of control on the forecasts. Maps produced with ShakeMap or other analogous tools are used world-wide both by research institutes and governmental agencies (e.g., Michelini *et al.* 2008; Weginger *et al.* 2017; Wald *et al.* 2007).

This thesis proposes both an alternative definition of instrumental intensity, more data-compliant than GMICEs, and a data-based estimation of spectral amplification curves at selected sites through spectral inversion. It fits into the context of improving the quality of seismic monitoring products available in near real-time after an event. Application is exemplified in the case of Italy, a seismically active country where monitoring is carried out by numerous Institutions using both national and regional seismic networks.

1.2 State of the art

1.2.1 Instrumental macroseismic intensity

The correlation between peak ground motion measures (usually PGA or PGV) and intensity has been explored for decades. In earlier studies, a link was initially established between observed intensity at a location and the few available instrumental data in the form of so-called Intensity-to-Ground-Motion Conversion Equations (IGMCEs; Gutenberg and Richter 1942; Trifunac and Brady 1975; Murphy and O'Brien 1977). With the technological and infrastructural development and the consequent growing abundance of strong-motion seismic networks, the tendency was reversed and observed ground motion values associated to intensity through Ground Motion to Intensity Conversion Equations (Wald *et al.* 1999b; see Zanini *et al.* 2019 for a comprehensive review of GMICEs based on peak GMPs at global level). Most of these studies were limited by the non-reversibility of the developed models, so more recent studies explored the possibility of invertible relationships (Faenza and Michelini 2010; Worden *et al.* 2012).

Most of the aforementioned studies choose to define the instrumental intensity I as a linear function of the logarithm of a single ground motion parameter, in the form:

$$I = a + b \log_{10} GMP \,.$$

<div align="right">(1-2)</div>

Some authors prefer to use a bilinear function instead (Caprio *et al.* 2015; Moratalla *et al.* 2021). Only Gómez Capera *et al.* (2020) recently tried to propose a nonlinear functional form, predicting macroseismic intensity as a function of $\log_{10} GMP$ and vice versa by performing separate regressions. Different intensity scales are used depending on the region and on the availability of data. For instance, the Medvedev–Sponheuer–Karnik scale (MSK-64; Medvedev *et al.* 1964) was widely used in Eastern Europe and USSR, while the Mercalli-Cancani-Sieberg (MCS; Sieberg 1930) was more employed in Western Europe; both were later formally replaced by the European Macroseismic Scale (EMS-98; Grünthal 1998). The Modified Mercalli scale (MMI; Wood and Neumann 1931; Stover and Coffman 1993) is most used for macroseismic observations in the United States, while most Japanese measurements are reported using the Japan Meteorological Agency scale (JMA). Under certain circumstances, some scales overlap and can be used indiscriminately (Tiberi *et al.* 2018); a complete review of different scales and their characteristics can be found in Musson *et al.* (2010).

Many GMICE formulations exist in literature also for the specific Italian case (Figure 1-1). In fact, Italy counts one of the most comprehensive intensity databases thanks to its long tradition in the collection of macroseismic observation. The filing of macroseismic intensity reports dates back to 1783, when a task force of scientists (including geologists, architects and engineers) was sent to Calabria (Southern Italy) on behalf of the Bourbon kingdom to investigate the damages produced by the destructive seismic sequence that struck the region. Various GMICE models have been proposed in the past for Italy (Chiaruttini and Siro 1981; Margottini *et al.* 1992; Panza *et al.* 1997; Faccioli and Cauzzi 2006). More recently, Faenza and Michelini (2010, 2011), Zanini *et al.* (2019) and Masi *et al.* (2020) took advantage of both the Italian Macroseismic Database (DBMI) and the ITalian ACceleration Archive (ITACA) to propose new conversion equations. In particular, Faenza and Michelini (2010) developed a stable methodology that uses binning as a form of smoothing to mitigate the intrinsic variability of both instrumental and macroseismic data. They obtained linear

orthogonal relationships between MCS intensity and a set of ground motion parameters, including both PGV and PGA and selected spectral ordinates.
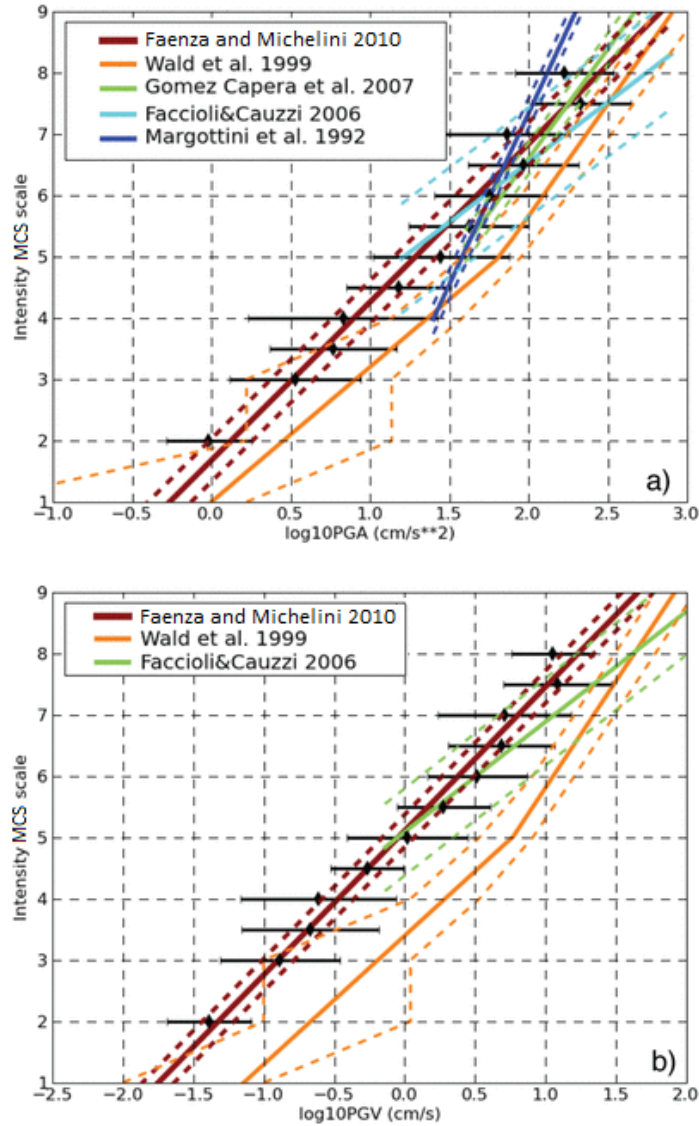
Figure 1-1: Comparison between the intensity versus GMP regressions obtained for Italy by different authors, for PGA (a) and PGV (b). GMICEs by Faenza and Michelini (2010), Wald *et al.* (1999b), Gómez Capera *et al.* (2007), Faccioli and Cauzzi (2006) and Margottini *et al.* (1992) are shown. Modified from Faenza and Michelini (2010).

All the aforementioned GMICEs use half-integer intensity values in their input database, as a consequence of their wide presence in macroseismic databases as a

form of uncertainty assignment. Moreover, linear regressions are defined as continuous functions and provide an output in the form of decimal values. This aspect has to be carefully taken into consideration when applying such results in the context of seismic monitoring, as displaying decimal values could lead to misinterpret the actual precision of the forecast. Decimal values could be interpreted to assign a quantified uncertainty between two intensity classes, but the most common choice is to simply round the output values to the nearest integer (e.g., Masi *et al.* 2020).

An important application of GMICE relations is inside the production of shakemaps around the globe. In the Italian case, GMICEs by Faenza and Michelini (2010, 2011) are considered the default choice. Because of the incomplete coverage of ground-motion records, shakemaps are often used as prediction tools, relying on GMPEs to estimate the ground shaking and then applying GMICEs to convert them into instrumental intensity. Shakemaps displaying instrumental intensity are widely used when communicating with the nontechnical community. Consequently, uncertainties both on GMPEs and GMICEs must be taken into account when estimating the uncertainty on intensity. In the case of GMICEs, the associated uncertainty is usually an average value describing the overall performance of the fit. No detailed information on class-specific uncertainty is supplied, even if the distribution of calibration data over different classes is usually highly variable because of the differing availability of observations for high and low intensity classes (lower being more populous).

Even if they are widely employed, GMICEs are not the only possible way to define intensity as a function of ground motion parameters. Naïve Bayes classifiers have been recently proposed as an alternative approach (Kuehn and Scherbaum 2010; Lancieri *et al.* 2015). GNB methodology uses supervised learning to calculate the conditional probability distribution of intensity class C given the value of the observed variables V according to Bayes' rule:

$$\Pr(C|V) = \frac{\Pr(C)\Pr(V|C)}{\Pr(V)}. \tag{1-3}$$

Being a Bayesian procedure, it requires some assumptions to be applied (Wasserman 2004):

- choosing a probability density $\Pr(C)$ expressing the belief on parameter (class) C before any data is seen (the so-called prior distribution);
- defining a statistical model $\Pr(V|C)$ expressing the belief about V data given the parameter C;
- updating the beliefs after observing the data and calculating the posterior distribution $\Pr(C|V)$.

The aim is to estimate the probability of class i given a set of N_V observed variable values, $\Pr(C = c_i | V = [v_1, \dots, v_{N_V}])$, as:

$$\Pr(C = c_i | V = [v_1, \dots, v_{N_V}]) = \frac{\Pr(C = c_i)\Pr([v_1, \dots, v_{N_V}] | C = c_i)}{\sum_j \Pr(C = c_j)\Pr([v_1, \dots, v_{N_V}] | C = c_j)}, \quad (1\text{-}4)$$

where c_j is a partition of the event space. In its naïve form, the theorem assumes that the V variables are conditionally independent on C, so the conditional probability $\Pr([v_1, \dots, v_{N_V}] | C = c_i)$ can be expressed as the likelihood product:

$$\Pr([v_1, \dots, v_{N_V}] | C = c_i) = \prod_{j=1}^{N_V} \Pr(v_j | C = c_i), \quad (1\text{-}5)$$

where $\Pr(v_j | C = c_i)$ is the conditional probability of observing the value assumed by the j-th variable V on the class c_i. This hypothesis of conditional independence actually implies that the value assumed by the j-th variable V does not depend on the value assumed by the other $N_V - 1$ variables. Even if this assumption may not be realistic from a physical point of view, Kuehn and Scherbaum (2010) showed that it works well since the intention of the Bayesian classifier is not to provide a physical model to explain the data-generating process, but rather to serve as a simple statistical predictor of the class C given the variable V, based on the empirical probability curves. This way, GNB can be easily applied to investigate also a multi-parametric definition of the instrumental intensity.

The main advantage of GNB is that ordinal data is easily handled, with guaranteed respect of the defined classes. Moreover, its result is not a single value but a full discrete probability over the available intensity classes. The combination of these aspects makes it easier to extract a definition-compliant, integer forecast value from the results, the most natural choice being the class with highest probability of occurrence. Additionally, the associated probability is automatically available and can be used to easily assess the uncertainty in the forecast.

1.2.2 Site-specific amplification and Fourier spectral inversion

Depending on geological conditions, site effects might cause a considerable amplification of the ground shaking and consequently increase the damage suffered during earthquakes. One of the most known examples is the case of the 1985 Michoacán, Mexico earthquake, when seismic waves originated from an event over 350 kilometres away from Mexico City underwent resonation only in the urban areas built on lacustrine sediments, causing localized major damages. Site effects are the variations produced on seismic wave propagation by local geological conditions in superficial soil layers (e.g., Aki 1993) and their measure is called the site transfer function. In most cases, site effects are linear, meaning that the transfer function is independent of the seismic input, and the soil response is self-similar regardless of the earthquake characteristics. Different, nonlinear effects can be observed in the case of strong ground motion.

Several factors influence the presence of site-specific amplification effects. The soil properties themselves play an important role, as softer rocks slow down the seismic wave velocity with a consequent increase in amplitude and build-up of seismic energy. Moreover, the geometry of the soil deposits can cause additional effects such as refraction, diffraction and focalization, and may result in trapped waves causing amplification due to constructive interference. Trapped waves also increase the duration of the shaking (and consequently the potential damage) and are capable of triggering resonance with vibrational modes of buildings. If the soil geometry is simple (1D case, single layer), vertical reverberations enhance the amplification at well-defined frequencies, whereas if the geometry is more

complex (2D and 3D cases) the additional lateral reverberations may cause broad band effects (e.g., Faccioli *et al*. 2002).

The study of local site effects plays an important role in the assessment of strong ground motions and seismic hazard. Many different techniques can be used to estimate the site transfer function. In the presence of simple (1D) soil structures, the quarter-wavelength approximation can provide estimates of frequency-dependent average seismic parameters (velocity and density) and the corresponding amplification factors (Poggi *et al*. 2011). Other techniques perform an empirical estimation by either using earthquake or noise measurements. In urban areas, the identification of the fundamental resonance frequency and the corresponding amplification is often carried out through site response analysis. One of the most common techniques used to estimate the site response is the use of Standard Spectral Ratios (SSR; Borcherdt and Gibbs 1976) of earthquake records, which consists in dividing the frequency spectrum obtained from the observed waveform at a site by that observed at a nearby reference station, preferably located on bedrock. When no reference station is available, the H/V approach (also called receiver function method) allows to estimate the site characteristics by computing the spectral ratio of the horizontal and vertical components on the S-wave portion of different traces recorded at the station (Lermo and Chavez-Garcia 1993; Field and Jacob 1995). The receiver function method is effective in identifying the fundamental frequency of a site but does not correctly retrieve the amplification factor at higher frequencies (Lermo and Chavez-Garcia 1993; Bonilla *et al.* 1997). Supplementary analyses employ seismic noise measurements (e.g., Nakamura 1989) and proved to be especially effective in the identification of the fundamental frequency for alluvial sites (e.g., Duval *et al*. 2001), but has limitations in estimating the amplification magnitude (Lermo and Chavez-Garcia 1993). Spectral studies are often performed on weak motion data, as they are more easily available, especially for areas of moderate seismicity. However, weak motion data might fail in providing information on the possible nonlinear site effects such as inelastic soil behaviour.

A different approach is represented by the reconstruction of site transfer functions through Fourier spectral inversion of a set of observations at the investigated site. This method is based on the assumption that the Fourier Amplitude Spectrum (FAS) $FAS_{ij}(f,r)$ observed at a recording station j originating from an earthquake i can be described as:

$$FAS_{ij}(f,r) = (2\pi f)^n \, \Omega_i(f) \, D_{ij}(f) \, S_{ij}(r) \, T_j(f) \, I_j(f) \,, \qquad (1\text{-}6)$$

where f is the frequency, r is the hypocentral distance, n is a factor accounting for the used component (acceleration, velocity or displacement), $\Omega_i(f)$ is the source model describing the amplitude spectrum at source, $D_{ij}(f)$ is the attenuation along the ray path, $S_{ij}(r)$ is the apparent geometrical spreading describing amplitude decay with distance, $T_j(f)$ is the site transfer function and $I_j(f)$ is the instrument response function. The problem of separating source, path, and site effects in terms of physical phenomena is a nontrivial and nonlinear problem (Scherbaum 1990; Poggi *et al*. 2011). Depending on the used parametrization, strong trade-off effects may influence the inversion results. Both parametric (e.g., Salazar *et al*. 2007; Drouet *et al*. 2008) and non-parametric (e.g., Parolai *et al*. 2000, 2004; Oth *et al*. 2008, 2011) inversion techniques have been developed to solve the inverse problem represented by Equation (1-6). All algorithms require the application of some constraints, both on the attenuation operator (e.g., Castro *et al.* 1990) and either on the source or site response (Andrews 1986). The non-parametric approach has the advantage of providing more insights into the potential complexity of attenuation but needs a very good distance coverage in order to employ its full potential.

Parametric Fourier amplitude inversion is a flexible tool that can be applied in a variety of cases. Many different implementations exist in literature, differing in the choice of parameters used in the modelling and in the details of the model functional form, in the area of application and in the algorithm used to perform the inversion. Edwards *et al.* (2008) developed a procedure calibrated on United Kingdom data based on mixed multi-step inversion, Drouet *et al*. (2008) used a single-step inversion on data collected by the French accelerometric network.

Bora *et al.* (2017) fine-tuned a parametric inversion procedure analysing acceleration traces recorded in Europe, the Mediterranean, and the Middle East and combined it with Random Vibration Theory to additionally estimate predicted response spectra. Throughout all works based on parametric inversion, two main strategies are defined to retrieve the site transfer function, which is usually parametrized as:

$$T_j(f) = A_j a_j(f) e^{-\pi f \kappa_j} \, , \qquad (1\text{-}7)$$

where A_j is the average site amplification relative to the chosen reference (usually a rock profile), $a_j(f)$ is the frequency-dependent site amplification function, and κ_j is a constant, site-related attenuation operator (Anderson and Hough 1984). Some authors (e.g., Drouet *et al.* 2008) preferred to directly include the site transfer function as a whole as a parameter in the inversion. Another common strategy consists in a combined approach, using direct inversion of earthquake recordings to estimate parameters A_j and κ_j together with residual analysis to reconstruct the frequency-dependent amplification functions in a non-parametric way. The latter approach has been shown to provide amplification values consistent with predicted 1D behaviour at rock sites and more complex 2D or 3D behaviour at soil sites (Edwards *et al*. 2013; Michel *et al*. 2014) and with results obtained through SSR (Drouet, Triantafyllidis *et al*. 2008).

Many studies investigated parametric inversion with specific calibration for the Italian case, both in non-parametric (e.g., Castro *et al*. 1997; Pacor *et al*. 2016) and parametric (e.g., Zollo *et al.* 2014) forms. Multiple analyses are available about the North-Eastern region of Italy in particular (e.g., Castro *et al.* 1996; Castro *et al*. 1997; Malagnini *et al*. 2002; Franceschina *et al.* 2006), but are mainly focused on other spectral parameters such as attenuation and the quality factor, with little coverage on site-specific transfer functions.

2 Data processing

The quality of data used to calibrate seismic monitoring tools has a high impact on the efficiency of their empirical results. This is particularly evident in the case of the GNB methodology, in which the total population of the database is determined by the limited availability of macroseismic measures that can be safely associated to instrumental data. Great effort was put into data collection and processing in order to ensure the highest possible accuracy of products derived from it.

Pre-selection of data is the first step towards correct building of databases. In the effort of ensuring a reliable definition of instrumental intensity, only data clearly pertaining to main shocks was considered in order to exclude the presence of observations describing cumulative damage. In fact, Grimaz and Malisan (2017) highlighted the potential bias due to the cumulative damage in the correlation between intensity and ground motion parameters, leading to a significant difference in intensity assessment when the same damage is associated to single or multiple seismic shocks. Only macroseismic observations obtained by expert assessment were considered and retrieved from specialized macroseismic databases and reports. In principle, it would be possible to also include crowdsourced intensity data gathered through internet "Did you feel it?" questionnaires (e.g., Wald *et al.* 1999a). Nevertheless, Lesueur *et al.* (2013) proved that the inclusion of crowdsourced data is not a trivial process, especially when it comes to associating it to the available instrumental data. To guarantee the homogeneity of the database, crowdsourced data was excluded from the analysis. Equal care was applied to the pre-selection of instrumental seismic recordings. As explained in more detail in §5.2, precise information on the instrumental response at each station is fundamental to avoid introducing unmodelled contributions that would affect the results of spectral inversion. All recordings obtained by seismic stations without certainty on the correctness of instrumental information, or with no associated information at all, were excluded from the preliminary selection.

Instrumental data obtained from the pre-selection were organized into BRTT, Inc. Antelope® databases. This choice was also motivated by the fact that most data was already available in such format. In fact, Antelope® is a data acquisition and processing software routinely used by many of the largest seismic networks in the world, as well as the Italian national strong motion network and many local networks in the North-Eastern Italy region (online Kinemetrics portfolio, https://kinemetrics.com/portfolio/; cf. §5.1 for a complete description). Even if the processing scheme was applied based on this specific arrangement, waveforms were collected in the standard *miniseed* format and the whole process can be adapted to different database management systems. The base-level processing procedure was modified from Gallo *et al*. (2014) and applied on both the national-level database used for GNB classification and the regional-level database used in spectral inversion. Its main features include an automatic calculation of the usable frequency range between 0.1 and 50 Hz based on signal to noise ratio, consequent Butterworth filtering on the selected range, and a trend removal used to compute peak ground motion values, including velocity and displacement. The possibility of calculating and saving Fourier amplitude spectra was added and tested for stability on the correct use of spectral amplitude scaling.

A supplementary selection was operated on the database after the base-level processing. The accuracy of extracted GMP values was tested and the case of triggered recordings was dealt with. For long time, it was common practice in seismic monitoring to save only the portion of waveforms immediately preceding and following an event. Even if modern instruments now greatly increased the availability of continuous recording, most waveforms recorded until the Nineties come in a triggered (i.e., cut) form. This usually allows a good estimate of the associated peak parameters, as they prevalently fall inside the recorded time window, whereas the missing of part of the signal and consequently the lack in given frequency bands might influence the extraction of spectral parameters. To avoid possible underestimation, spectral parameters associated to triggered waveforms were marked for exclusion. Similarly, the duration of the noise window used to calculate signal-to-noise ratios determines the actual frequency band on which noise can be safely estimated. As a consequence, the investigated

frequency band was accordingly reduced to avoid possible underestimation of the noise Fourier spectra. A more detailed description of this selection process is reported in §5.2.

The obtained databases served as input for the specific data processing described in §3.1 and §5.2. Particular effort was put into the correct handling of macroseismic data contained in the dataset used for GNB classification. This is due to the awareness that even expert-assessed intensity measurements frequently include half-integer or "in between" values to express the uncertainty in class attribution (e.g., Locati *et al*. 2019). If used as-are, such values constitute a distortion of the original definition given by macroseismic scales, which usually distinguish ordinal classes based on a combination of earthquake effects suffered by people, buildings and environmental features. An exemplification of actual class definition according to the Mercalli-Cancani-Sieberg (MCS) macroseismic intensity scale is reported in Table 2-1. In order to be compliant to class definitions and to correctly include the uncertainty associated to intensity measures, an ulterior specific data processing is needed, as described §3.1.

Table 2-1: Excerpts from the MCS scale elaborated by A. Sieberg in 1930. Modified from Ferrari and Guidoboni (2000).

Degree (class)	Description
I – Imperceptible	Noticed only by seismographs.
II – Very light	Felt only by rare, nervous subjects who are perfectly quiet or very sensitive, and more exactly almost only on the upper floor of buildings.
III – Light	Even in densely inhabited areas the earthquake is felt as a shock by only a small part of the inhabitants who are inside their houses, like the quick passing of an automobile. By some it is recognized as an earthquake only after a reciprocal exchange of ideas.
IV – Moderate	Not many of the people who are outside of the buildings feel the earthquake. Inside the houses it is certainly identified by numerous, but not all, people, in consequence of the trembling or slight oscillatory movement of the furniture […].
V – Relatively strong	[…] The seism is felt by numerous people in the street or anyway outside. In flats the earthquake is noticed because of the shaking of the entire building. Plants as well as the thin branches of bushes and trees move visibly […]. Free hanging objects start to oscillate [...].
VI – Strong	The earthquake is felt by all with fear: therefore many run outside […]. Rather stable furnishings and even isolated pieces of furniture are moved or fall […]. In single, solidly built houses there is small damage […]. Heavier, but still harmless damage on badly constructed buildings. […]
VII – Very strong	Remarkable damage is caused to a higher number of furnishings of the flats […]. Watercourses, ponds and lakes begin to wave and become turbid because of the moving slime […]. There is moderate damage to numerous solidly built buildings […]. Isolated collapse of badly built or badly preserved houses.
VIII – Destructive	Entire tree trunks wave vividly or are even uprooted. Also the heavier furniture is partially moved and turned over […]. About one fourth of the houses reports heavy destructions; some collapse, many become uninhabitable […]. Sand and slime come out of wet grounds.
IX – Ruinous	About half of the stone houses is heavily destroyed; a certain number of them collapses; the largest part becomes uninhabitable. Houses with wooden frames are pulled up from their own foundations […].
X – Annihilating	Heavy destruction of about three fourths of the buildings; the largest part collapses. Even solid wooden constructions and bridges undergo heavy damage. Some are also destroyed […]. In the soft, and especially wet ground, fissures […] are formed […]. Big rocks break off the river banks and the steep coasts […].
XI – Catastrophe	Collapse of all stone buildings […]. Even the bigger and safer bridges collapse due to the breaking of the stone pillars or to the yielding of the iron ones […]. The ground undergoes various, considerable changes, which are determined by the nature of the soil […].
XII – Big catastrophe	No man-made work resists. The transformation of the ground takes on enormous dimensions. Accordingly water flows, under and above the ground, undergo the most various changes: waterfalls are formed, lakes stagnate, rivers divert.

3 Instrumental macroseismic intensity

Linear regressions are the most used form in implementing a definition of instrumental macroseismic intensity based on observed ground motion observations. Their main downside is that they treat intensity as a continuous numerical value, whereas intensity classes are by definition ordinal objects. For this reason, predicted outcomes are not directly meaningful and need further manipulation to be interpreted. Most commonly, they are either rounded to the nearest integer value, or read as reflecting an uncertainty between two intensity classes. As one goal of this work is to improve the instrumental intensity definition by expressing forecasts in the form of ordinal values, the Gaussian Naive Bayes classification methodology was selected as preferred, and GMICE-like linear regressions were only meant as a comparison. GNB allows to correctly handle ordinal data throughout the whole inversion process by estimating a discrete conditional probability distribution that links the (ordinal) intensity I to any GMP class (Pedregosa *et al*. 2011). This methodology also has the additional benefit of directly providing a probability associated to the intensity forecast.

3.1 Data

The accelerometric input database is composed of two parts, according to the availability and ownership of the data.

The first part of the database refers to 72 events from 1972 to 2004 (Table A-1 in the Appendix), with $M_L \geq 3.4$, and is composed of 193 GMP-intensity pairs. These are part of the set of 87 events (266 GMP-intensity pairs) used by Faenza and Michelini (2010) to derive the GMICEs used to date for the Italian territory. The original dataset includes older recordings, not always of high-quality data,

originally being analog traces with no information on the starting time. A thorough analysis was conducted in order to discard all the recordings with evident artificial signals arising from analog-to-digital conversion, or with no clearly identifiable peaks in acceleration, velocity or displacement. Recordings with multiple sequential events were also discarded. Data for the remaining 72 events are taken from the ITACA 2.0 database (ITalian ACcelerometric Archive, version 2.0, Luzi *et al*. 2008; Pacor *et al*. 2011), and now all belong to the RAN network (Rete Accelerometrica Nazionale, Gorini *et al*. 2010; Costa *et al*. 2014), managed by the Italian National Civil Protection (DPC).

The second part of the database comes from a selection of 18 events (82 GMP-intensity couples) from 2002 to 2016, again with $M_L \geq 3.4$. This dataset is a re-elaboration of the one used by Tiberi *et al*. (2018), and is composed of high-quality accelerometric data collected by the CE3RN (Central Eastern European Earthquake and Research Network, Costa *et al*. 2010; Bragato *et al*. 2014) and RAN stations.

Macroseismic intensity data-points associated to the event dataset are mainly taken from the 2015 version of the Italian Macroseismic Database (DBMI15; Locati *et al*. 2016), with supporting information from expert-assessed macroseismic reports (Azzaro *et al*. 2016; Galli *et al*. 2016; Tertulliani and Azzaro 2016). All intensity values are expressed in the Mercalli-Cancani-Sieberg macroseismic scale. Only main shocks were analysed, in order to exclude cumulative damage which would require an *ad hoc* separate modelling. To guarantee the homogeneity of the database, only macroseismic intensity measures issued from expert surveys were used. In fact, the inclusion of crowdsourced intensity data is not a trivial process (Lesueur *et al*. 2013) and would go beyond the scope of this work.

A set of ground motion parameters were considered as candidates for instrumental intensity definition, namely peak ground acceleration (PGA), velocity (PGV) and displacement (PGD), Arias intensity (I_A), Housner intensity (I_H), and spectral acceleration at 0.3 s (PSA03), 1.0 s (PSA10) and 3.0 s (PSA30). The GMPs for all events were extracted after homogeneously processing the waveforms according

to the procedure described in Chapter 2. In the case of badly triggered records, for which the pre-event noise window was not sufficient to correctly evaluate the signal to noise ratio, the Butterworth filter corner frequencies were taken from the ITACA database directly. GMPs were taken from the maximum value between the horizontal components. A minimum distance criterion was used to associate the parameters with the observed intensity values, with a maximum distance limit of 3 km. The complete database counts 90 events (Figure 3-1) in the time-window from 1972 to 2016, corresponding to 275 associated GMP-intensity pairs. Intensity values range between $I_{MCS} = $ II and $I_{MCS} = $ X. Epicentral distances range between 1.6 and 150.7 km (Figure 3-2) and are well distributed, especially for the central values of intensities (IV − VI).
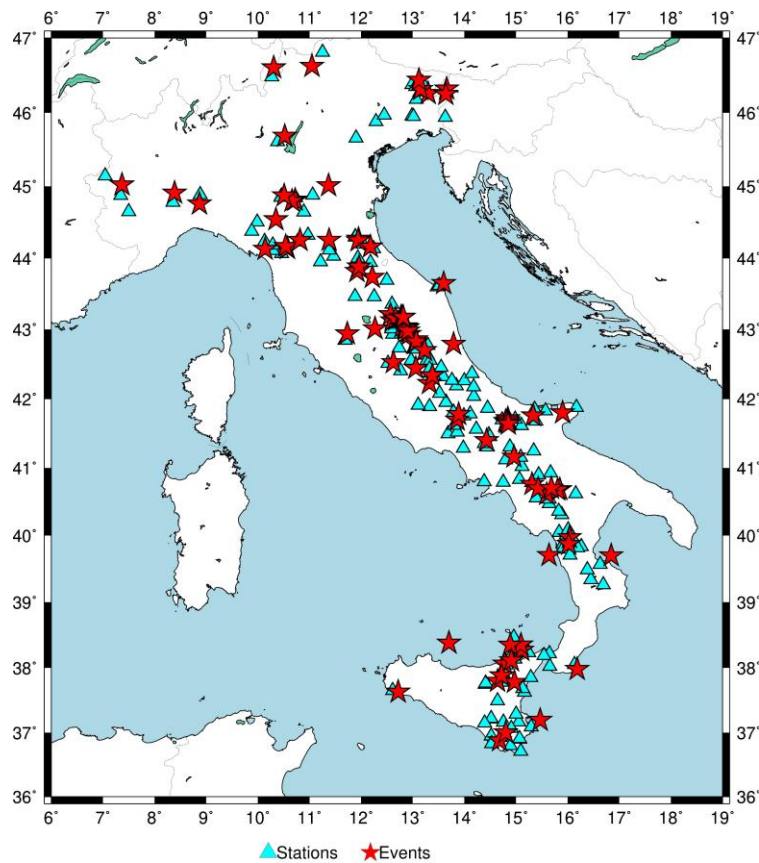
Figure 3-1: Spatial representation of the dataset used for instrumental intensity estimation. Red stars mark the epicentral locations of the analysed events; cyan triangles mark the station sites for which the GMPs are estimated with an associated observed intensity value. From Cataldi *et al.* (2021).
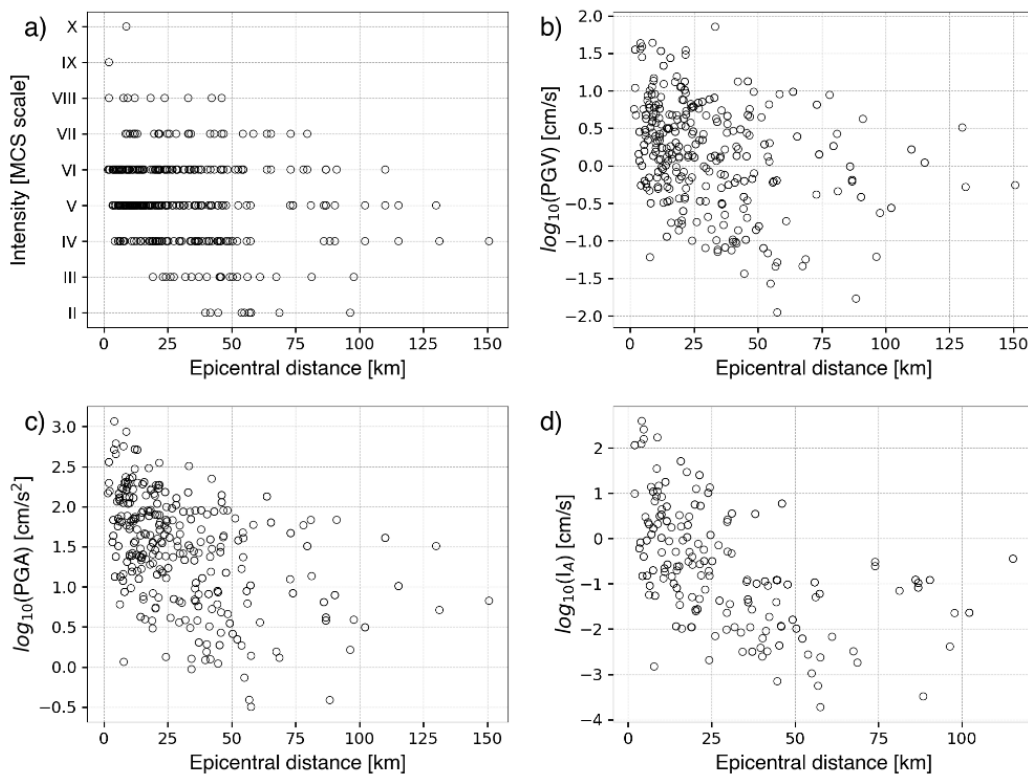
Figure 3-2: Distribution of intensity points (a) and parameters PGV (b), PGA (c) and I$_A$ (d) as a function of epicentral distance. Plots for remaining parameters are reported in the Appendix (Figure A-1).

A specific data processing was performed to correctly handle macroseismic intensity values. By definition, macroseismic intensity is an ordinal quantity, which means that it has natural, ordered categories and the distances between the categories are not known (Agresti 2013; Kuehn and Scherbaum 2010). Regardless of the chosen scale, intensity classes are always defined as a "collective" measurement, coming from the observation of many factors which are not linearly dependent on any single, directly measurable value. This implies that classes are not proportional to one another. While the proportionality between the energy released by a $M_W = 4.0$ and a $M_W = 5.0$ earthquake can be exactly defined, that there is no assurance that the effects observed for a degree V are proportional to those of a degree IV. Moreover, intensity measures have a high error content by definition, pushing strongly towards the impossibility of interpreting decimal values as an improvement in the actual intensity estimate. Even so, it is common practice to treat intensity as a continuous value, so this kind of data is widely

present in dedicated macroseismic surveys and is also found in the Italian Macroseismic Database. In particular, the DBMI custom is to use half integer classes as a way to assign uncertain MCS intensity values. Following Kuehn and Scherbaum (2010), in order to be consistent with the class definitions given by the MCS scale, half integer values were reassigned to the nearby integer classes, with the use of some weights, so that integer classes only would appear in the actual dataset used in calculations. More specifically, all data originally corresponding to half integer classes were assigned both to the above and below integer classes, with a weight $w = 0.5$, whereas all data originally corresponding to integer classes were assigned a weight $w = 1$. This procedure also allowed to include the uncertainty information coming from DBMI measures.

The resulting data distribution over different intensity classes is reported in Figure 3-3; the corresponding decomposed distributions for each parameter class can be found in the Appendix (Figure A-2). The weighted dataset consists in 376 points for the peak parameters (PGA, PGV, PGD), 174 of which originally belonging to an integer class and thus kept unchanged. It should be stressed that, in the case of partial or cut recordings, the extracted parameters were limited to peak amplitudes (PGA, PGV, PGD) and all integral quantities were discarded as not reliable, as a consequence to the risk of underestimating them due to missing part of the record. It is the case of Arias intensity, PSA03 and PSA10, with 220 usable points (100 with $w = 1$), Housner intensity, with 200 usable points (94 with $w = 1$), and PSA30, with 195 usable points (91 with $w = 1$). The difference in the number of available points for each parameter also comes from the fact that they are calculated on different frequencies (or frequency ranges), so each parameter was kept or discarded based on the specific corner frequencies applied to filter the waveform originating it. For example, in some cases the used high-pass filter was so high that it would filter out the frequency values used in the Housner and PSA parameter calculation.

Figure 3-3: Cumulative distribution of the weighted dataset points binned into classes at integer intensity intervals.

It is interesting to analyse the distribution of the GMP standardized values with respect to Normal curves with zero mean and a unit standard deviation (Figure 3-4). From that, it clearly emerges how after the application of the logarithm in base 10 the data follow the normal Gaussian curves, so the logarithmic form of the dataset was used for the calculations.
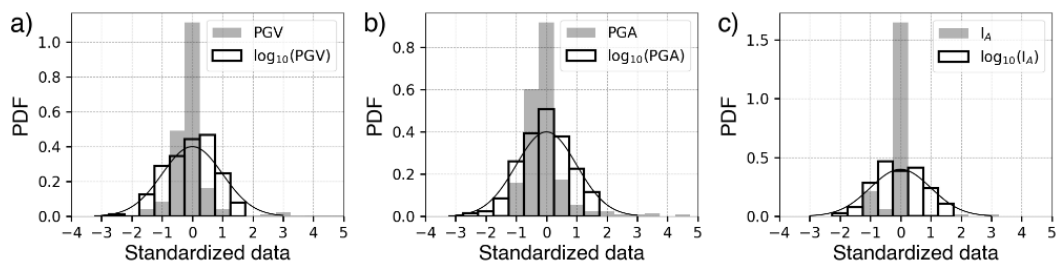
Figure 3-4: GMP value distributions for PGV (a), PGA (b), and I_A (c). In each graph the standardized data with zero mean and unit standard deviation is represented, in grey for the original data and as black boxes for the logarithm in base 10 of those. As a reference, the Gaussian normal distribution is depicted (solid black line). From Cataldi *et al.* (2021).

23

3.2 Method and results

3.2.1 Gaussian Naïve Bayes classifiers

GNB is part of a set of supervised learning algorithms based on applying Bayes' theorem in the Naïve form, that is, with the assumption of conditional independence between every pair of features given the value of the class variable (Zhang 2004). In this application, only one feature at a time was considered, so that the used formulation actually coincides with the full Bayes' theorem. Even so, the use of a flexible tool was preferred as it could in principle be easily reconverted to treat multi-parametric inversion, if needed. An extensive description of the procedure itself and of the underlying statistics can be found in Lancieri *et al.* (2015) and references therein; a synthetic overview of its application is provided here.

For any variable X (taken among the eight selected GMP classes), and the categorical variable I which is dependent on X, a Naïve Bayes classifier predicts the conditional probability distribution of I given $\log_{10} X$ according to Bayes' rule:

$$\Pr(I|\log_{10} X) = \frac{\Pr(\log_{10} X \,|I) \Pr(I)}{\Pr(\log_{10} X)}, \tag{3-1}$$

where $\Pr(\log_{10} X \,|I)$ is the conditional probability of observing $\log_{10} X$ on class I, and $\Pr(I)$ and $\Pr(\log_{10} X)$ are the *a priori* probabilities for I and $\log_{10} X$, respectively.

Equation (3-1) can be used to predict the probability of having intensity class k when the variable X takes the value x_i, as:

$$\Pr(I = k|\log X = \log_{10} x_i)$$
$$= \frac{\Pr(\log_{10} X = \log_{10} x_i \,|I = k) \Pr(I = k)}{\sum_j \Pr(\log_{10} X = \log_{10} x_i \,|I = j) \Pr(I = j)}, \tag{3-2}$$

where summation over j covers the whole event space, that is, all possible intensity classes. The prior distribution for the discrete variable I on each class k, $\Pr(I = k)$, was estimated in a frequentist way as the relative frequency of observations on the dataset:

$$\Pr(I = k) = \frac{N_k}{N_{tot}}, \tag{3-3}$$

in which N_k is the amount of data in class $I = k$ and N_{tot} is the total amount of data. Estimation of the conditional probability $\Pr(\log_{10} X \,|\, I)$ was obtained from statistical analysis of the dataset, based on the mean and standard deviation values associated to the parameter distributions (cf. Figure 3-4). Mean values μ_k^* were obtained as the weighted arithmetic mean of the logarithm in base 10 of the parameter ($\log_{10} X$), for each intensity class k between II and X:

$$\mu_k^* = \frac{\sum_{j=1}^{N_k} w_{jk} \log_{10} x_{jk}}{\sum_{j=1}^{N_k} w_{jk}}, \tag{3-4}$$

where w_{jk} is the weight assigned to the j-th point among the N_k data points with $I_k = k$. As for the associated errors, there is an evident lack of data for some classes with respect to others (e.g., classes IX and X versus class V; cf. Figure 3-3) which prevents from providing a robust estimate of the regular standard deviation for those classes. The classical definition of the standard deviation values would produce very low or even null values and would not reflect the actual distribution of the underlying data. Following the approach proposed by Kuehn and Scherbaum (2010), a standard deviation σ_{CSD} (common to all intensity classes for each GMP) was estimated as the square root of the pooled variance:

$$\sigma_{CSD} = \sqrt{\frac{\sum_k \sum_{j=1}^{N_k} (\log_{10} x_{jk} - \mu_k^*)^2}{\left(\sum_{j=1}^{k} N_j\right) - 9}} = \sqrt{\frac{\sum_k \sum_{j=1}^{N_k} (\log_{10} x_{jk} - \mu_k^*)^2}{N_{tot} - 9}}, \tag{3-5}$$

where the denominator is corrected for the number of different classes in which data were binned (nine). Values of μ_k^* and σ_{CSD} for each parameter are reported in the Appendix (Table A-2). The corresponding normal distribution for each

intensity class k, with a mean value of μ_k^* and common standard deviation σ_{CSD}, was thus used as an estimate of the conditional probability $\Pr(\log_{10} X \,|I)$.

The *python* algorithm *pomegranate* (Schreiber 2018) was used to perform GNB classification on the whole dataset. Figure 3-5 shows an example of the corresponding conditional probability distribution $\Pr(\log_{10} X = \log_{10} x_i \,|I = k)$ for each intensity class, for the PGA case. GNB fits the probabilities on the whole dataset to obtain a discrete conditional probability distribution on all intensity classes for each input GMP value; the class with the highest associated probability value is then chosen as the best estimate of I.
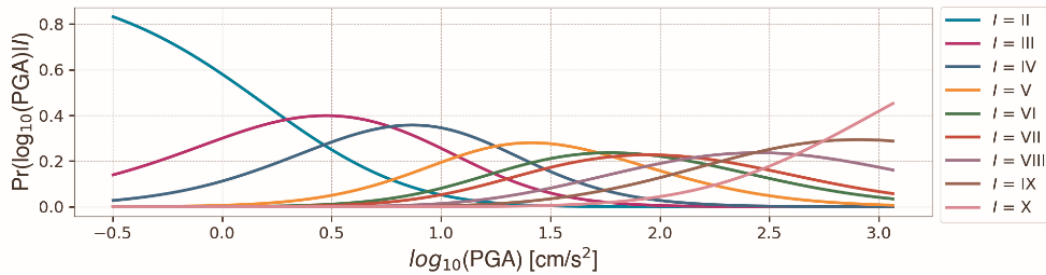
Figure 3-5: Example of the a priori probability distribution $\Pr(\log_{10} X = \log_{10} x_i \,|I = k)$ used by the GNB Classification for the PGA parameter, for each intensity class. From Cataldi *et al.* (2021).

The resulting intensity predictions are represented in Figure 3-6, colour-coded from lower (white) to higher (black) associated probabilities. They are obtained by applying the model to a linear space covering all values of the input parameters; for each value on the x-axis, the corresponding colour-coded probability values along the vertical (intensity) axis sum up to one.
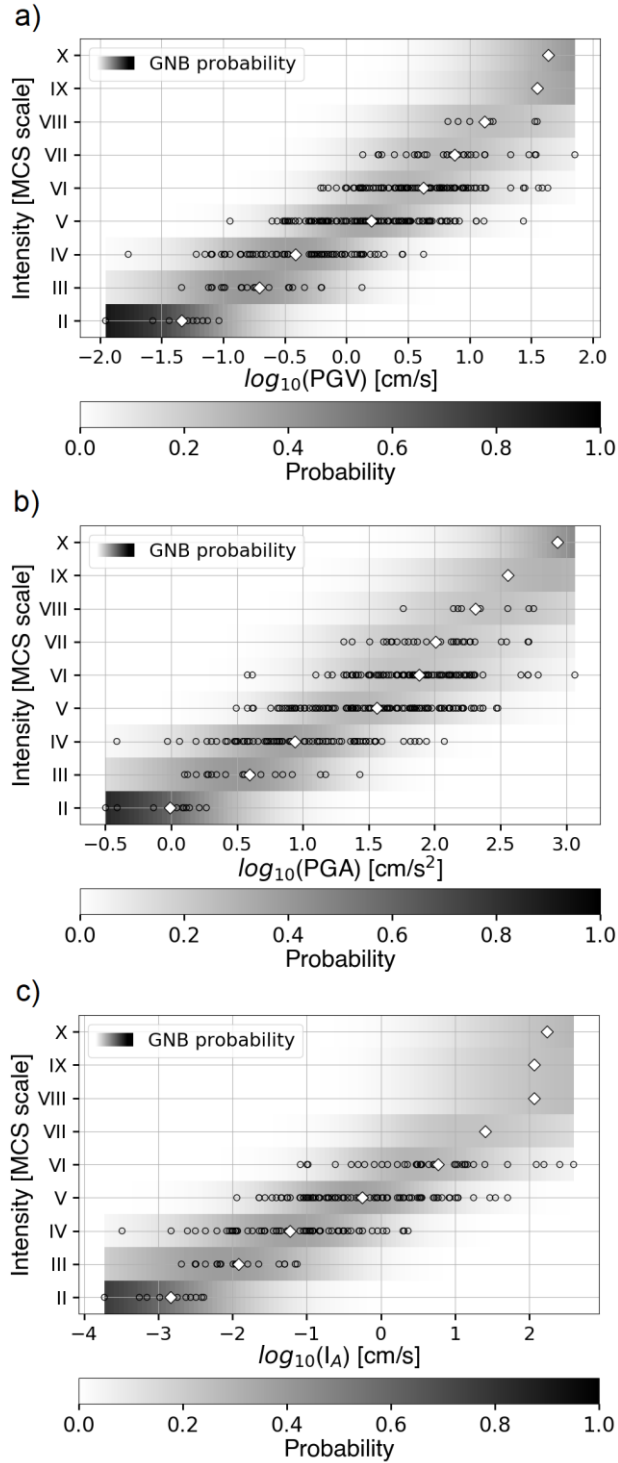
Figure 3-6: Probabilities obtained from GNB Classifiers (grey scale) for PGV (a), PGA (b) and I_A (c). Black circles depict the underlying dataset, white diamonds represent the mean GMP values used to derive the a priori probability distribution $\Pr(\log_{10} X = \log_{10} x_i \,|\, I = k)$. For each value on the X axis, the corresponding probability values along the Y axis sum up to one.

3.2.2 Ground Motion to Intensity Conversion Equations

Even if GNB is the preferred methodology selected for the definition of instrumental intensity, it is interesting to compare its results with the more standard formulation provided by GMICEs. The log-linear functional form itself was kept (Equation (1-2)), but pre- and post- processing was applied to try and make the methodology as intensity-compliant as possible. The first part of the pre-processing is described in detail in §3.1, while the resulting forecast values were rounded up to the nearest integer as a form of post-processing.

In order to consider both the dependent and the independent variables as affected by sampling variability, which is more correct given the nature of the used data, the ODR methodology (Boggs *et al*. 1988) was used to extract the intercept and gradient parameters appearing in the GMICE equations (*a*, *b* in Equation (1-2)) from the sample. ODR is a common technique for fitting data to models, as it minimizes the weighted orthogonal distances from the curve, taking into consideration both the vertical (σ_y) and horizontal (σ_x) uncertainties. In its simplest form, the ratio of the standard deviation of the errors on dependent and independent data (σ_y/σ_x) is assumed to be known and fixed. This use of uncertainties also makes it possible to directly invert the relation, so the regression coefficients could be likewise used to express the GMPs as a function of intensity.

Both intensity and strong motion data are characterized by an intrinsically high spatial variability, so pre-processing was needed to mitigate this feature. This is especially true for instrumental data, which is intrinsically punctual and strongly connected to specific, local geological conditions, whereas in the case of intensity this effect is mitigated by the definition of macroseismic classes as a collective measurement. The polar character of intensities also adds to the disequilibrium in the input dataset, mostly concentrated in the lower-central classes (IV − VI). A possible solution to address this variability is to perform a preliminary smoothing of data to filter out effects related to regional variability, random components, and geological conditions. Following Faenza and Michelini (2010), GMP-intensity couples were binned into integer intensity classes as a form of smoothing of the

instrumental data. The regressions were thus performed on the mean of the logarithmic GMP values in each bin, taking advantage of the fact that logarithmic data follow the normal Gaussian curves for which it is possible to estimate the underlying statistics. The weighted mean values μ_k^* and common standard deviations σ_{CSD} were estimated using the same procedure previously described in §3.2.1. The corresponding intensity values I were also assigned an error σ_I to account for the dispersion of data. After testing different binning on the GMP data, in order to check the corresponding discrete distribution of intensity values, a conservative value of $\sigma_I = 1.0$ was adopted as a common standard deviation associated with all intensity classes and all parameters.

For each of the eight considered parameter classes, GMICEs were calculated on nine data couples $(\mu_k^*, I_k = k)$, with associated errors (σ_{CSD}, σ_I) and intensity classes ranging from II to X (cf. Figure A-3 in the Appendix for a graphic representation). Regression parameters are reported in Table 3-1. To allow for a qualitative comparison of the results, for each equation the R squared value (R^2), representing the proportion of explained variance of I, and the standard deviation of the bins (σ) and of the data (σ_d) were calculated. The standard deviation of the bins is defined as:

$$\sigma = \sqrt{\frac{\sum_{k=2}^{10} \varepsilon_k^2}{9-2}}, \tag{3-6}$$

where $\varepsilon_k = I_k - \widehat{I_k}$ is the residual between predicted intensity value ($\widehat{I_k}$) and true intensity value (I_k) corresponding to μ_k^*. Due to the low sample population, a reduced form was preferred in which the number of intensity points used in the regression (nine) is reduced by the number of fitted parameters (*a* and *b*). Just like R^2, σ depends on residuals calculated on the binned dataset (ε_k). For this reason, it does not fully catch the actual underlying variability in *I*, and its values are way lower than the prior ones assigned to the input data ($\sigma_I = 1.0$). Following Gómez Capera *et al*. (2020), the standard deviation of the data σ_d was also calculated, as:

$$\sigma_d = \sqrt{\frac{\sum_{n=1}^{N_{tot}}(I_n - \widehat{I_n})^2}{N_{tot} - 1}} \ , \tag{3-7}$$

where $I_n - \widehat{I_n}$ is the residual calculated for the n-th input point. Values of R^2, σ and σ_d are reported in Table 3-1. σ_d values are close to 1 and provide a better measure of the variability in I for a given GMP value. Even so, since I is an ordinal variable, they cannot be used as-are and require some degree of interpretation. One possibility is to define a probability associated to each $\widehat{I_n}$, in the form of a Gaussian distribution centred on the forecasted intensity and with σ_d as standard deviation (cf. §3.1).

Table 3-1: The resulting GMICEs parameters (a, b) obtained through ODR regression for each considered GMP class, with associated R^2 value and the standard deviation values of the bins (σ) and of the data (σ_d).

		a	b	R^2	σ	σ_d	#records
GMP	PGD	7.01 ± 0.17	2.33 ± 0.15	0.97	0.49	1.24	376
	PGV	4.96 ± 0.17	2.65 ± 0.16	0.97	0.47	1.19	376
	PGA	1.32 ± 0.35	2.85 ± 0.19	0.97	0.51	1.36	376
	I_A	5.63 ± 0.23	1.46 ± 0.13	0.95	0.67	1.22	220
	I_H	3.58 ± 0.30	2.46 ± 0.21	0.95	0.66	1.20	200
	PSA03	0.65 ± 0.56	2.69 ± 0.25	0.94	0.73	1.32	220
	PSA10	2.73 ± 0.35	2.41 ± 0.20	0.95	0.64	1.28	220
	PSA30	4.78 ± 0.27	2.31 ± 0.22	0.94	0.74	1.31	195

3.3 Appraisal of the results

3.3.1   Leave-One-Out cross validation

The best way to assess the performance and reliability of the resulting intensity predictions, both from ODR and GNB, would be to test them on an "unseen" dataset, different from the one used to extract them. In the Italian case, the available database itself does not contain enough data to properly build both a training set and a testing set, so Leave-One-Out cross-validation (LOOCV) was applied as a proxy to assess the equations performance on unseen data.

LOOCV works by repeatedly dividing the whole dataset into two subsets: the one used to train the equations, containing N-1 points, plus a single point which is left out to be used for validation. The LOO system was constrained so that only points associated to integer intensity values (i.e., weight equal to one) would be left out as a test case. Each of these points was dropped in turn, so that the regressions were performed on the remaining data and used to estimate intensity on the left-out data point ($\hat{I}$). The classification ability with respect to the actual values (I) was then scored using the Cross-Entropy loss function $\mathcal{L}$ (C.E.; also called log-loss):

$$\mathcal{L}[P] = -\frac{1}{N} \sum_c \delta_{o,c} \, \log_{10}(P_{o,c}) \,, \tag{3-8}$$

where $\delta_{o,c}$ is 1 if the intensity value of observation o belongs to class c and 0 otherwise, and $P_{o,c}$ is the predicted probability that observation o has intensity class c.

The Cross-Entropy loss takes into consideration the probability associated to each intensity class, which should reflect the intrinsic variability in intensity values for a given ground-motion input. This allows to compare models not only on their average classification ability, but also on how well they capture the uncertainty. C.E. loss score increases as the predicted probability deviates from the actual label; it would be 0 for a perfect model. In the case of ODR forecasts, the

predicted probability was not directly available and had to be estimated by integrating the normal density centred on $\hat{I}$ over the interval $[I - 0.5, I + 0.5]$.

Table 3-2: The resulting Cross-Entropy (C.E.) loss scores for ODR regressions and for GNB Classifiers. Lower values indicate better performance of the algorithm.

	GMP							
	PGD	PGV	PGA	I_A	I_H	PSA03	PSA10	PSA30
C.E. (ODR)	1.70	1.65	1.82	1.67	1.71	1.80	1.75	1.77
C.E. (GNB)	1.49	1.42	1.53	1.39	1.38	1.42	1.44	1.50

GNB classifier models score better than ODR regressions for all ground motion parameters (Table 3-2), indicating an overall better performance of the GNB models. Among other parameters, the equations regarding PGV (in agreement with Kuehn and Scherbaum 2010), PSA03, I_A and I_H provide the best performances.

3.3.2   Spectral parameters

Both methodologies have a way to address the weakness arising from less populated classes. Even so, it should be kept in mind that models obtained from spectral parameters are still resenting the lack of data in high intensity classes ($I >$ VII). As discussed in Chapter 2, the available database included many waveforms for which only the peak amplitude parameters could be extracted without risk of underestimation. The consequent pre-processing led to less populated datasets for the case of spectral parameters. The distribution of such data (e.g., Figure 3-6.c; Figure A-2 in the Appendix) highlights how they especially lack in high intensity classes, which could lead to inconsistencies in the related forecasts. This holds true for both methodologies and is simply more evident in the case of GNB where it translates into not well resolved probability values. For this reason, it is

advisable that only the resulting models for PGA and PGV should be used for forecasts.

3.3.3 Sensitivity study

Both GNB and ODR models were tested against the underlying sample dataset by comparing the predicted classes with the observed ones, to check in which data ranges each performed better. Results are shown in form of weighted confusion matrices (Figure 3-7). The elements distributed along the highlighted diagonal are the number of data correctly categorized, while the off-diagonal elements are the misclassified data. GNB models provide more realistic outcomes for all classes with respect to ODR models, which also tend to a class overestimation (more elements on the right side of the diagonal). In both cases, PGV-based classification is more robust than the PGA-based one.
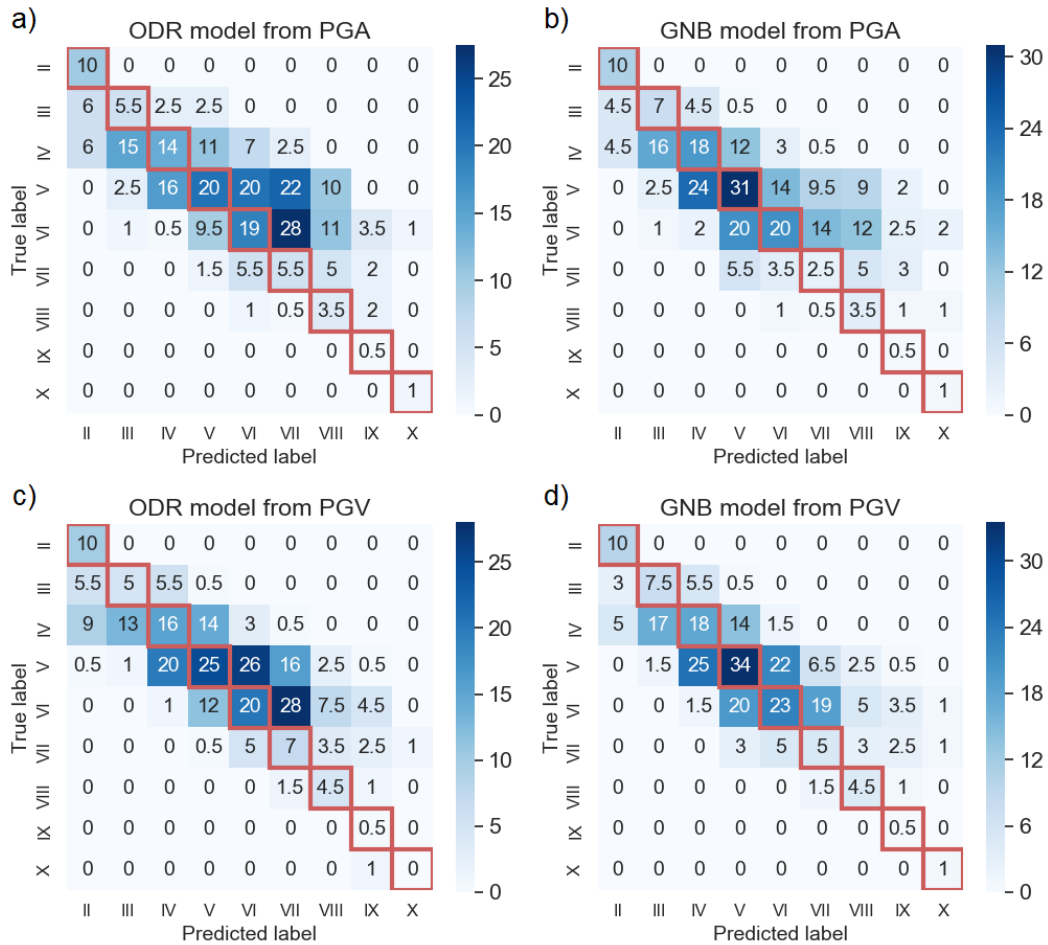
Figure 3-7: Weighted confusion matrices representing method classification on the training dataset. The true (observed) class labels are reported on the Y axis, the predicted labels are reported on the X axis. Elements distributed along the highlighted diagonal are the number of data correctly categorized. Results refer to the ODR (a) and GNB (b) methods for the PGA parameter (upper panels) and for the ODR (c) and GNB (d) methods for the PGV parameter (lower panels). From Cataldi *et al.* (2021).

3.3.4 Application of GNB forecasts

In order to be directly applicable into shaking intensity maps, GNB classification models have to be converted to GMICE-like objects. A single instrumental intensity value was assigned to each input ground motion parameter value in the database range. The forecast was chosen as the class with the highest associated probability (corresponding to the darkest colour in Figure 3-6). The resulting linear trend is comparable to the corresponding ODR equation. Results for the PGA and PGV cases are reported in Figure 3-8; they can be used as a guide in defining parameter ranges for each instrumental intensity class (cf. Table 3-3), to be preferred to the GMICE formulation. An example of application to data from Turkey can be found in Ertuncay *et al.* (2021).

Table 3-3: Table of PGA and PGV value intervals for calculation of intensity *I*, derived from GNB analysis.

		PGA_{min} [cm/s^2]	PGA_{max} [cm/s^2]	PGV_{min} [cm/s]	PGV_{max} [cm/s]
I_{MCS}	**II**	0.32	1.91	0.01	0.10
	III	1.91	6.31	0.10	0.28
	IV	6.31	17.78	0.28	0.74
	V	17.78	52.48	0.74	2.57
	VI	52.48	85.11	2.57	5.75
	VII	85.11	141.25	5.75	9.77
	VIII	141.25	269.15	9.77	21.38
	IX	269.15	575.44	21.38	39.81
	X	575.44	1148.15	39.81	70.79
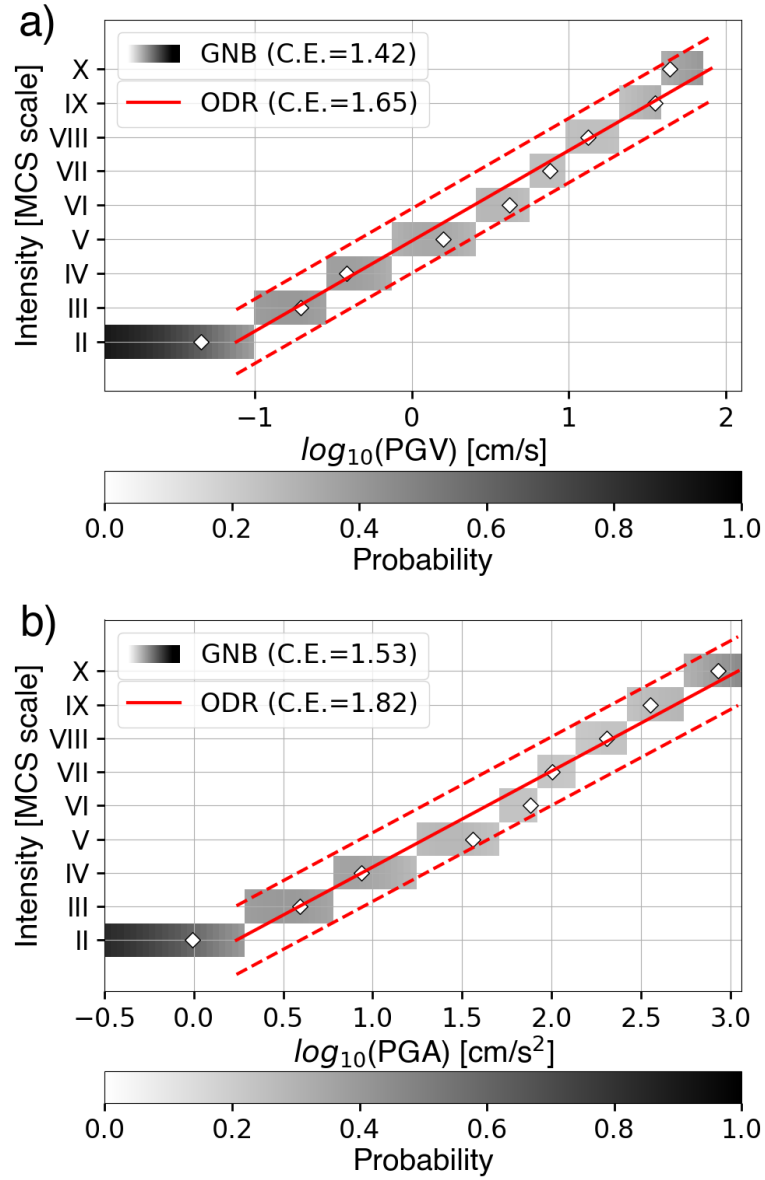
Figure 3-8: Intensity classes with highest associated GNB probability (grey scale) for each PGV (a) and PGA (b) value in the database range. The corresponding ODR equation with associated $\pm 2\sigma$ error is reported for comparison (red lines), together with the mean GMP values used to derive it (white diamonds).

3.3.5   Comparison with GMICEs from literature

GMICEs obtained in this study use integer classes only and are somehow more compliant to the MCS intensity scale than standard formulations that use half-integer values. Results are compared with a selection of relationships obtained for Italy and available in literature, namely the ones reported by Gómez Capera *et al*. (2020), Caprio *et al*. (2015), Faenza and Michelini (2010), and Faccioli and Cauzzi (2006). A summary of the characterizing parameters is reported in Table 3-4 and Table 3-5 for the PGV and PGA case, respectively.

The relations are consistent to each other inside the common standard deviation values estimated with the dataset used in this study (Figure 3-9). The main difference is in the reliability and range of validity of these laws, as those estimated using integer intensity classes have higher values of R^2 and a wider range of validity. In fact, the resulting equations present high R^2 values (over 0.90) for all the studied GMPs, rendering it impossible to choose a single parameter as providing the best estimate of intensity. The lowest standard deviation of data is associated to the regression line obtained for PGV ($\sigma_d = 1.19$). However, it should be kept in mind that results from cross validation and sensitivity tests indicate that the GNB models should be preferred to the ODR ones in any case.

Table 3-4: Comparison with Intensity - PGV relationships proposed by previous studies.

Author	Relationship	R^2	σ	*I* range
This study (ODR)	$I_{MCS} = (4.96 \pm 0.17) + (2.65 \pm 0.16) * \log_{10} PGV$	0.97	0.47	II to X
Faccioli and Cauzzi (2006)	$I_{MCS} = (5.09 \pm 0.22) + (1.80 \pm 0.17) * \log_{10} PGV$	0.61	0.71	V to VIII-IX
Faenza and Michelini (2010)	$I_{MCS} = (5.11 \pm 0.07) + (2.35 \pm 0.09) * \log_{10} PGV$	-	0.26	II to VIII
Caprio *et al.* (2015)	$I_{MCS} = (4.424 + 1.589 * \log_{10} PGV)$ if $\log_{10} PGV \leq 0.3$ $I_{MCS} = (4.018 + 2.671 * \log_{10} PGV)$ if $\log_{10} PGV > 0.3$	-	-	II to VIII
Gómez Capera *et al.* (2020)	$I_{MCS} = 4.514 * \exp(0.502 * \log_{10} PGV)$	-	0.36	II to X-XI

Table 3-5: Comparison with Intensity - PGA relationships proposed by previous studies.

Author	Relationship	R^2	σ	*I* range
This study (ODR)	$I_{MCS} = (1.32 \pm 0.35) + (2.85 \pm 0.19) * \log_{10} PGA$	0.97	0.51	II to X
Faccioli and Cauzzi (2006)	$I_{MCS} = (2.62 \pm 0.10) + (1.96 \pm 0.29) * \log_{10} PGA$	0.38	0.89	V to VIII-IX
Faenza and Michelini (2010)	$I_{MCS} = (1.68 \pm 0.22) + (2.58 \pm 0.14) * \log_{10} PGA$	-	0.35	II to VIII
Caprio *et al.* (2015)	$I_{MCS} = (2.270 + 1.647 * \log_{10} PGV)$ if $\log_{10} PGA \leq 1.6$ $I_{MCS} = (-1.361 + 3.822 * \log_{10} PGV)$ if $\log_{10} PGA > 1.6$	-	-	II to VIII
Gómez Capera *et al.* (2020)	$I_{MCS} = 2.276 * \exp(0.546 * \log_{10} PGA)$	-	0.31	II to X-XI
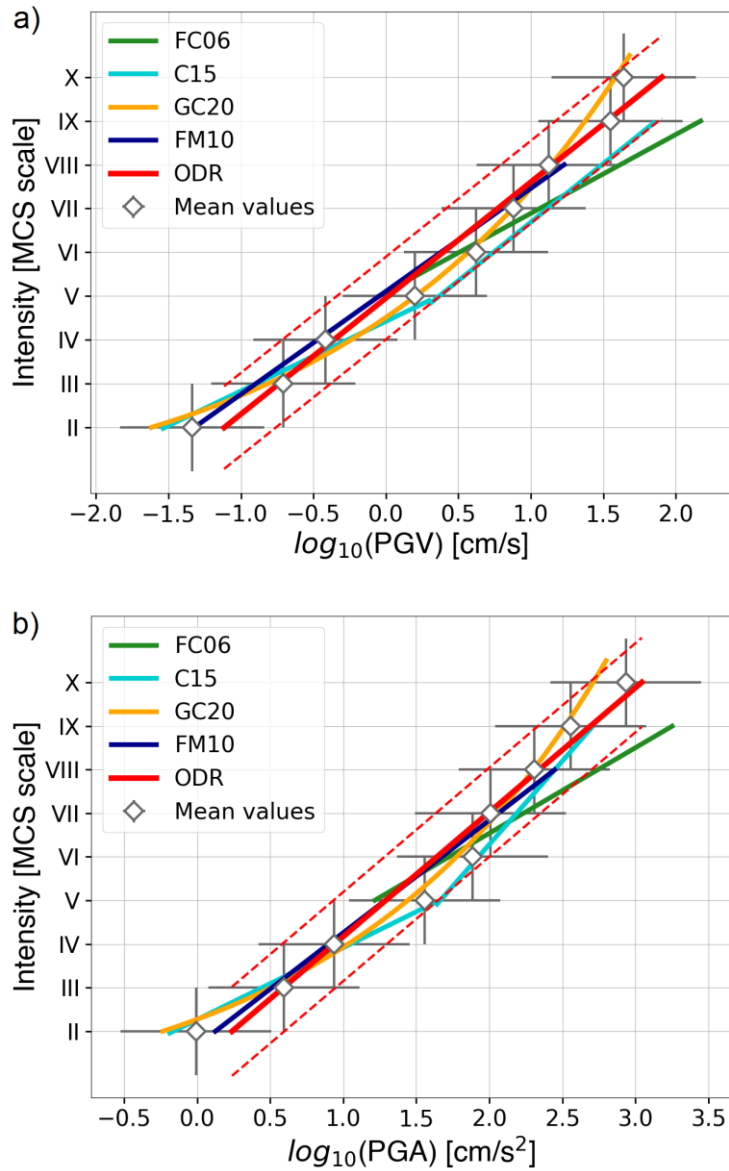
Figure 3-9: Comparison of the Intensity - PGV relationship obtained in this study through ODR regression on integer intensity classes and four previous studies that used half-integer classes: Faenza and Michelini (2010), FM10; Faccioli and Cauzzi (2006), FC06; Caprio *et al.* (2015), C15; Gómez Capera *et al.* (2020), GC20; b) Same as a), for the Intensity - PGA relationship. Dotted red lines are the ±2σ uncertainty associated to the ODR GMICEs.

3.4 Discussion

Macroseismic intensity is by definition an ordinal quantity and its classes are always defined as a "collective" measurement, coming from the observation of many factors which are not linearly dependent on any single, directly measurable parameter. Even so, it is common practice to treat intensity as a continuous value, for example by defining the instrumental intensity as a continuous linear function of the ground motion parameter values used to calculate it. A different approach was explored and applied to the Italian case by means of the Gaussian Naïve Bayes classifiers methodology. GNB allows to correctly handle ordinal data throughout the whole inversion process by estimating a discrete conditional probability distribution over integer intensity classes for any input value of the GMP parameter. Expert-assessed MCS intensity values in the range II-X were used, together with high quality accelerometric data. Results obtained in this work were published in Cataldi *et al.* (2021).

A set of ground motion parameters were considered as candidates for instrumental intensity definition, three peak parameters (PGA, PGV, PGD) and five spectral quantities (I_A, I_H, PSA03, PSA10, PSA30). The used dataset, composed of GMP-intensity couples observed for Italy, underwent a specific pre-processing procedure. This process was aimed at re-assigning half-integer intensity values to the nearby integer classes with the aid of a weighting factor. This ensured that only integer intensity values would be used in the calculations. The database was converted to its logarithmic form for its GMP part in order to gain better control on the underlying data distribution. Even if the results obtained for spectral parameters are promising, as they have good statistic scores, the associated available dataset used to derive them had lacks in higher-intensity classes. For this reason, it is advisable to use only the results obtained for peak parameters when a practical application is needed.

A set of linear equations in form of GMICEs were also estimated to provide a comparison for the GNB results. To make them more intensity-compliant and to

mitigate the intrinsic GNB variability, they were calculated on a binned version of the integer-valued database.

Instrumental intensity was defined as the class with the highest associated probability according to GNB classification and compared with the forecast obtained by rounding the GMICE output to the nearest integer value. Cross-Entropy loss scores, obtained through Leave-One-Out cross validation, indicated an overall better performance of GNB results both in terms of average performance on unseen data and of capability to capture data uncertainty. This result was confirmed by the sensitivity study carried out on the underlying dataset, which showed how GMICE results tend to an overestimation of the forecasts. Overall, GNB models perform better than GMICE ones on the whole considered intensity range, in terms of classification scores. Among the considered parameter classes, forecasts based on PGV appear to provide the best scores.

The GNB methodology goes beyond providing a single-valued intensity estimate, as it calculates a full discrete probability distribution for the MCS intensity classes in the considered range. GNB classification models were converted to GMICE-like objects to be directly suited for applicative contexts such as the creation of shaking maps. The resulting linear trends are comparable with GMICE curves but have the advantage of directly providing integer intensity values with an associated probability and overall better performance. As for GMICEs themselves, they are compatible with formulations available in literature for Italy, with the advantage of higher statistic scores.

The possibility to increase the estimate accuracy with respect to "standard" GMICEs might be extremely useful in some applications, such as shaking intensity maps. In fact, GMICEs are the default choice in generating shakemaps with the USGS-ShakeMap software (Wald *et al*. 1999c). Ertuncay *et al*. (2021) substituted them with the proposed conversion of GNB results to GMICE-like objects in the shakemap procedure in a case-study application on Turkish macroseismic data. Moreover, the GNB-based methodology is a machine learning oriented procedure that can be easily updated as more data is collected. In the era of big data, it can be included in the effort to efficiently analyse incoming data in

near-real time. A possible future development could include testing and calibration of the procedure on different areas, even at regional level, as soon as new, independent intensity data on new events becomes available. For example, it would be interesting to calibrate regional models for the South-eastern Alps region to improve the local real-time generation of shakemaps developed by Moratto *et al.* (2009). A wider database would also allow more focused studies both on multi-parametric relationships and on damages observed in near-fault areas.

4 Spectral inversion: modelling

A novel procedure for spectral inversion was elaborated based on preceding literature. In a general way, the velocity FAS of ground motion observed at a station j, originating from an earthquake i, for any frequency point k, can be represented as the product of a source (Ω), a propagation (D), and a site (S) term:

$$FAS_{ijk}(r_{ij}, f_k) = 2\pi f_k \, \Omega_i(f_k) \, D_{ij}(r_{ij}, f_k) \, S_j(f_k) \, I_j(f_k) \,. \qquad (4\text{-}1)$$

Equation (4-1) can be linearized by taking the logarithm:

$$\log FAS_{ijk}(r_{ij}, f_k) = \log(2\pi f_k) + \log \Omega_i(f_k) + \log D_{ij}(r_{ij}, f_k) + \qquad (4\text{-}2)$$
$$\log S_j(f_k) + \log I_j(f_k) \,.$$

The inverse problem represented by Equation (4-2) is a classical linear system of the form $\boldsymbol{Ap} = \boldsymbol{y}$, where $\boldsymbol{p}$ is the vector of the model parameters, $\boldsymbol{y}$ is the data vector of the logarithmic spectral amplitudes, and $\boldsymbol{A}$ is the system matrix relating them. Some constraints are needed to solve this system, both on the attenuation operator (e.g., Castro *et al.* 1990) and either on the source or site response (Andrews 1986). Parametric and non-parametric approaches are both viable strategies to handle this inverse problem.

The choice of the solving approach depends on the characteristics of the data to be used. Non-parametric inversion methods allow to gain more insight in the potential complexity of some of the spectral features, such as the attenuation terms, but need a dataset with excellent distance and azimuth coverage for their results to be reliable. Parametric algorithms need less data restrictions, as they work by iteratively perturbing the parameter set to minimize the difference between observed data and synthetic parametric models.

A flexible parametric inversion strategy was developed and converted into a *python* algorithm to be released as an open-source project. The parametric

approach was preferred in order to obtain a tool that could efficiently work on any dataset, regardless of the completeness of azimuth and distance coverage. The aim was to apply it to the case study area of Northeast Italy, as extensively described in Chapter 5.

4.1 Method

The parametric method iteratively tries to fit the observed FAS with respect to the chosen parametric model by minimizing the loss function, which describes the difference between observed data and the synthetic parametric model. At each iteration, the parameter set is perturbed using information from the Jacobian function and the corresponding loss function is calculated. The process is repeated until a threshold condition is met, and the parameters are kept as a solution.

4.1.1 Selection of the forward model

As discussed in the Introduction (§1.2.2), the Fourier velocity spectrum observed at each of N_j stations, originating from any of N_i earthquakes, for each of N_k frequency points, can be represented as the product of a source (Ω), a propagation (D), and a site (S) term. The actual parametrization of each term depends on the used dataset and consequently on the assumptions on source, site and path characteristics. If the data is correctly deconvolved with the seismograph response functions before inversion, the instrument response function term $I(f)$ can be set equal to one.

Parametrization provided by Bora *et al*. (2017) for European shallow crustal earthquakes and by Edwards *et al*. (2008) for the United Kingdom were considered as candidates for the forward modelling function, together with the one presented in Malagnini *et al*. (2002) which was specifically developed for the Northeast Italy region.

The source term $\Omega(f)$ is modelled using Brune (1970, 1971) point source model with a single corner frequency, to represent the far-field velocity Fourier spectrum:

$$\Omega_i(f_k) = \Omega_{0i} \frac{1}{1 + \left(\frac{f_k}{f_{ci}}\right)^2} , \qquad (4\text{-}3)$$

where f_c is the source corner frequency and Ω_0 is the long-period plateau value at the source, given by:

$$\Omega_{0i} = \frac{\xi F \Theta_{\lambda\varphi}}{4\pi v_S^3 \rho R_0} M_{0i} , \qquad (4\text{-}4)$$

The quantities on which Ω_0 depends are the seismic moment M_0, the average radiation pattern $\Theta_{\lambda\varphi}$ (= 0.55 for S waves; Boore and Boatwright 1984), the near surface amplification factor F (= 2), the factor to account for the partition of total shear-wave energy into two horizontal components ξ (= $1/\sqrt{2}$), the reference distance used for normalization R_0, the average density near the source ρ (= 2800 kg/m^3; Boore 1983, 2003) and the shear wave velocity in the proximity of the source v_S (= 3500 m/s). As the database is composed of low magnitude events ($M_L \le 4.6$), the far-field approximation holds even for shorter distances (Bora *et al*. 2017).

The propagation term $D(r, f)$ is composed of two parts:

$$D_{ij}(r_{ij}, f_k) = B(r_{ij}, f_k)\, G(r_{ij}, f_k) , \qquad (4\text{-}5)$$

where $B(r, f)$ is the attenuation along the path of propagation and $G(r, f)$ is the apparent geometrical spreading function.

Assuming whole path attenuation, $B(r, f)$ is parametrized as:

$$B(r_{ij}, f_k) = e^{-\pi f_k r_{ij}/(Q v_s)} , \qquad (4\text{-}6)$$

where Q is the quality factor related to high-frequency attenuation. Some studies, Malagnini *et al*. (2002) and Drouet *et al*. (2008) among them, suggested a frequency dependent Q factor in the form:

$$Q(f_k) = Q_0 \left(\frac{f_k}{f_0}\right)^{\eta},$$

(4-7)

where Q_0 is the reference value at $f_k = f_0 = 1$ Hz and η ranges from 0 to 1. However, other studies (e.g., Morozov 2008, 2009) have underlined how the distinction between a frequency-dependent quality factor ($\eta > 0$) and geometric attenuation is ambiguous, from a modelling point of view. Edwards *et al.* (2008) even showed how a frequency-dependent quality factor can lead to a strong trade-off with the geometrical spreading parameters. The model formulation for this work was thus restricted to a frequency-independent $Q = Q_0$ model by imposing $\eta = 0$.

The apparent geometrical spreading term $G(r, f)$ collects many different contributions, such as phase interference, scattering, and dispersion (e.g., Atkinson and Mereu 1992). Many models have been proposed in literature in the form of a piecewise function where each segment presents constant exponential decay. A generic representation of its frequency-independent formulation is given by:

$$G(r_{ij}) = \begin{cases} \left(\dfrac{R_o}{r_{ij}}\right)^{\lambda_1} & r_{ij} \leq \tilde{r}_1 \\ G(r_1)\left(\dfrac{r_1}{r_{ij}}\right)^{\lambda_2} & \tilde{r}_1 \leq r_{ij} \leq \tilde{r}_n \\ \vdots & \vdots \\ G(r_n)\left(\dfrac{r_n}{r_{ij}}\right)^{\lambda_n} & r_{ij} \geq \tilde{r}_n \end{cases}$$

(4-8)

where $\boldsymbol{\lambda} = [\lambda_1, \dots, \lambda_n]$ is a set of constant exponential decay factors corresponding to distances $\tilde{\boldsymbol{r}} = [\tilde{r}_1, \dots, \tilde{r}_n]$. It should be noted that R_0 is the same reference distance appearing in Equation (4-4). It represents the maximum distance for which apparent geometrical spreading is assumed to be null. It is used to take into account the rupture source dimension as it deviates from the point-source

description; for this reason, it also appears in the formula used to compute the seismic moment to restore the point-source assumption. Following Edwards *et al.* (2008), the source size normalization was set to $R_0 = 1$ km without affecting the geometrical spreading or moment values. Drouet *et al.* (2008) suggested a simple frequency-independent function in the form $G(r) = 1/r^\lambda$. Bora *et al.* (2017) derived a slightly more structured model with two sub-functions, using low frequency Fourier spectral amplitudes to minimize the trade-off with the attenuation factor. Malagnini *et al.* (2002) proposed a more complex model describing a slightly frequency-dependent geometrical spreading, specific to Northeast Italy, in the form:

$$G(r_{ij}) \,|\, f_k \leq 1\,\text{Hz} = \begin{cases} \left(\dfrac{1}{r_{ij}}\right)^{1.0} & r_{ij} \leq \tilde{r}_1 \\[2mm] G(r_1)\left(\dfrac{r_1}{r_{ij}}\right)^{1.6} & \tilde{r}_1 \leq r_{ij} \leq \tilde{r}_2 \\[2mm] G(r_2)\left(\dfrac{r_2}{r_{ij}}\right)^{1.2} & \tilde{r}_2 \leq r_{ij} \leq \tilde{r}_3 \\[2mm] G(r_3)\left(\dfrac{r_3}{r_{ij}}\right)^{1.3} & \tilde{r}_3 \leq r_{ij} \leq \tilde{r}_4 \\[2mm] G(r_4)\left(\dfrac{r_4}{r_{ij}}\right)^{0.5} & r_{ij} > \tilde{r}_4 \end{cases} \qquad (4\text{-}9)$$

$$G(r_{ij}) \,|\, f_k > 1\,\text{Hz} = \begin{cases} \left(\dfrac{1}{r_{ij}}\right)^{0.95} & r_{ij} \leq \tilde{r}'_1 \\[2mm] G(\tilde{r}_1)\left(\dfrac{\tilde{r}_1}{r_{ij}}\right)^{1.2} & \tilde{r}'_1 \leq r_{ij} \leq \tilde{r}'_2 \\[2mm] G(\tilde{r}_2)\left(\dfrac{\tilde{r}_2}{r_{ij}}\right)^{1.8} & \tilde{r}'_2 \leq r_{ij} \leq \tilde{r}'_3 \\[2mm] G(\tilde{r}_3)\left(\dfrac{\tilde{r}_3}{r_{ij}}\right)^{1.2} & \tilde{r}'_3 \leq r_{ij} \leq \tilde{r}'_4 \\[2mm] G(\tilde{r}_4)\left(\dfrac{\tilde{r}_4}{r_{ij}}\right)^{0.5} & r_{ij} > \tilde{r}'_4 \end{cases} \qquad (4\text{-}9b)$$

where $\tilde{r} = [50, 60, 80, 100]$ km and $\tilde{r}' = [40, 50, 60, 100]$ km.

The site transfer function can be described as:

$$S_j(f_k) = A_j \, a_j(f_k) \, e^{-\pi f_k \kappa_{0j}/f_k^{\eta}} \,, \tag{4-10}$$

where A_j is a frequency-independent correction factor relative to the chosen reference profile, κ_{0j} is a constant site-related attenuation operator (e.g., Anderson and Hough 1984), and $a_j(f_k)$ is the frequency-dependent site amplification function that takes into account resonant frequencies due to the layered, fractured subsurface (e.g., Steidl *et al.* 1996). A constraint on the used reference must be imposed to ensure the physical meaning of the correction factor terms A_j; an example is provided in §5.3. In agreement with the choice made on Q factor parametrization, frequency-independent whole-path attenuation $(\eta = 0)$ is assumed. As for the frequency-dependent site amplification function, it noticeably represents the most complex parameter in the model, being it a set of amplification values at each sampled frequency, for each station (i.e., $N_j N_k$ elements). Preliminary stability tests were run to check if it was possible to reproduce a known empirical amplification function with the available dataset composition. Station CARC was used as a benchmark, for which empirical amplification curves from noise analysis are available from literature (Fitzko *et al.* 2007). Synthetic tests showed that the frequency-dependent amplification term could not be fully solved by direct parametrization with the used database configuration. The preferred approach was thus to exclude the term $a_j(f_k)$ from the forward model and leave it for calculation through residual analysis instead, as described in §4.1.4. Consequently, the actual local site amplification term used for the parametrization becomes:

$$S_j(f_k) = A_j \, e^{-\pi f_k \kappa_{0j}} \,. \tag{4-11}$$

In addition, three uncertainty terms were added to model the overall error related to source (ε^{SO}), propagation (ε^P) and site (ε^{SI}) modelling blocks respectively.

The forward spectral model used for inversion is obtained by putting all the contributions together and taking the logarithm:

$$\log FAS_{ijk}(r_{ij}, f_k) = \log(2\pi f_k) + \log\left(\frac{\Theta_{\lambda\varphi}F\xi}{4\pi\rho v_s^3 R_0}\right) + \log M_{0i}$$

$$- \log\left(1 + \left(\frac{f_k}{f_{ci}}\right)^2\right) + \log G(r_{ij}, f_k) \qquad (4\text{-}12)$$

$$- \frac{\pi f_k r_{ij}}{v_s Q_0} + A_j - \pi f_k \kappa_{0j} + \varepsilon_i^{SO} + \varepsilon^P + \varepsilon_j^{SI}$$

where $G(r_{ij}, f_k)$ is defined in Equation (4-9). Notice that $\boldsymbol{\varepsilon^{SO}}$, ε^P and $\boldsymbol{\varepsilon^{SI}}$ act as a collector for all the uncertainties related to the correspondent contribution, as a mean value over all frequencies. For example, ε^P represents the average propagation uncertainty term over all events and sites, while $\boldsymbol{\varepsilon^{SO}} = [\varepsilon_1^{SO}, \dots, \varepsilon_{N_i}^{SO}]$ models the overall source uncertainties (one for each event) coming from both the seismic moment and the corner frequency estimates. Other solutions could be used to estimate the uncertainties than this parametrization choice. For example, the error on source localization could be considered negligible and set to zero, or a model without uncertainty parametrization could be used and Generalized Inversion Technique applied on the inversion residuals to retrieve uncertainties related to each term. The software used to implement the model was purposely built in a flexible way so that uncertainty parametrization can be modified or not used at all, depending on the specific application. An example of different choices for uncertainty estimation is presented in the chapter dedicated to Validation (§6.1.1).

4.1.2 Selection of the loss function

Any optimization problem seeks to minimize a loss function that maps the difference between true and estimated values for a given instance of data. As the optimization process is dependent on the loss function, the choice for its functional form should be tailored to the quality and type of data for the problem at hand.

The norm used to define the parameter estimates plays an important role; in general, it can be defined based on the class of L_P norms (Rice and White 1964):

$$\|y_i - a_P\|_P = \left[\sum_{i=1}^{m}|y_i - a_P|^P\right]^{1/P}, \tag{4-13}$$

where $1 \leq P \leq +\infty$, y_i is one data point among a dataset of m elements and a_P is the L_P estimator of the data. If $P = 1$, the norm provides an estimate of the median value among the dataset. If $P = 2$, it provides an estimate of the mean value. Depending on the characteristics of data errors, one formulation or the other should be preferred. The loss function based on L_1 is known as least absolute deviations and is more robust in terms of resistance to outliers in the dataset, but may allow multiple solutions. The loss function based on L_2, also called least squares error, provides a single and more stable solution. When applied to the FAS inversion context, the absolute amplitude fit (L_1) can be written as:

$$LF = \sum_{i,j,k}\frac{\left|FAS_{ijk}^o - FAS_{ijk}^m\right|}{\sigma_{ijk}}; \tag{4-14}$$

the least-squares fit (L_2) can be written as:

$$LF = \sum_{i,j,k}\frac{\left[FAS_{ijk}^o - FAS_{ijk}^m\right]^2}{\sigma_{ijk}}; \tag{4-15}$$

and their respective log-space formulations ($\log - L_1$, $\log - L_2$) are:

$$LF = \sum_{i,j,k}\frac{\left|\log\left[FAS_{ijk}^o\right] - \log\left[FAS_{ijk}^m\right]\right|}{\log\left[\sigma_{ijk}\right]}; \tag{4-16}$$

$$LF = \sum_{i,j,k}\frac{\left[\log\left[FAS_{ijk}^o\right] - \log\left[FAS_{ijk}^m\right]\right]^2}{\log\left[\sigma_{ijk}\right]}; \tag{4-17}$$

where superscripts o and m denote spectral values for the observed and the modelled data, respectively.

The L_1 and L_2 estimators result in a possibly biasing automatic weighting that gives preference to the higher amplitudes, whereas the log-space minimization

functions homogeneously fit the different sections of the spectra. A careful and homogeneous data processing is assumed, complete with noise subtraction, to ensure that spectra have low level of noise over all different frequency bands. Under this condition it is safe to use the log-space estimators, and the variance of each spectral data point (σ_{ijk}) can be assumed equal so that it does not affect the minimization.

The loss function selected for the inversion procedure is a modified version of the $\log - L_2$ formulation (Equation (4-17)):

$$LF = \sum_{i,j,k} \frac{\left[M_{ijk}\left(\log\left[FAS_{ijk}^o\right] - \log\left[FAS_{ijk}^m\right]\right)\right]^2}{N_{TOT}\,W_{ijk}}, \qquad (4\text{-}18)$$

where M is a matrix of Boolean masks indicating which frequency points of each spectrum should be used for inversion (for example, based on Signal-to-Noise Ratio; cf. §5.2), $N_{TOT} = N_i N_j N_k$ is the normalization factor related to the number of points used in the inversion, and W is an optional weight matrix.

The term W was introduced to increase the flexibility in the choice of the functional form. If $W_{ijk} = 1$, the formulation reverts to an almost standard $\log - L_2$ loss function. A frequency-independent choice such as $W_{ijk} = W_{ij}$ allows to give more weight to certain stations and/or events. This could be useful in the case of non-homogeneous datasets or to calibrate the inversion on specific parameters; for example, W_{ij} could be proportional to the noise level of each record. Frequency-dependent W_{ijk} could be used to focus the inversion on specific frequency bands. For example, setting $W_{ijk} = W_k = 1/f_k$ would down-weight the high-frequency content of the spectra in favour of fitting the low-frequency plateau and source corner frequency.

4.1.3 Possibility of multi-step inversion

As the method had to be applied to the Northeast Italy dataset, a preliminary synthetic test was made to check if the use of frequency-dependent weights W_{ijk}

could improve the solution accuracy for a comparable dataset. A detailed description of the Northeast Italy dataset characteristics is provided in §5.2. The test consisted in a two-step inversion, and it was made feasible by the possibility of *a priori* fixing parameters during the inversion process (cf. software description in §4.2.2). Two specular low-pass and high-pass Butterworth filters were built for each spectrum, defined with respect to the corner frequency f_c estimated for that event (Figure 4-1). For the first step of the inversion, the weight matrix W_{ijk} to be used in Equation (4-18) was populated using the gain values of the low-pass filters, so that the inversion itself was performed on low frequency bands only. This allowed to focus the inversion on the parameters most affected by frequency values below f_c, namely seismic moments M_0 and the quality factor Q_0. Parameters obtained from step one were used as input for step two, with the constraint that values obtained for M_0 and Q_0 should be kept as constants. For step two of inversion, the weight matrix was instead built from the high-pass filters.
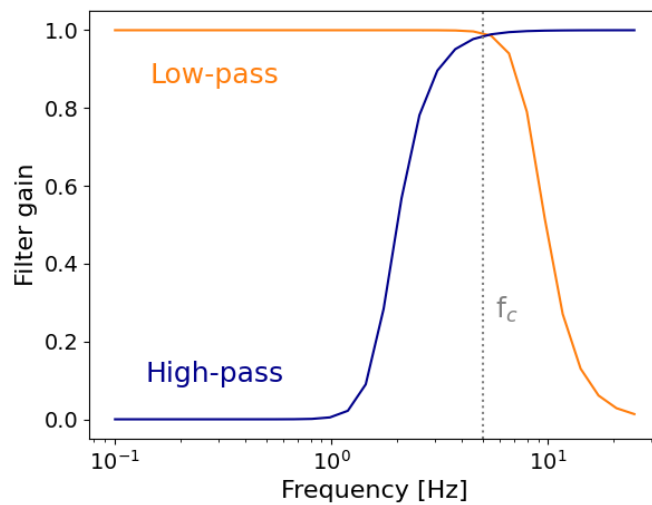
Figure 4-1: Low-pass and high-pass Butterworth filters built for a sample spectrum related to an event with $f_c = 5$ Hz. Filter gain values are used as frequency-dependent terms W_k that build the weight matrices W for the two-step inversion.

A two-step inversion could in principle provide an improvement in the definition of parameters, as different parameters might be more sensible to different frequency bands. However, for the used dataset composition, the synthetic test did

not lead to a meaningful improvement in terms of precision of the output parameters, and consequently also in terms of loss function score.

4.1.4 Inversion for frequency-dependent site functions

As discussed in §4.1.1, site response curves are left unmodelled by Equation (4-12) and must be treated separately. The factorial residual of the inversion, given by:

$$\vartheta_{ij}(f_k) = \frac{FAS^o_{ijk}}{FAS^m_{ijk}},\tag{4-19}$$

is used to reconstruct the frequency-dependent site functions by taking the log-space geometric mean of the factorial residuals at each discrete frequency f_k over all events ($i = 1, \dots, N_i$), at each station j (e.g., Scherbaum 1990):

$$\log(a_j(f_k)) = \frac{1}{N_i}\left\{\sum_{i=1}^{N_i} \log(\vartheta_{ij}(f_k))\right\}.\tag{4-20}$$

Even if these site response functions are not directly obtained during the inversion, they are by definition the residuals of the previous inversion and should be interpreted carefully. The algorithm cannot distinguish between actual residuals to the model coming from the frequency-dependent amplification and those coming from noise and instrument calibration errors. For this reason, a careful data processing is needed to ensure that instrument effects at least are correctly subtracted.

For the same reason, the uncertainty terms $\boldsymbol{\varepsilon}^{SO}$, ε^P and $\boldsymbol{\varepsilon}^{SI}$ were added to the model, so that the algorithm could distinguish residual contributions tied to a specific event or site. As previously noted, these uncertainty terms should not be read as errors related to a single parameter, but rather to a whole parametric class, as in source, propagation or site. A preliminary sensitivity test was conducted to check the need for a site uncertainty term. Even if the frequency-dependent component related to site uncertainty should remain unmodelled and contribute to

the factorial residuals, a frequency-independent term could gather the uncertainty related to the parametrized site contributions A_j and κ_{0j}. This feature might be particularly helpful in the light of the possible bias introduced by trade-off between κ_{0j} and Q_0. Consequently, $\boldsymbol{\varepsilon^{SI}} = [\varepsilon_1^{SI}, \dots, \varepsilon_{N_j}^{SI}]$ was modelled as only depending on the stations, so that it could collect uncertainties related to other site parameters which are in fact frequency-independent. This might in principle introduce a different trade-off, between A_j and ε_j^{SI}, as both terms have the same functional form and are only dependent on the station index. This problem is actually solved from the beginning by the use of the reference site constraint, which is only imposed on site correction terms A_j rendering them mathematically distinguishable from ε_j^{SI}.

4.2 Software

Many algorithms for parametric inversion of FAS exist in literature, some of which are publicly available as source code (e.g., *SpecMod*, Edwards *et al*. 2010). Some are tailored too strictly to the case studies they were developed for and therefore cannot easily be modified for application on a different dataset. In other cases, the software is built as a flexible and modular structure, but some inversion functionalities are missing.

A versatile parametric spectral inversion tool was developed that could also be applied to the problem at hand. The software was built in a modular way to ensure the possibility of different parametrization choices. It relies on the package *scipy* (Virtanen *et al*. 2020) and makes use of the built-in function *scipy.optimize.minimize* to perform the inversion itself. The algorithm recursively fits the observed FAS with respect to the chosen parametric model. At each iteration, the parameter set is perturbed using information from the Jacobian function and the corresponding loss function is calculated, until a threshold condition for minimization is met. A loss function weighting feature was added that is not present in software available from literature, as well as the possibility to

54

fix or free parameters. As discussed in more detail in §4.2.2, these implementations open the possibility of performing multi-step regressions without need to resort to additional software and of easily imposing constraints on the inversion procedure.

4.2.1 Optimization methodology

The chosen parametrization used for Equation (4-12) makes the problem nonlinear and requires that the equations should be linearized prior to the fit.

The methodology selected as preferred to perform the inversion is the Sequential Least Squares Programming (SLSQP; Kraft 1988). SLSQP is an iterative constrained nonlinear optimization methodology that minimizes a function of multiple variables with any given bound and equality/inequality constraint. At any iteration, it defines an appropriate search direction in the parameter space by applying Broyden–Fletcher–Goldfarb–Shanno algorithm (BFGS; Nocedal and Wright 2006), a quasi-Newton method that uses first derivatives only, to optimize the quadratic model of a Lagrange function consisting of both loss function and applied constraints.

4.2.2 Software structure

Even if it was oriented towards a specific application, the rationale behind the structure of the software was to provide a generic framework to the iterative minimization procedure. The *python* inversion software based on the *scipy.optimize.minimize* package was built with a flexible and highly customizable structure, so that it can be used either with the SLSQP algorithm or any other nonlinear optimization methodology. A flexible build also allows to modify the choice for the forward model, and to similarly perform parametric inversion on any other dataset beyond the one at hand.

In this perspective, python classes are defined for easily handling different object types. A *Spectrum* class is used to store the spectral amplitude values and their

related information. Such metadata include: frequency point values; the Boolean mask indicating which frequency points should be used for inversion, and the corresponding cut-off frequencies; hypocentral distance; station name and channel; event number and magnitude. A *Parameter* class is defined to easily access and modify variable quantities used in the parametric function that models the spectra. Its properties include the parameter value, upper and lower bound values for its range of definition and a Boolean indicator for possibly fixing the parameter value. A *StaticParameter* class allows to store matrices containing all *Spectrum* objects in the dataset and their corresponding Boolean arrays and hypocentral distances, together with information on the number of events, stations, and frequency points.

The software developed to perform the inversion is based on a modular framework, so that each component can be easily modified without affecting the whole structure. Input observed data (previously stored as a *pickle* object) is read, and correspondent synthetics are generated using the forward model with given initial parameter values. The loss function describing the difference between observed and synthetic models is evaluated. If its value is below a certain threshold, the algorithm stops and the parameters are kept as a solution, otherwise the parameter set is perturbed and the process is repeated for the new point, until the threshold condition is satisfied. The four main modules that are used throughout this process are the forward modelling module, the loss function and Jacobian function modules, and the wrapper module to *scipy.optimize*. Each of these components makes use of the previously defined object classes.

Module for forward modelling

The forward modelling tool in turn has a modular structure that treats each component (source, site, and path) independently. It represents a generalization of the model presented in Equation (4-12):

$$\log FAS_{ijk}\left(r_{ij}, f_k\right) = n\log(2\pi f_k) + \log\left(\frac{\Theta_{\lambda\varphi}F\xi}{4\pi\rho v_s^3 R_0}\right) + \log M_{0i}$$

$$- \log\left(1 + \left(\frac{f_k}{f_{ci}}\right)^2\right) + \log G\left(r_{ij}, f_k; R_0, \tilde{r}, \lambda\right) \quad \text{(4-21)}$$

$$- \frac{\pi f_k r_{ij}}{v_s Q_0 f_k^\eta} + A_j + a_j(f_k) - \frac{\pi f_k \kappa_{0j}}{f_k^\eta}$$

$$+ \varepsilon_i^{SO} + \varepsilon^P + \varepsilon_j^{SI}.$$

Its structure relies on four classes of parameters:

- Source parameters: seismic momentum (M_0) and corner frequency (f_c);
- Path parameters: quality factor parameters (Q_0 and η), parameters characterizing the piecewise apparent geometric spreading function ($R_0, \tilde{r}$ and λ);
- Site parameters: frequency-dependent ($a(f)$) and frequency-independent (A) amplification terms, site-related attenuation factor parameters (κ_0 and η);
- Uncertainty parameters: $\boldsymbol{\varepsilon^{SO}} = [\varepsilon_1^{SO}, \dots, \varepsilon_{N_i}^{SO}]$ for the source block, ε^P for the propagation block, and $\boldsymbol{\varepsilon^{SI}} = [\varepsilon_1^{SI}, \dots, \varepsilon_{N_j}^{SI}]$ for the site block.

Each term is implemented separately so that it can be easily modified depending on the chosen model: for example, the number of $\boldsymbol{\lambda}$ instances appearing in the path term can be modified to reflect the regional geometric spreading function of the area under investigation. Similarly, the uncertainty parameters can be turned off. As the model refers to the logarithm of the FAS, terms related to each parameter are calculated separately and then simply added to build the corresponding spectral value. An extra parameter n (i.e., the exponent to the term $2\pi f$) allows to choose the spectrum component, either acceleration, velocity, or displacement.

Modules for calculating loss function and Jacobian function

The choice of the loss function determines the criterion used to identify the best fitting model and therefore is an important tool in the inversion process. The loss

function depends both on the used data and on the choice for the forward model. The module for its calculation is structured to allow a choice among the most used log-norms, and can be easily modified to add other norm calculation methods. Besides the standard normalization by dividing for the number of used data points, the possibility of dividing by weights is also implemented. Weights can either be uniform in the frequency domain or differentiated, so that multiple configurations could be easily applied depending on the studied dataset and on the desired result. For example, applying a weight equal to $2\pi f$ is similar to transforming the parameter space into the log-log domain (Poggi *et al*. 2011) and can be used to down-weight the higher frequency content of the spectra (cf. §4.1.3).

It should be noted that almost all inversion methods implemented in *scipy.optimize* support an automatic calculation of the Jacobian function of the provided loss function. This usually consists in using either 2-point or 3-point interpolation to calculate partial first derivatives, with consequent increase in the computational time and loss of precision in the result. A specific module was implemented to pass the actual analytic formulation of the Jacobian function to the *scipy* wrapper for inversion. If the loss function is modified, its analytic Jacobian function can be calculated and readily updated in the corresponding module, to ensure more control on the inversion process and higher numeric precision.

Wrapper module to *scipy.optimize*

The inversion itself is handled by a wrapper module that acts as interface between the *scipy* tools and the desired parametric setup. The user can define the desired inversion method (among those implemented in *scipy.optimize*), loss function and Jacobian function to be passed to the *scipy* inversion tool. Additionally, this module exploits the use of the *Parameter* class. Upper and lower bound values for each parameter are read from the input *Parameter* objects and passed to the inversion tool to confine the parameter space of the inversion results. This feature can be further pushed to fix a parameter so that the algorithm treats it as a

constant. When the Boolean indicator is set to False, the wrapper sets the upper and lower bounds for inversion equal to the parameter value itself. With this setup, the inversion tool works on a reduced version of the problem, where some of the formerly parameter objects are in fact constant objects, so that the result is a real solution of the constrained problem. The Boolean indicator is also shared with all other modules. This way, once the parameters to be fixed are selected, the loss function and the Jacobian function are automatically updated before being passed to the inversion tool. The possibility to treat parameters as fixed opens the possibility of performing multi-step regressions without need to resort to additional software. Moreover, the wrapper module also allows to impose constraints on the inversion procedure. This feature is implemented for use with the SLSQP methodology, which is the only *scipy.optimize* method that currently supports constraints; if any other inversion method is selected, it is ignored. The constraint imposition can be used to set the reference station: for example, it can force the sum of all site amplification terms at all used sites to be zero, so that the site amplification parameters are defined as relative to the network average.

Additional modules

Besides the four modules composing the inversion software, a separate tool was developed to perform the residual analysis, reconstruct the frequency-dependent site functions as described in §4.1.4, and produce a graphic rendering.

5 Spectral inversion: application

The modelling and inversion techniques described in Chapter 4 were applied to the Northeast Italy case study area. The actual parametrization of the forward model described by Equation (4-12) was also implemented based on the specific setting to be used. The aim was both to test the goodness of the developed algorithm and to extract information on site response functions via analysis of residuals.

5.1 Case study area

A case study area for spectral inversion was selected in the Northeast part of Italy, at the boundary with Slovenia and Austria.

The main orographic feature of the region is represented by the eastern part of the Southalpine Chain, where the Southern Alps interact with the Dinarides mountain belt (Cuffaro *et al*. 2010). A description of the many seismogenic sources populating the area is provided by *Zonazione Sismogenetica ZS9* (Meletti *et al*. 2004), a seismic source model for Italy developed in the framework of compiling the national reference map for probabilistic seismic hazard assessment. The main sources are identified in the Trieste-Monte Nevoso (*ZS9* 904), Friuli-Veneto Orientale (*ZS9* 905) and Garda-Veronese (*ZS9* 906) seismogenic zones, defined based on a 3D structural model of Italy, on the spatial distribution of historic and modern earthquake for classes of magnitude, and on a kinematic model for the Mediterranean area. The area has significant seismic activity and suffered from destructive historical earthquakes, mostly located in the Veneto and Idrija regions. Figure 5-1 illustrates the historic seismicity in the area for the period 1976-2017 and the related main seismogenic sources. The majority of events has low to moderate magnitude, with some higher magnitude cases like the 1976 Friuli

(Italy) earthquake and the 1998 and 2004 Bovec-Krn (Slovenia) sequences. Most events have shallow hypocentral depths lower than 20 km: the depth distribution is clustered around 7.5 and 13.5 km in the western part of Friuli-Venezia Giulia region (Bressan *et al.* 1998), around 7.5 and 11 km in the central active area, and down to about 20 km in the eastern area (Bressan *et al*. 2016; Bressan *et al*. 2019). The Alpine area shows well defined compressive style, whereas by moving to the West more strike-slip patterns appear in Veneto and Austria, and in Slovenia to the East (Bressan *et al.* 2003).
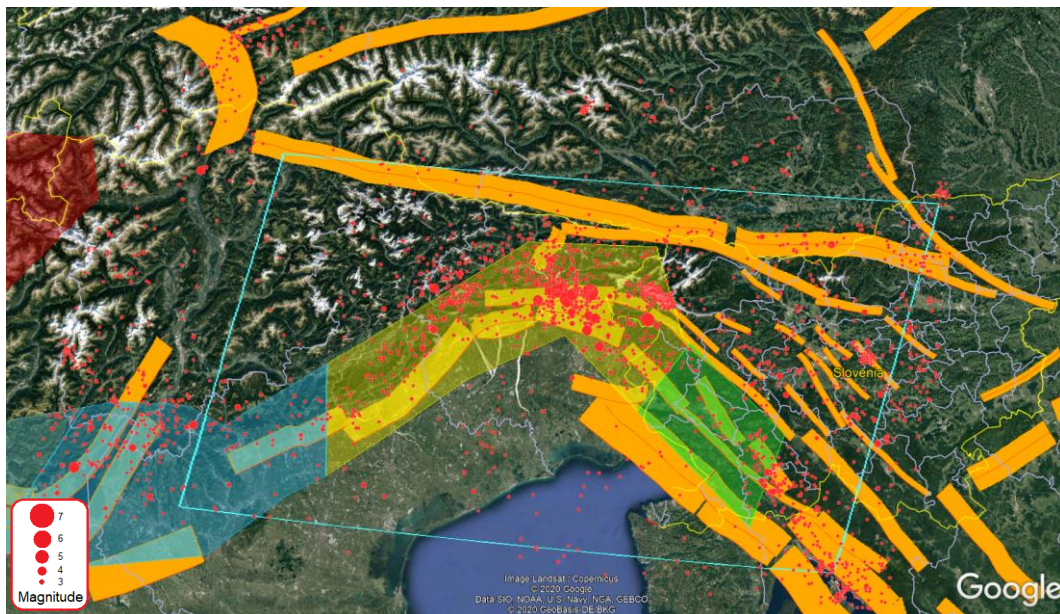
Figure 5-1: Seismicity of the study area in the period 1976-2017 (red circles). The main seismogenic sources populating the area are also pictured. Orange rectangular polygons are composite seismogenic sources as defined in the Individual Seismogenic Sources database (DISS; DISS Working Group 2018); other coloured polygons are *ZS9* seismogenic zones. The cyan polygon is the area chosen for the case study application (cf. Figure 5-3).
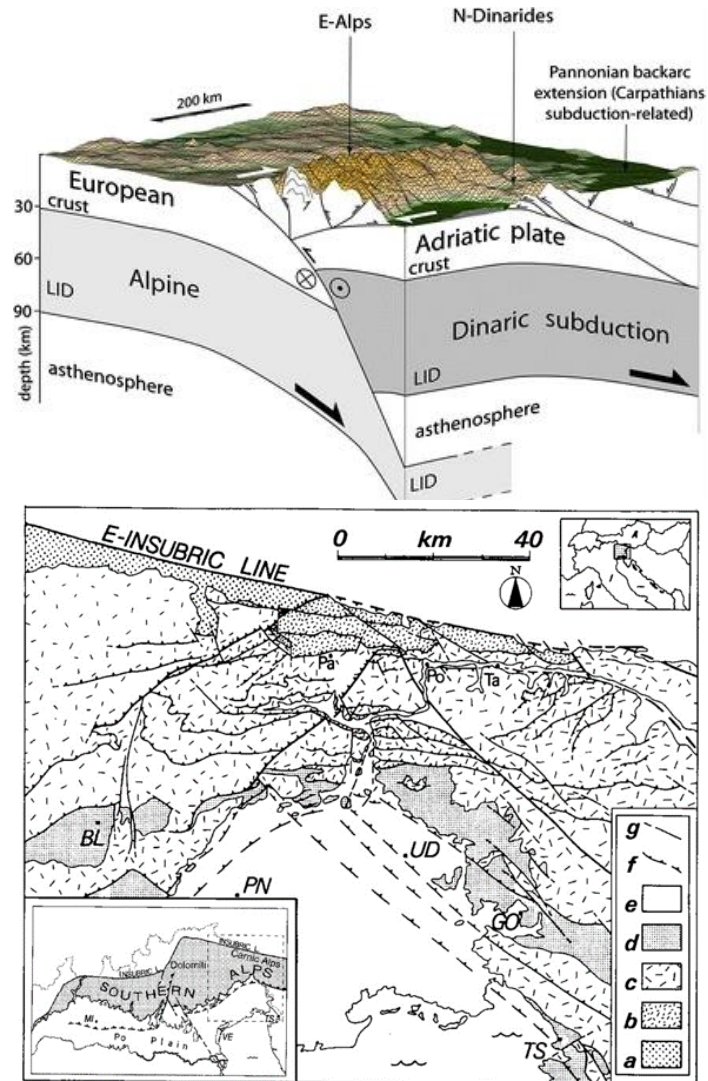
Figure 5-2: a) 3D reconstruction at the interference between the Alpine and Dinaric subductions; modified from Cuffaro *et al*. (2010). b) Schematic geologic map of the eastern Southern Alps (modified from Malagnini *et al*. 2002): a, Hercynian low metamorphic basement (Ordovician–Carboniferous); b, Paleocarnic nonmetamorphic Chain (Upper Ordovician-Carboniferous) and Upper Carboniferous- Lower Permian covers; c, Permo- Mesozoic mainly carbonatic successions; d, Flysch (Upper Mastrichtian-middle Eocene) and molassic sequence (Miocene); e, Quaternary alluvial deposits and moraines; f, thrust; g, subvertical fault; BL, Belluno; PN, Pordenone; UD, Udine; GO, Gorizia; TS, Trieste; Pa, Paluzza; Po, Pontebba; Ta, Tarvisio; MI, Milan; VE, Venice.

Figure 5-2 shows the tectonic and geologic features described in this paragraph. They reflect the collision between the Adria microplate and the European plate (Mantovani *et al*. 2006). The strongest seismicity is clustered in the Friulian Alps area, which is characterized by the greatest shortening. The orogeny of this

section of the Alpine chain interests a stratigraphic sequence about 14 km thick (Pondrelli *et al.* 2020), which contains almost all the background seismicity of the area. The Moho depth increases smoothly from South to North and reaches a depth of 40 km below the Southern Alps front (Bianchi *et al.* 2014).

After the destructive sequence of the 1976 Friuli earthquake, the number of seismic recording stations monitoring the area started to rise. Current seismic activity has been recorded since late Seventies by many networks (cf. Table 5-1), both national and regional, which are densely present in the Friuli-Venezia Giulia area, with coverage also towards Veneto, Slovenia, and Austria. Most stations belong to the Italian Accelerometric Network (code IT; Rete Accelerometrica Nazionale – RAN; Gorini *et al.* 2010; Costa *et al.* 2014; in light blue in Figure 5-3), managed by the Italian National Civil Protection. Other national networks are the Slovenian (code SL; Slovenian Environment Agency 2001), Austrian (code OE; Austrian Seismic Network; ZAMG 1987) and Croatian ones (code CR; Croatian Seismograph Network; University of Zagreb 2001). These national networks are supported by denser local networks such as NorthEast Italy BroadBand (code NI; OGS and University of Trieste 2002), OGS (code OX; North-East Italy Seismic Network; OGS 2016) and RAF (code RF; Friuli Venezia Giulia Accelerometric Network; Costa *et al.* 2010) networks. Stations belonging to the HAREIA Interreg project (code HA), and one station belonging to the Mediterranean network MedNet (code MN; MedNet Project Partner Institutions 1990) are also present. All stations are equipped with accelerometric and/or velocimetric sensors.

Table 5-1: List of seismic monitoring networks covering the study area, appearing also in Figure 5-3.

Network Code	Description	Label
CR	Croatian National Network	▲
HA	Hareia Project Network	▲
IT	Italian Accelerometric Network – RAN	▲
MN	MedNet	▲
NI	NorthEast Italy BroadBand Network	▲
OE	Austrian National Network	▲
OX	OGS Network (Italy)	△
RF	RAF Network (Italy)	▲
SL	Slovenian National Network – SNRS	▲

The case study area was defined as a simple polygon to encase all the main seismogenic features present in the region and to be covered by as many stations as possible. It is represented in Figure 5-3 as a black polygon which extends from the margin of the Veneto plain to the West to the Slovenian border to the East, and from the Venetian lagoon to the South to the Austrian borders on the Alps chain to the North. To increase the database population, a buffer area to be used as additional source for events only was defined around the selected case study area. This event buffer area forms an almost regular frame around the case study polygon (9.8 – 16.3 °E, 45.0 – 47.7 °N). Its extension was limited to the South to exclude events originated in the Po plain, for which a characteristic enhancement of ground motion has been observed at distances between 90 and 150 km as an effect of the reflection of S waves at the Moho (Bragato *et al.* 2011; Sugan and Vuan 2014). Figure 5-3 also displays the Italian seismogenic zones, according to *ZS9*, specifically zones 903 to 907. According to the zonation, they mainly host events with reverse or strike-slip focal mechanisms with focal depths in the range 5-12 km.
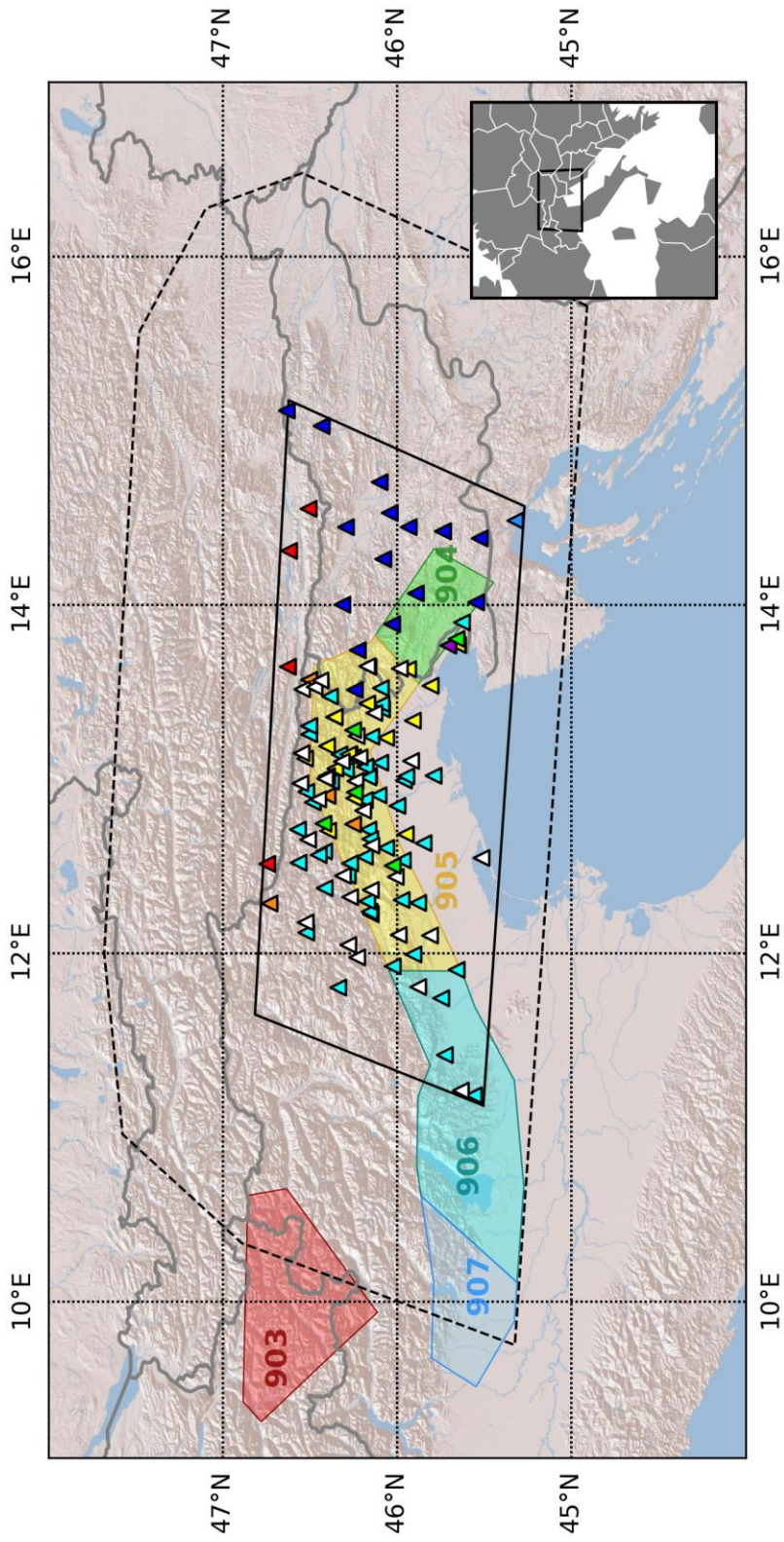
Figure 5-3: The case study area (solid black polygon) in Northeast Italy. The outer black polygon marks the corresponding buffer area. Filled polygons represent seismogenic zones as defined according to ZS9. Filled triangles mark the location of seismic stations in the area; corresponding colour codes are reported in Table 5-1.

5.2 Data

The area of interest is a trans-frontier region between three different Countries, so the use of national seismic bulletins could limit the coverage and precision of data related to earthquakes happened beyond national borders. The reviewed ISC web bulletin (International Seismological Centre 2020) was used as a reference to obtain parametric information on the events occurred inside the buffer area. It results from expert revision on localizations provided by different European agencies and covers seismic events up to 2019. For each event, the suggested prime localization was used, which includes information on latitude, longitude, depth, and origin time. Based on the information on the availability of recording stations in the area through the years, the search was limited to seismic events occurred after 2009, when the broadband station coverage started to increase significantly, together with the quality of the installed instruments. Most of the earthquakes happened in the last decade in the area have low to moderate seismicity, so a constraint was imposed that the information should include the local magnitude (M_L) value of the event, with $M_L \geq 2$. The preliminary event selection consisted of 392 earthquakes occurred in the study area after 2009, with $2 \leq M_L \leq 4.6$.

The initial selection of events retrieved from the ISC reviewed bulletin had to be matched with instrumental data recorded by stations inside the case study area. It should be stressed how, even if the coverage for the area is good, retrieving both datasets and metadata (namely instrumental corrections) from some stations may not be a trivial procedure due to the different ownerships. Other stations, many of which belonging to NI network, were only recently installed (from 2018 on) and did not have enough data associated to the selected events. In the light of such considerations, a conservative decision was made to use the waveform database available from SeisRaM group at the University of Trieste, which relies on the trans-frontier European network CE3RN (Bragato *et al*. 2014) and gathers data collected from IT, NI, RF, SL, CR, AT and MN networks starting from 2009. This choice currently limits the event time span to the period 2012-2017, for which the

revised database is available, but it matches the requirement that all waveforms have correct assigned information on seismometer calibration and response functions. This requirement ensures that response functions can be correctly deconvolved with the signal and noise spectra, so that the term $I_j(f)$ in Equation (1-6) can be set equal to one.

Some constraints were applied to the SeisRaM 3-component accelerometric dataset before use to improve the stability of the results. Following Edwards *et al.* (2008), only recordings with at least one associated pick (either P or S) were kept; if only one pick was recorded, the other arrival time was estimated using a P:S velocity ratio of 1.73. The composition of the database ($M_L \leq 4.6$ and hypocentral distances over 10 km) ensured that criteria for the use of point source spectral models were met (Brune 1970, 1971; Boatwright 1978). Stations with less than five records related to the event bulletin were discarded.

Data from the SeisRaM catalogue was pre-processed following the procedure described in Chapter 2 to determine the actual number of traces containing useful information related to the selected events. The only processing difference consists in the lack of Butterworth filtering. A 10-second noise window (prior to P-wave arrival) was extracted along with a signal window with length corresponding to 5%-95% of the energy integral (beginning at the P-wave onset). The process formerly used to select the filter cut-off frequencies, based on Signal-to-Noise Ratio (SNR), was used only to determine the range of frequencies with sufficient SNR level for each waveform. No actual filtering was applied to the waveforms. Only traces with SNR values greater than 2.8 over the bandwidth 2-10 Hz were kept. Since the database is limited to the period 2012-2017, all seismometers are high broadband seismometers with sampling frequency above 100 Hz, most of them at 200 Hz. A maximum frequency range of 0.5-25 Hz was imposed to focus the inversion on the range of interest for site amplification effects. The lower bound was selected based on the availability of the noise window, lasting 10 seconds only, as most waveforms came from triggered recordings. As a consequence, the lowermost frequency sampled by the noise window is around 0.1 Hz. This conservative choice ensured that at least a few cycles of each

sampled frequency were contributing to the noise spectra, without risk of underestimation. This preliminary waveform database counted 1299 3-component accelerometric waveforms related to 37 stations in the case study area.

After pre-processing, the waveform database was matched with information from the selected event catalogue by using a recursive filter. Events with less than five associated recordings were discarded, then only stations with at least five recorded events were kept, until both conditions were matched at the same time. After all selection criteria were applied, a total of 234 3-component records was available for use in the spectral inversion, recorded at 24 stations from 23 events. Table 5-2 and Table 5-3 list the events and the stations, respectively. The database is composed mainly by low and intermediate magnitude events, with corresponding moderate hypocentral distances. This reflects the sparseness of the station coverage and the characteristic seismicity for the area. Figure 5-4 presents the distribution of ray paths, stations and earthquake epicentres. Metadata for the dataset is summarized in Figure 5-5. Magnitude and hypocentral distance ranges are shown in Figure 5-5.a, with $2.3 \leq M_L \leq 4.5$ and hypocentral distances up to 204 km. All events are shallow crustal events with focal depths below 20 km (Figure 5-5.b). The distribution of events over the magnitude range is shown in Figure 5-5.c. The station v_{S30} values range from 312 to 1082 m/s, all inferred from topographic methods. Most stations are on *EC8-A* class soil, as defined in Eurocode-8 (EC8 2004), with some exceptions belonging to classes *EC8-B* and *EC8-C*.
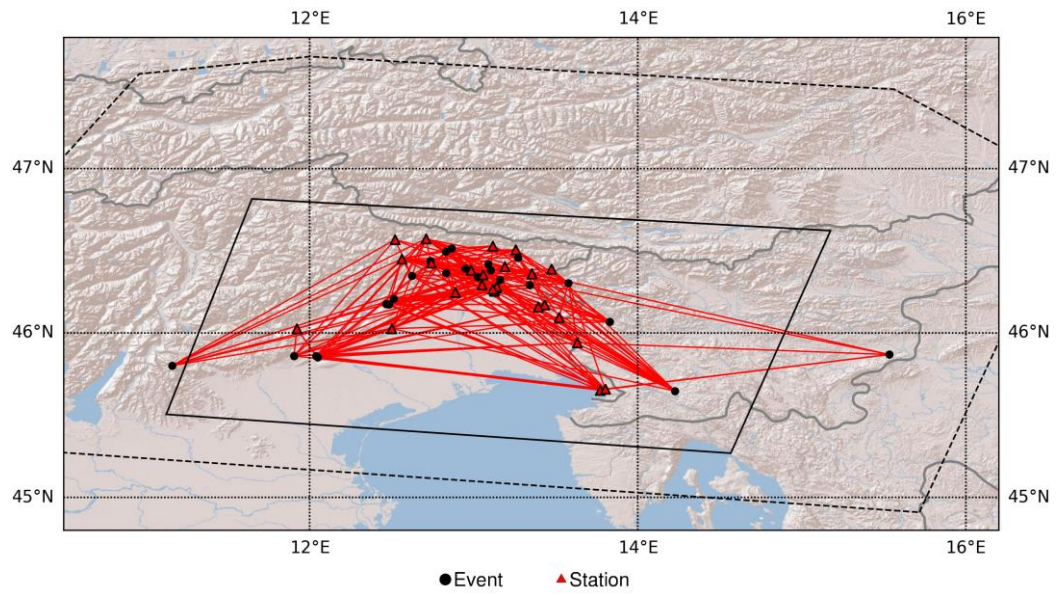
Figure 5-4: Map of the case study area with the used earthquakes (circles), stations (triangles), and rays (red lines).
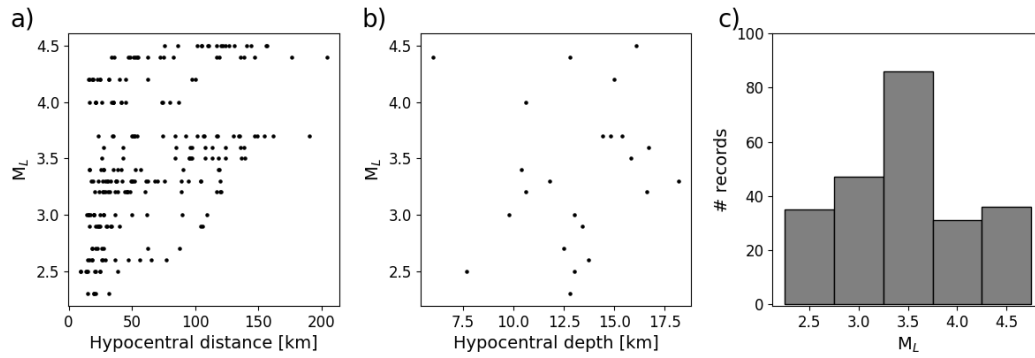
Figure 5-5: a) Magnitude and hypocentral distance ranges, b) hypocentral depths of the events, and c) distribution of events over the used magnitude range.

Table 5-2: Events in the selected database used for spectral inversion.

Event	Date	Time	Latitude (°N)	Longitude (°E)	Depth (km)	M_L	# Records
1	2012-06-09	02:04:57	46.1759	12.4661	12.8	4.40	14
2	2013-08-24	13:59:01	46.1743	12.4841	14.4	3.70	8
3	2013-09-06	15:01:35	46.3661	12.8309	9.8	3.00	6
4	2013-10-31	18:46:22	46.2083	12.5133	10.6	3.20	5
5	2014-04-22	08:58:28	45.6467	14.2267	16.1	4.50	17
6	2014-05-29	07:24:19	46.0679	13.8304	18.2	3.30	10
7	2015-01-30	00:45:50	46.3818	13.1038	15.0	4.20	15
8	2015-05-12	02:02:50	45.8601	12.0389	15.8	3.50	13
9	2015-05-15	05:35:46	45.8522	12.0489	16.7	3.60	10
10	2015-08-17	00:15:33	46.4579	13.2708	13.7	2.60	12
11	2015-08-18	20:10:02	45.8601	11.9049	14.8	3.70	9
12	2015-08-29	18:47:04	46.3029	13.5776	10.6	4.00	16
13	2015-11-01	07:52:33	45.8689	15.5325	6.0	4.40	5
14	2015-11-11	19:46:37	46.4951	12.8281	16.6	3.20	14
15	2015-11-11	21:20:31	46.5143	12.8651	12.5	2.70	8
16	2015-11-21	11:52:38	46.4392	12.7366	11.8	3.30	16
17	2015-12-08	15:05:01	46.3486	12.6223	10.4	3.40	13
18	2016-07-19	22:36:50	46.4190	13.0883	13.0	2.50	5
19	2016-08-10	02:38:05	46.3926	12.9490	13.4	2.90	13
20	2016-12-22	08:43:54	46.2919	13.3403	12.8	2.30	5
21	2017-02-09	08:14:08	45.8017	11.1616	15.4	3.70	7
22	2017-03-23	13:11:07	46.3245	13.1601	13.0	3.00	9
23	2017-03-24	17:47:07	46.3393	13.0263	7.7	2.50	5

Table 5-3: Stations in the selected database used for spectral inversion. All of them are high-quality broadband accelerometric stations.

Name	Network	Latitude (°N)	Longitude (°E)	Altitude (m)	Soil Classification
AUP	RF	46.5064	13.2563	905	Rock (*EC8-A*)
AVS	RF	46.2946	13.0497	256	Rock (*EC8-A*)
CARC	RF	45.6526	13.77	0	Soil (*EC8-C*)
CESC	RF	46.3557	13.0572	355	Soil (*EC8-B*)
CHF	RF	46.3887	13.4737	1194	Rock (*EC8-A*)
CMO	RF	46.0941	13.5207	680	Rock (*EC8-A*)
DANT	RF	46.5675	12.5199	1453	Rock (*EC8-A*)
DST2	NI	45.6589	13.8013	80	Rock (*EC8-A*)
FDS	RF	46.4513	12.5627	1795	Rock (*EC8-A*)
FLP	IT	46.0266	11.9225	294	Soil (*EC8-C*)
GEDE	RF	46.2544	13.1247	232	Soil (*EC8-B*)
GEPF	RF	46.2752	13.1385	255	Rock (*EC8-A*)
GESC	RF	46.2819	13.1409	373	Soil (*EC8-C*)
GORI	RF	45.9403	13.6308	141	Soil (*EC8-B*)
MASA	RF	46.1722	13.4313	693	Rock (*EC8-A*)
MOGG	RF	46.4048	13.1885	371	Rock *(EC8-A)*
PAUL	RF	46.5302	13.1162	640	Rock (*EC8-A*)
POLC	NI	46.0266	12.5005	150	Soil (*EC8-B*)
PRAD	RF	46.2476	12.8873	586	Rock (*EC8-A*)
PURA	NI	46.4258	12.7419	1420	Rock (*EC8-A*)
RST	RF	46.3629	13.3537	673	Rock (*EC8-A*)
SPP	IT	46.5725	12.709	1318	Soil (*EC8-C*)
STOL	RF	46.3605	13.3546	586	Soil (*EC8-C*)
TLM2	IT	46.3814	12.9839	519	Soil (*EC8-B*)

Velocity FAS were extracted after integrating and detrending each accelerometric record to obtain velocimetric waveforms. The maximum between horizontal components was selected, and a standard Konno Ohmachi smoothing (Konno and Ohmachi 1998), which is symmetric in log-space, was applied to reduce the level of high-frequency fluctuations in the spectra. The actual spectra used for inversion were then obtained by interpolation of the smoothed spectra over a set of 30 logarithmically spaced frequency values between 0.5 and 25 Hz. A Boolean array

was associated to each spectrum, to mark which frequency points had acceptable SNR values and thus should be used for inversion and which had to be discarded. As a final step, each spectrum was corrected by subtracting its associated pre-event noise spectrum from it. A schematic representation of this procedure is presented in Figure 5-6.
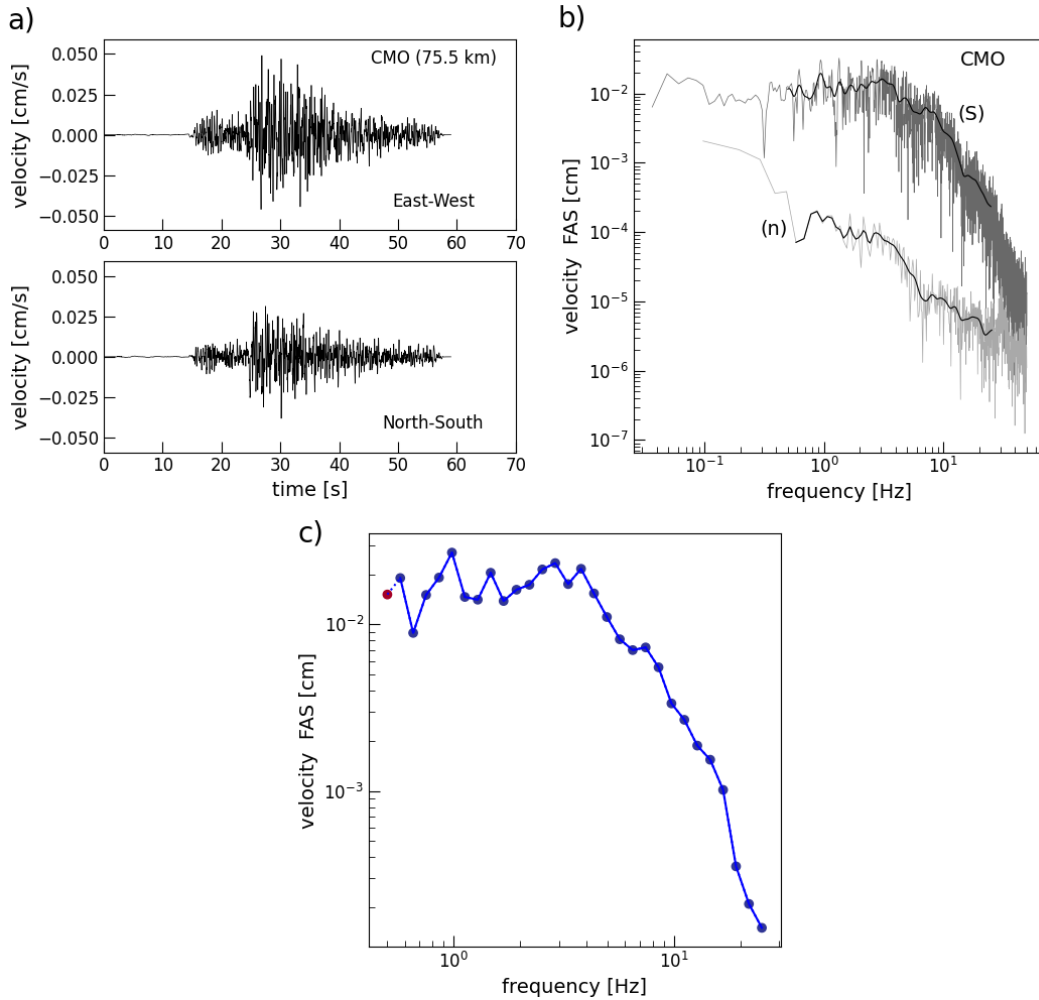
Figure 5-6: a) records of the velocity for the East-West and North-South components of the 22 April 2014 earthquake (# 5 in Table 5-2) at station CMO. b) Raw (grey lines) and smoothed (black lines) FAS of the velocity for signal (S) and noise (n) for the maximum horizontal component, in this case the East-West one. c) The final corresponding spectrum used for inversion (blue line), with 30 frequency points log-normally distributed in the range 0.5 – 25 Hz. Each point has an assigned Boolean weight, equal to 1 if the point has sufficient SNR score (blue circles) and equal to 0 otherwise (red circles).

5.3 Inversion setup

Parametric inversion was performed on a set of 234 velocity Fourier spectra, corresponding to $N_i = 23$ events recorded by $N_j = 24$ stations. Amplitudes were estimated on a set of $N_k = 30$ log-distributed frequency points in the range $0.5 - 25.0$ Hz. The maximum between East–West and North–South components is selected in the Fourier domain.

The parametric model used for inversion is the one described by Equation (4-12), where the used apparent geometrical spreading term is the one defined by Malagnini *et al.* (2002), obtained from a similar dataset in terms of used events and for the same tectonic and geological setting. The corresponding sets of exponential decay factors $\boldsymbol{\lambda} = [\lambda_1, \dots, \lambda_n]$ and distances $\tilde{\boldsymbol{r}} = [\tilde{r}_1, \dots, \tilde{r}_n]$, as defined in Equation (4-9), were thus kept as constants throughout the inversion. The source term is in the form of a simple far field Brune spectrum. The quality factor and the site-related attenuation operator are treated as frequency-independent. The frequency-dependent site response is left unmodelled. It should be stressed that the validity of this parametric form relies on the assumption that the data is correctly deconvolved with the seismograph response functions before inversion ($\log I(f) = 0$). For this reason, the preliminary choice of using only data in the SeisRaM database was made to ensure that instrument effects were correctly subtracted, even at the cost of limiting the temporal span of available data. The choice of fixing a priori some of the parameters ($\boldsymbol{\lambda}$, $\tilde{\boldsymbol{r}}$) in Equation (4-12) using trusted reference from literature was also aimed at limiting the trade-off between parameters and better constraining the inversion.

As a consequence of the aforementioned modelling choices, the parameter vector used in the inversion algorithm is:

$$
\boldsymbol{p} = [\; [\log M_{01}, \dots, \log M_{0N_i}], [f_{c1}, \dots, f_{cN_i}], Q_0, [\log A_1, \dots, \log A_{N_j}],
$$
$$
[\kappa_1, \dots, \kappa_{N_j}], [\varepsilon_1^{SO}, \dots, \varepsilon_{N_i}^{SO}], \varepsilon^P, [\varepsilon_1^{SI}, \dots, \varepsilon_{N_j}^{SI}] \;] \, .
\tag{5-1}
$$

It is composed of N_i terms for each of source spectra, corner frequency and source uncertainty vectors, one term each for the attenuation factor and the propagation uncertainty term, and N_j terms for each of frequency-independent site amplification, site-related attenuation and site uncertainty vectors.

A single-step inversion based on the SLSQP methodology was performed, using the loss function described by Equation (4-18) with no weighting applied, as discussed in §4.1.3. Individual bounds were applied to each parameter to limit the search space to a physically meaningful region and avoid fictitious local minima. A constraint was imposed on the constant amplitude correction terms A_j to define them with respect to a reference condition. Ideally, the optimal reference would be a rock site for which no amplification is experienced, at any frequency. Alternatively, a site with detailed known site response could be used (e.g., Parolai *et al.* 2004), or a virtual mean reference site could be defined so that amplification refers to the network average (Edwards *et al.* 2008), that is:

$$\sum_{j=1}^{N_j} \log(A_j) = 0 \,. \qquad (5\text{-}2)$$

Even if in principle any well-characterized reference site could be used, most engineering and seismic applications of site response curves use rock soil condition as a reference. The choice for a reference site was thus narrowed to stations belonging to class A according to EC8 classification. Station MOGG (Moggio Udinese, cf. Table 5-3) was initially chosen as a possible candidate, as it is installed directly on the outcropping Dolomite bedrock. In order to keep the whole inversion process more flexible and easily exportable to other datasets, the inversion was constrained by assuming:

$$\sum_{EC8-A} \log(A_j) = 0 \,, \qquad (5\text{-}3)$$

so that parameter A_j is defined as the amplification relative to the network average rock (*EC8-A*) site.

Initial guesses for source parameters were obtained from the reference ISC bulletin parametric information. Seismic moment starting guesses (M_0^{IN}) were built from bulletin values of M_L by combining the scaling law by Hanks and Kanamori (1979):

$$\log_{10}(M_0^{IN}) = 1.5\,M_w + 9.05\,,$$
(5-4)

where M_w is the moment magnitude and M_0 is expressed in Nm, with the empirical relationship developed by Munafò *et al.* (2016) to correlate M_w and M_L for Italian earthquakes with small magnitudes (up to $M_w \sim 4$):

$$M_w = 0.67\,M_L + 1.15\,.$$
(5-5)

Notice that the used linear approximation of the relationship between M_w and M_L is only valid within limited magnitude range. Even so, preliminary tests proved that it is the most reasonable for use with the analysed dataset, as it only serves as a guide to define the parameters search space. Other more general relationships, such as the quadratic one proposed by Grünthal *et al.* (2009) for Europe, would actually lead to a systematic increase in the inversion residual and thus compromise the quality of the extracted amplification curves (cf. §5.4). Brune's source model (Brune 1970, 1971) was used to calculate the initial guess for corner frequency values (f_c^{IN}) as a function of seismic moment and the stress drop parameter $\Delta\sigma$:

$$f_c^{IN} = 0.4906\,v_s \left(\frac{\overline{\Delta\sigma}}{M_0^{IN}}\right)^{1/3},$$
(5-6)

where $\overline{\Delta\sigma} = 0.73\text{MPa}$ is the average regional stress drop taken from literature (Franceschina *et al.* 2006). The corresponding upper and lower bounds for source parameters were obtained by propagating the magnitude uncertainty, conservatively set to 0.5. Bounds for the corner frequency values also took into consideration the estimated maximum span of stress drop values for the region ($0.1 - 5$ MPa; from Franceschina *et al.* 2006).

Starting values for propagation and site parameters were taken from region-specific literature. Different hypotheses were considered for the Q model from independent studies investigating seismic wave attenuation for the area of interest. Console and Rovelli (1981) estimated the quality factor of the Friuli-Venezia Giulia region using the strong motion accelerograms of the 1976 earthquake sequence (main shock $M_L = 6.4$), for epicentral distances up to about 200 km. They compared power spectra of entire accelerograms at different distances and obtained the relation $Q(f) = 80f^{1.1}$, valid for the range $0.1 - 10$ Hz. Malagnini *et al.* (2002) obtained $Q(f) = 260f^{0.55}$ based on spectral inversion for an area extending to Slovenia to Friuli-Venezia Giulia, together with an average spectral decay parameter $\bar{\kappa} = 0.045$ s. Given the specific (frequency-independent) modelling choice, the actual parameter space provided to the algorithm was wide enough to cover different possibilities for the Q_0 value, from low ($Q_0 \sim 50$), to intermediate ($Q_0 \sim 500$), to high ($Q_0 \sim 1500$), with a starting value of $Q_0^{IN} = 260$. The average regional site-related attenuation value provided by Gentili *et al.* (2011) was used to set $\kappa_0^{IN} = 0.047$ s for all stations, with boundaries allowing it to span in the range $0.01 - 1$ s to cover the many possible values reported in literature. As for the frequency-independent site amplification A^{IN} and the uncertainty terms, they were initially set to zero and constrained as per Equation (5-3) so that the network average rock site would be used as a reference. No bound other than being positively defined was imposed on the values for the uncertainty terms.

With this setting, the algorithm explores a broad model space, but still keeps the starting values not too far from the expected real values so that the linearized inversion remains valid. Different starting models inside this model space were tested to check that the results of the inversion were not dependent on the starting model. Figure 5-7 depicts the plots of observed and fitted velocity FAS for a selection of records to demonstrate the representative performance of the fitted model. Modelled spectra are represented as red lines, with an associated total uncertainty band (red shaded area) obtained by combining the inverted uncertainty terms.
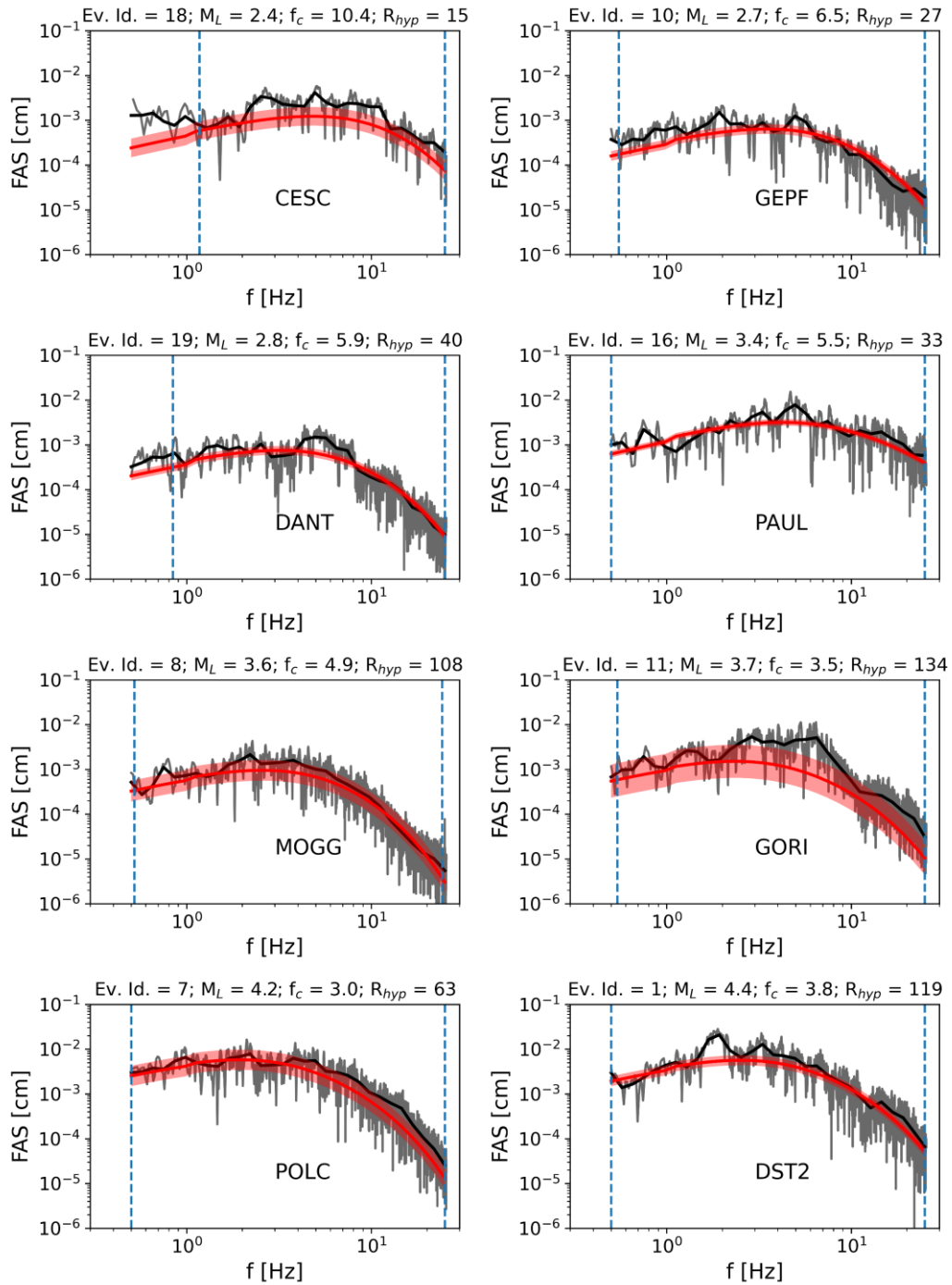
Figure 5-7: Plots depicting the fit between the observed velocity FAS (grey lines) and modelled FAS (thick red lines), with the lower and upper limits of the usable frequencies (blue vertical dashed lines). The associated total uncertainty band, obtained by combining the inverted uncertainty terms, is shown as red shaded area. Inversion was performed on the smoothed version of the spectra (black lines), using spectral amplitudes inside the usable frequency range.

5.4 Source parameters

Table 5-4 lists the source parameters of the 23 events analysed in this study. The dataset is composed of shallow earthquakes with moderate magnitudes. Direct comparison of M_L magnitudes is not feasible, so M_W values are compared instead. Observed M_L is converted to a "database" M_W value by using Equation (5-5).

Table 5-4: Source parameters of the analysed events. Reference values are taken from the reviewed ISC bulletin (International Seismological Centre 2020) for M_W and from the webservice of Istituto Nazionale di Geofisica e Vulcanologia (INGV; http://terremoti.ingv.it/) for M_0, along with the source classificaction (SS = strike-slip faulting, TF = thrust faulting, NF = normal faulting).

Event	Focal Mech.	M_L	M_W	Inverted M_W	M_0 [Nm]	Inverted M_0 [Nm]	Inverted f_c [Hz]	Inverted $\Delta\sigma$ [MPa]
1	TF	4.40	4.08	4.08	1.03E+15	1.47E+15	3.84	16.46
2	TF	3.70	3.62	3.61	1.05E+14	2.92E+14	5.82	11.37
3	n.d.	3.00	3.15	3.12	n.d.	5.42E+13	6.54	3.00
4	n.d.	3.20	3.28	3.19	n.d.	6.75E+13	7.63	5.92
5	SS	4.50	4.15	4.24	1.08E+16	2.60E+15	3.28	18.16
6	SS	3.30	3.35	3.49	1.66E+14	1.90E+14	4.20	2.77
7	TF	4.20	3.95	3.96	5.6E+14	9.84E+14	2.95	5.01
8	n.d.	3.50	3.48	3.54	n.d.	2.26E+14	4.91	5.27
9	TF	3.60	3.55	3.52	1.24E+14	2.13E+14	5.01	5.31
10	n.d.	2.60	2.88	2.93	n.d.	2.82E+13	6.48	1.51
11	NF	3.70	3.62	3.63	1.32E+14	3.10E+14	3.53	2.69
12	SS	4.00	3.82	3.91	1.14E+15	8.35E+14	2.24	1.86
13	TF	4.40	4.08	4.19	5.91E+15	2.16E+15	1.79	2.45
14	n.d.	3.20	3.28	3.27	n.d.	8.97E+13	5.13	2.39
15	n.d.	2.70	2.95	2.91	n.d.	2.56E+13	7.79	2.39
16	TF	3.30	3.35	3.44	1.54E+14	1.63E+14	5.54	5.45
17	n.d.	3.40	3.42	3.36	n.d.	1.25E+14	5.50	4.11
18	n.d.	2.50	2.82	2.77	n.d.	1.62E+13	10.39	3.58
19	n.d.	2.90	3.08	3.05	n.d.	4.21E+13	5.87	1.68
20	n.d.	2.30	2.68	2.72	n.d.	1.33E+13	9.00	1.92
21	n.d.	3.70	3.62	3.57	n.d.	2.51E+14	6.67	14.74
22	n.d.	3.00	3.15	3.00	n.d.	3.56E+13	12.98	15.41
23	n.d.	2.50	2.82	2.76	n.d.	1.55E+13	9.76	2.85

Inverted seismic moments range between 1.33×10^{13} and 2.60×10^{15} Nm and corner frequencies between 1.79 and 12.98 Hz. When a comparison with seismic moment estimates from literature is available, seismic moment values seem reasonable. For example, for the 2012 Pordenone event (#1 in Table 5-2), $M_0 = 1.47 \times 10^{15}$ Nm was obtained, in good agreement with the solution provided by Istituto Nazionale di Geofisica e Vulcanologia (INGV) of $M_0 = 1.03 \times 10^{15}$ Nm.
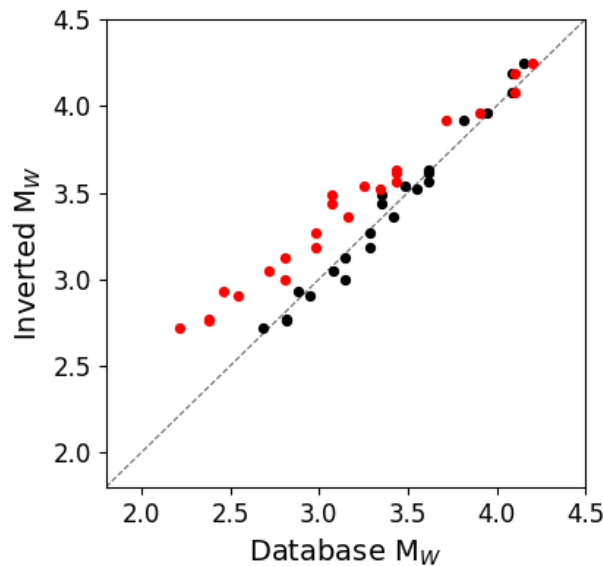
Figure 5-8: Comparison of inverted M_W with values obtained from database using Equation (5-5) (black circles). M_W values obtained from database using the quadratic relationship from Grünthal *et al*. (2009) are reported as a comparison (red circles).

Moment magnitude estimates are calculated from inverted seismic moments using the relationship described in Equation (5-4). Figure 5-8 shows the comparison between inverted moment magnitudes and database values calculated from the reviewed ISC bulletin data (black circles). M_W values are well reconstructed for the whole magnitude range. As a comparison, M_W values calculated using the Grünthal *et al.* (2009) relationship are reported as red circles in Figure 5-8. As briefly mentioned in the Inversion Setup description, the use of a quadratic relationship would have led to a systematic underestimation of the database M_W values, with a consequent systematic increase in the residual values (cf. Equation (4-19)).
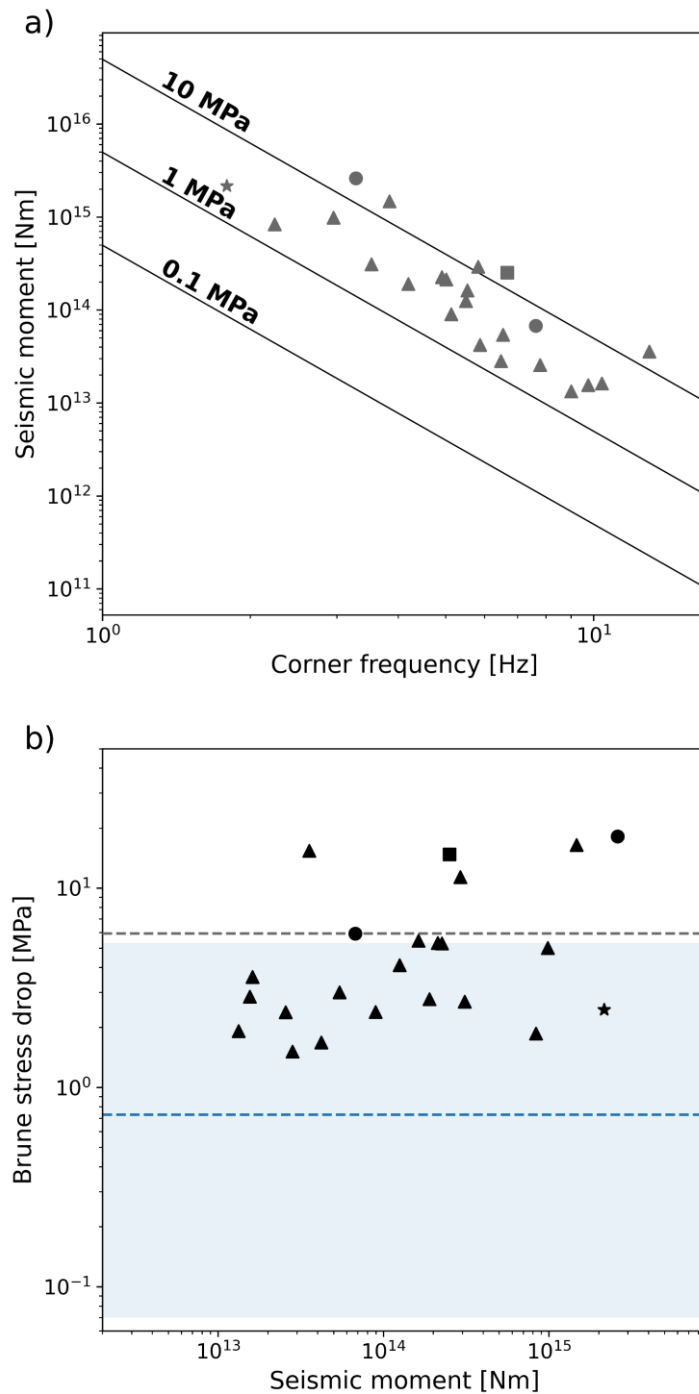
Figure 5-9: a) Inverted seismic moment versus corner frequency (grey markers); lines of constant Brune stress drop, computed with Equation (5-5), are also shown (black lines). b) Inverted Brune stress drop versus seismic moment (black markers) and related average value (dashed grey line). The range of values obtained by Franceschina *et al.* (2006) for the Friuli-Venezia Giulia region is shown for comparison (light blue area), together with its average value (dashed blue line). Circles, triangles and squares represent events occurred in *ZS904*, *ZS905* and *ZS906,* respectively. The star marks event #13, occurred outside of *ZS9* zonation.

Seismic moments are plotted against corner frequencies in Figure 5-9.a, together with lines of constant stress drop. Equivalent Brune stress drops $\Delta\sigma$ can be calculated from (Brune 1970):

$$\Delta\sigma = \frac{7}{16} M_0 \left(\frac{f_c}{0.37 v_s}\right)^3 . \qquad (5\text{-}7)$$

The stress drop values inferred from inverted parameters are scattered, but mostly lie between 1.51 and 5.92 MPa, as shown in Figure 5-9.b. Noticeably, higher values ($\Delta\sigma \geq 10$ MPa) are found for events occurred in seismic zones *ZS904* and *ZS906* (#5 and #21 in Table 5-2, respectively). In fact, if only events occurred inside *ZS905* are considered, most of the stress drop values fall in the range $0.1 - 5$ MPa obtained by Franceschina *et al.* (2006) for the Friuli-Venezia Giulia region (shaded blue area in Figure 5-9.b), which is roughly coincident with that seismogenic zone. This spatial pattern for stress drop values is visually rendered in Figure 5-10.
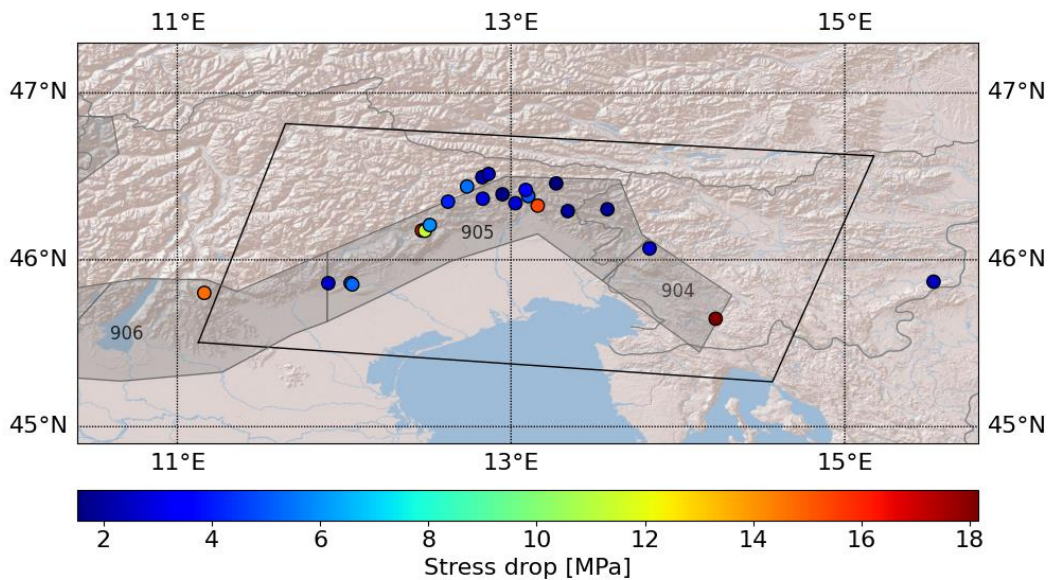
Figure 5-10: Spatial distribution of Brune stress drop estimated from inverted source parameters. *ZS9* seismogenic zones are plotted for reference (grey polygons; cf. Figure 5-1). Most values are below 10 MPa for events occurred inside zone *ZS905*.

The results obtained for source parameters prove the robustness of the inversion algorithm, which is capable to retrieve realistic information on the events used in the dataset. It should be noticed that more refined inversion techniques should be preferred if the focus were on the source parameters themselves, as corner frequencies of small events can be biased by erroneous correction of attenuation effects, bandwidth limitation of the recorded spectra and possible departure from the assumed model (Di Bona and Rovelli 1988).

5.5 Site parameters

Station-specific amplification effects can be represented in the form of site transfer functions. They result from the combination of elastic amplification and anelastic attenuation contributions in the frequency domain:

$$S_j(f_k) = A_j a_j(f_k) \, e^{-\pi f_k \kappa_{0j}} = ampl_j(f_k) \, e^{-\pi f_k \kappa_{0j}} \; . \qquad (5\text{-}8)$$

where $ampl_j(f_k)$ is the total elastic amplification composed by the constant amplitude correction resulting from the inversion (A) and the frequency-dependent response function obtained through residual analysis ($a(f)$).

A represents the mean amplification at the site over all frequencies with respect to the reference condition, which could be due to focusing of energy along wave guides at regional level (Edwards *et al.* 2008). The constraint for inversion was chosen so that amplification is referenced to the average network sites located on *EC8-A* class soil, which should correspond to rock condition (cf. Equation (5-3)). No distinct distribution of mean site corrections and geology emerges on a regional scale, based on the map view of frequency-independent amplification terms (Figure 5-11). Coherently with the chosen reference, sites located on softer soils (classes *EC8-B* and *EC8-C*; diamond markers in Figure 5-11) present higher mean site correction values with respect to *EC8-A* class sites (triangle markers in Figure 5-11), corresponding to rock conditions. The average value of constant amplitude corrections for soil sites is $A_{soil} = 1.33$.
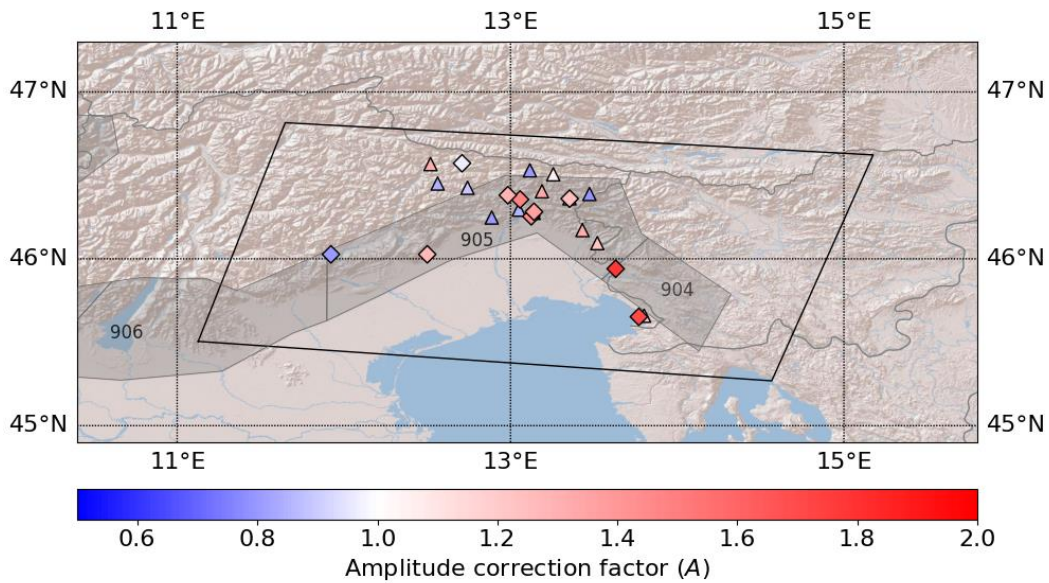
Figure 5-11: Spatial distribution of mean amplitude correction factors (*A*) obtained from inversion. Triangles mark stations catalogued as rock sites (*EC8-A*), diamonds mark stations on soil (*EC8-B* and *EC8-C*). The average values for the two categories are $A_{rock} = 1.0$ and $A_{soil} = 1.33$, respectively.

Frequency-dependent response functions were constructed from the residuals obtained from the inversion (Edwards *et al*. 2008; Drouet *et al*. 2010; Edwards and Fäh 2013) and may therefore suffer from the biases introduced into the overall residuals. The residuals are calculated as data/model, where the model is obtained as the exponential form of Equation (4-12) minus the uncertainty terms. A visual representation of the residual distribution (Figure 5-12) shows no evident dependency of the residual values on hypocentral distance, depth, or magnitude. The models seem to slightly underestimate the lowest and highest frequencies, whereas in the central frequency range $(1-20\,\text{Hz})$ the residuals are correctly centred on one.
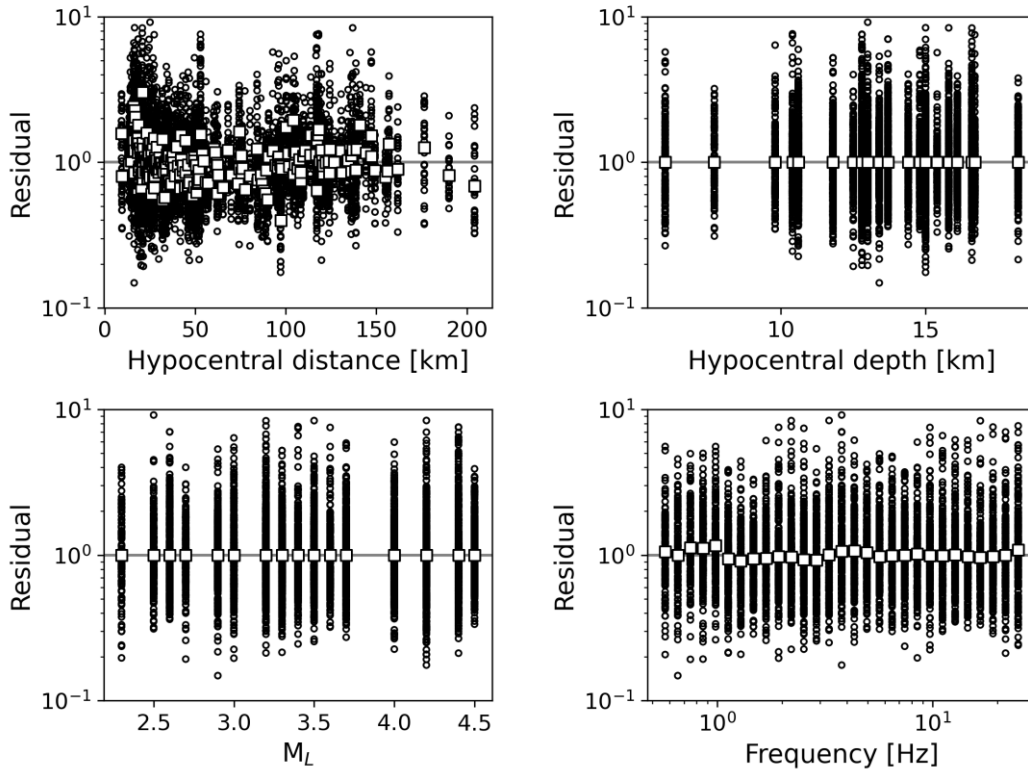
Figure 5-12: Overall residuals (data/model) plotted against hypocentral distance, source depth, local magnitude, and frequency. White squares indicate the log-mean residual at each hypocentral distance, source depth, local magnitude, and frequency value.

For each site, the amplification was computed frequency-wise following the methodology described in §4.1.4, that is, as the geometric mean of all factorial residuals at a given frequency over all spectra recorded at that station. This approach has been widely used in literature and has been proven to provide amplification values consistent with predicted 1D behaviour at rock sites and more complex 2D or 3D behaviour at soil sites (Edwards *et al.* 2013; Michel *et al.* 2014). For each station, the mean amplification value at each frequency ($a(f_k)$) has an associated geometric standard deviation, calculated as:

$$\sigma_j(f_k) = \exp\left(\sqrt{\frac{\sum_{i=1}^{N_i}\left(ln\frac{\vartheta_{ij}(f_k)}{a_j(f_k)}\right)^2}{N_i}}\right). \qquad (5\text{-}9)$$

The initial choice of using only stations with at least five associated recordings mitigates the aleatory variability captured by the estimation. This minimum requirement was ensured to also hold for each used frequency. Since spectra were calculated on a different number of frequency values, with upper and lower bounds dictated by signal-to-noise ratio, site amplification was calculated only for frequency points with at least five associate FAS values. All amplifications are relative to the average regional *EC8-A* site, defined as the generic rock reference. Figure 5-13 reports site amplification curves for a selection of stations (PAUL, class *EC8-A*; GEDE, class *EC8-B*; CARC and STOL, class *EC8-C*). It should be kept in mind that the robustness of site amplification curves as estimate of the true site amplification due to upper soil layering beneath the station still depends on the number of available recordings at each site. In fact, amplification curves also include contributions due to different effects, such as unmodelled phenomena (e.g., residual path effects), as well as inherent variability and noise.
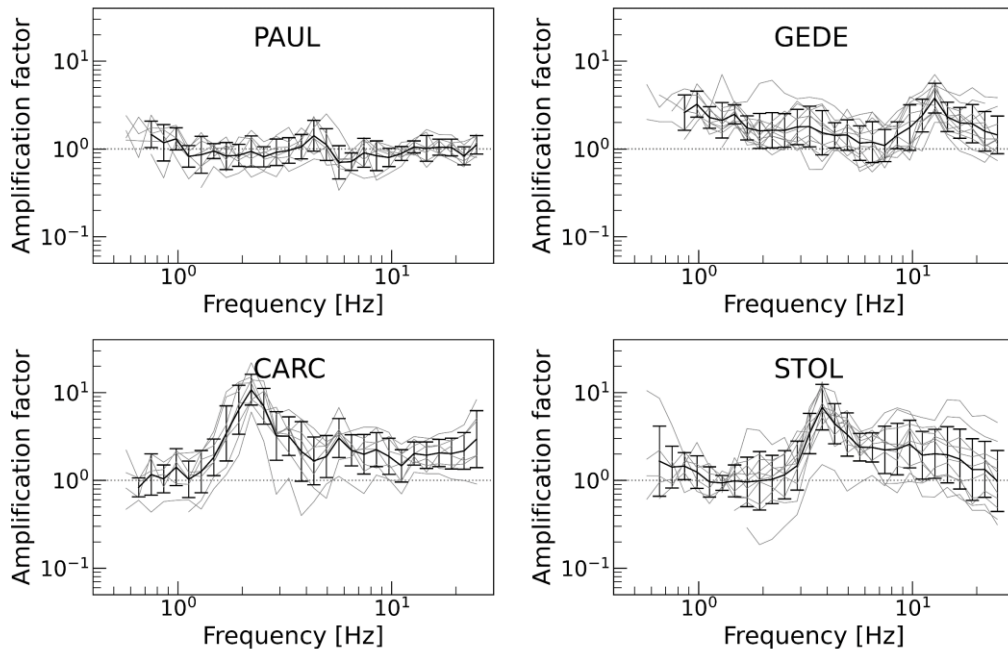
Figure 5-13: Sample plots of frequency-dependent site response for stations PAUL, GEDE, CARC and STOL. Grey lines indicate response functions from individual recordings, black lines represent the geometric mean site response function with error bars of one standard deviation.

The selection of a different constraint condition to be used as regional reference is investigated based on the site response function independent of attenuation ($ampl(f)$ in Equation (5-8)). From these, complete site transfer functions can be reconstructed for each station by multiplying them for the correspondent anelastic attenuation factor, which is a function of the inverted κ_0 value. Resulting curves for $ampl(f)$, including both the constant amplitude correction and the frequency-dependent curves, are plotted in Figure 5-14 for all available stations together with associated error bars of one standard deviation. Stations showing site responses affected by large site effects, described as large deviation above and below the average, are excluded. Remaining sites with flat response functions are indicated by grey boxes around the station names in Figure 5-14. Almost all of these stations belong to class *EC8-A*, which largely overlaps with rock site conditions, except for stations SPP and GESC. Station SPP is categorized as site class *EC8-C* according to geology and as class *EC8-A* according to topography. It shows no significant amplification, in agreement with noise H/V measures used to characterize the station (ITACA, D'Amico *et al.* 2020). In principle, this could lead to believe that it could be used in the reference. The same condition holds also for station GESC (site class *EC8-C*), and yet it has been known for showing strong amplification during the Bovec 1998 and 2004 earthquakes (Costa *et al.* 2006). As a precautionary measure, both stations are excluded as they are not located on rock according to geological maps and therefore high amplifications cannot be excluded, even if from a limited number of events. On the other hand, it is known that near-surface weathering and cracking can induce significant amplification also in stations identified as rock sites (Steidl *et al.* 1996), and that *EC8-A* classification does not always fully correspond to rock site condition. *EC8-A* stations for which no precise geological information is available are thus excluded fom reference as a conservative measure. In light of these results, it is possible to identify stations with a flat response close to 1 and located with no ambiguity on rock according to geological maps (AUP, AVS, DST2, FDS, GEPF, MASA, MOGG, PRAD, PURA, and RST). The suggested usage for these stations is to serve as a tool in imposing reference site conditions for the region, to be used for example for the classical spectral ratio method.
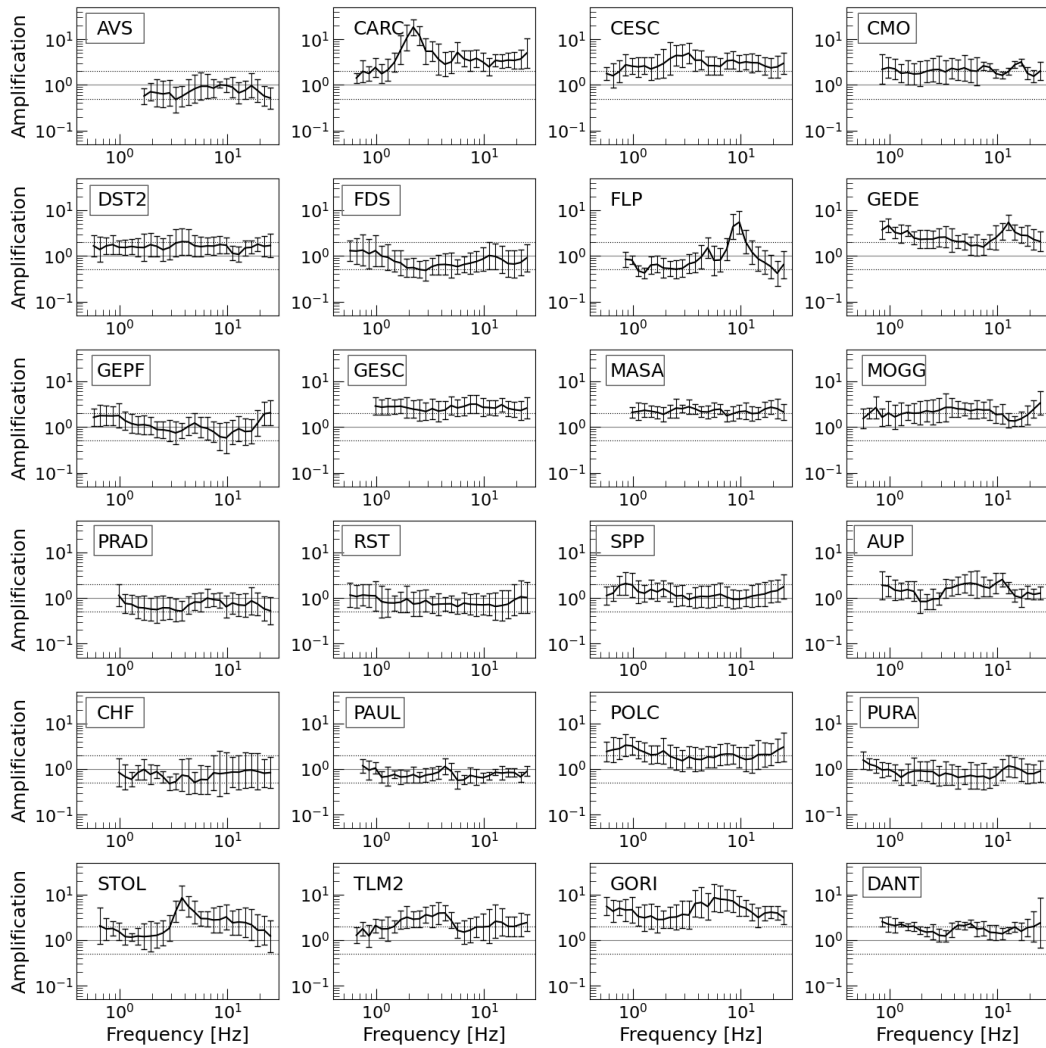
Figure 5-14: Elastic site response functions for the stations used in this study (cf. Table 5-3). Error bars are 1σ confidence levels; dashed horizontal lines mark the range of amplification values for which the function is considered as flat (amplification between 0.5 and 2). The stations with boxes around their names are those initially considered for building a new regional reference condition.

Site effects derived from the inversion were compared with those obtained with other methods, where available. The frequency values correspondent to the main amplification peak for stations FLP and STOL, which are located on alluvium and fluvial sediments, are in good agreement with values reported in ITACA (9 Hz versus 8.5 Hz for FLP; 3.9 Hz versus 3.7 Hz for STOL). Station CARC in particular has been accurately characterized by Fitzko *et al.* (2007). It is located in the old city centre of Trieste (NE Italy), in the basement of a 3-story historical building. The site is on a former salt pan, with 27-m thick sedimentary cover

87

composed by almost plane layers of clay and silts. Fitzko *et al.* (2007) compared multiple methodologies, namely noise horizontal-to-vertical (H/V) (Nakamura 1989), earthquake H/V (Lermo and Chavez-Garcia 1993) and spectral ratios with respect to a reference station. They used Standard Spectral Ratio to retrieve the transfer function of CARC related to the nearby bedrock reference site TRI (MN network) for frequencies up to 8 Hz. Fourier spectra for the full signals were averaged and smoothed by computing the mean amplitude for a 0.5 Hz frequency moving window, with 0.1 Hz overlap. Earthquake H/V response spectra (HVSR) were calculated by Fitzko *et al.* (2007) for a selection of five regional events, including the 2004 Bovec earthquake. The receiver function method is useful to identify the fundamental frequency of a site and to evaluate the amplification factor at low frequencies (e.g., Lermo and Chavez-Garcia 1993). They also applied Nakamura technique on microtremors (HVNSR), which is a reliable tool to find the resonant frequency of a site but has limitations in estimating the amplification magnitude (Lermo and Chavez-Garcia 1993). For the CARC station, results obtained by Fitzko *et al.* (2007) from SSR, HVSR and HVNSR method are reported in Figure 5-15 together with the site response functions determined from spectral inversion in this study. There is a good agreement of all results at low frequencies, with a strong amplification around 2 Hz due to the sedimentary structure, followed by a minor peak around 3 Hz and by another peak around 5.5 Hz. In particular, the frequency-dependent amplification alone ($a(f)$) shows amplification values in agreement with HVSR curves, which are obtained without imposing a reference condition, whereas total elastic amplification curves provide values compatible with SSR curves, which in turn relate the amplification to the bedrock reference.
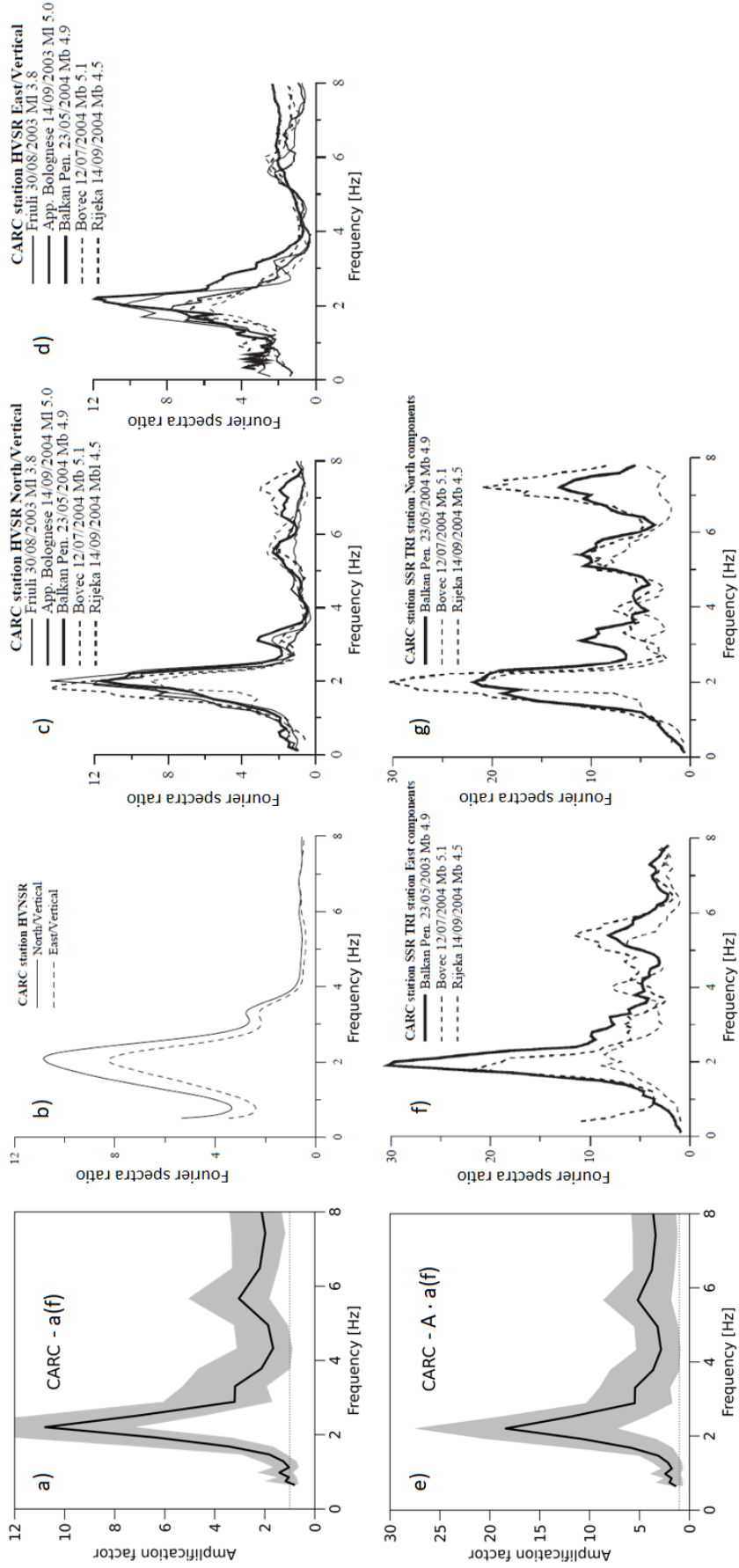
Figure 5-15: Site response functions determined from spectral inversion for station CARC (*a*, *e*), compared with results obtained by Fitzko *et al.* (2007) from HVNSR (*b*), HVSR (*c,d*) and SSR (*f,g*) methods. The black line in panels *a* and *e* represents the average site response function, while the shaded grey area represents the 1σ uncertainty range. Modified from Fitzko *et al.* 2007.

6 Spectral inversion: validation

The robustness of the inversion methodology and of the modelling choices was verified by running separate tests on the handling of uncertainty and path terms. As an example of possible application of the developed inversion framework, the case scenario of the 1998 Bovec-Krn earthquake was investigated.

6.1 Sensitivity tests

6.1.1 Use of uncertainty estimators

The uncertainty parameters $\boldsymbol{\varepsilon^{SO}}$, ε^{P} and $\boldsymbol{\varepsilon^{SI}}$ were added to the base formulation of FAS modelling with the aim of directly measuring the variability on source, path and site building blocks through the inversion. The results presented in Chapter 5 were derived by keeping all three terms in the forward model, so that the uncertainty estimated on the residuals (cf. Equation (5-9)) should represent the variability of the frequency-dependent amplification function only. The used parametrization choice is just one possibility among other possible solutions to estimate the uncertainties, such as leaving them unmodelled to extrapolate them in a second step using Generalized Inversion Technique.

The base formulation without epsilon terms was investigated, and the main inverted parameters compared to check that the choice on uncertainty handling does not affect the results of the inversion. The inversion was performed on the same dataset and under the same constraints and boundary conditions used in Chapter 5, except that all epsilon parameters were constrained to be null. A selection of inverted spectra is depicted in Figure 6-1, together with the reference model obtained by including uncertainty terms in the parametrization. All

modelled spectra fall within the 1σ uncertainty range associated to the results presented in Chapter 5.
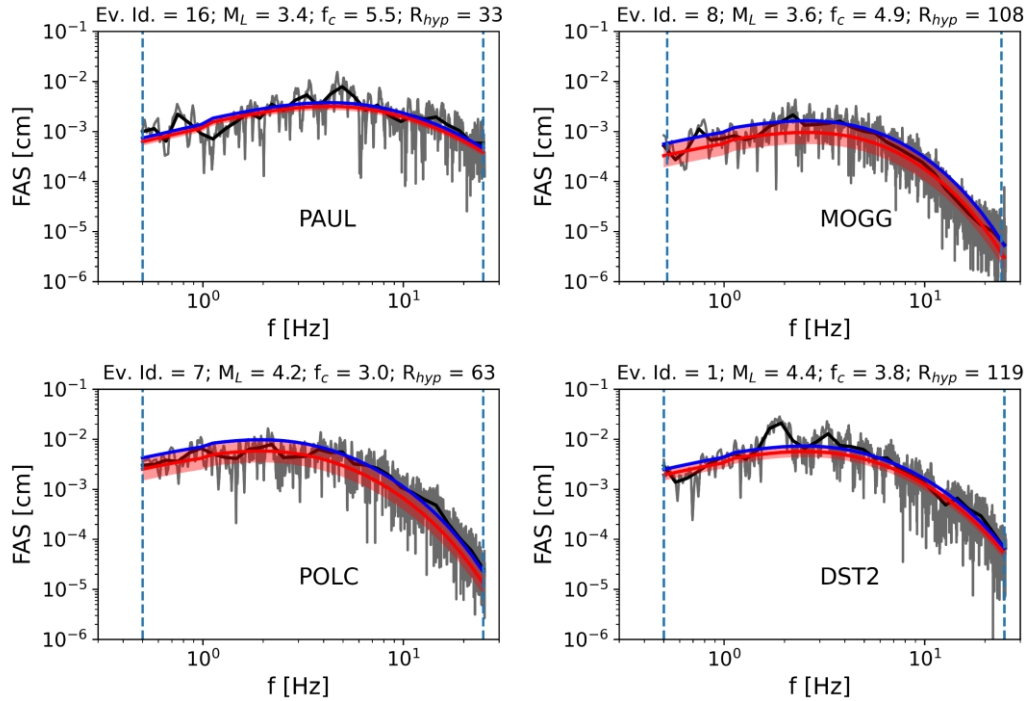
Figure 6-1: Example of fit curves for the observed velocity FAS (grey lines) obtained when the uncertainty terms are included in the model used for inversion (red lines) and when they are excluded (blue lines). The associated total uncertainty band obtained for the first case is shown as shaded red area. Inversion was performed on the smoothed version of the spectra (black lines), using spectral amplitudes inside the usable frequency range (vertical blue dashed lines).

The resulting source parameters are reported in Table 6-1, together with those obtained with the parametrization that uses epsilon terms. Figure 6-2 provides a visual representation of source parameter comparison. Both models are equally capable of reproducing the database magnitude and seismic moment values. Corner frequency values are also in good agreement, whereas the calculated stress drop is mostly comparable but slightly higher in the case without uncertainty modelling for events with highest magnitude. The spatial distribution of such stress drop values is shown in Figure 6-3.

Table 6-1: source parameters of the analysed events obtained with (ε) or without (no ε) uncertainty modelling. reference values are derived from the reviewed isc bulletin (international seismological centre 2020) for M_W and from the webservice of istituto nazionale di geofisica e vulcanologia (ingv; http://terremoti.ingv.it/) for M_0.

Event	Database M_W	M_W (ε)	M_W (no ε)	f_c [Hz] (ε)	f_c [Hz] (no ε)	Database M_0 [Nm]	M_0 [Nm] (ε)	M_0 [Nm] (no ε)	$\Delta\sigma$ [MPa] (ε)	$\Delta\sigma$ [MPa] (no ε)
1	4.08	4.08	4.09	3.84	3.84	1.03E+15	1.47E+15	1.55E+15	16.46	17.25
2	3.62	3.61	3.62	5.82	5.83	1.05E+15	2.92E+14	3.05E+14	11.37	11.92
3	3.15	3.12	3.11	6.54	6.69	n.d.	5.42E+13	5.19E+13	3.00	3.07
4	3.28	3.19	3.10	7.63	7.90	n.d.	6.75E+13	5.02E+13	5.92	4.88
5	4.15	4.24	4.36	3.28	3.26	1.08E+16	2.60E+15	3.86E+15	18.16	26.47
6	3.35	3.49	3.64	4.20	4.21	1.66E+14	1.90E+14	3.24E+14	2.77	4.76
7	3.95	3.96	3.99	2.95	2.95	5.6E+14	9.84E+14	1.10E+15	5.01	5.59
8	3.48	3.54	3.61	4.91	4.89	n.d.	2.26E+14	2.90E+14	5.27	6.70
9	3.55	3.52	3.51	5.01	4.99	1.24E+14	2.13E+14	2.06E+14	5.31	5.04
10	2.88	2.93	3.00	6.48	6.45	n.d.	2.82E+13	3.60E+13	1.51	1.91
11	3.62	3.63	3.66	3.53	3.51	1.32E+14	3.10E+14	3.46E+14	2.69	2.95
12	3.82	3.91	4.03	2.24	2.24	1.14E+15	8.35E+14	1.26E+15	1.86	2.78
13	4.08	4.19	4.32	1.79	1.79	5.91E+15	2.16E+15	3.34E+15	2.45	3.75
14	3.28	3.27	3.27	5.13	5.13	n.d.	8.97E+13	9.12E+13	2.39	2.43
15	2.95	2.91	2.88	7.79	7.66	n.d.	2.56E+13	2.38E+13	2.39	2.11
16	3.35	3.44	3.55	5.54	5.52	1.54E+14	1.63E+14	2.39E+14	5.45	7.93
17	3.42	3.36	3.34	5.50	5.44	n.d.	1.25E+14	1.13E+14	4.11	3.60
18	2.82	2.77	2.75	10.39	10.21	n.d.	1.62E+13	1.50E+13	3.58	3.16
19	3.08	3.05	3.04	5.87	5.87	n.d.	4.21E+13	4.02E+13	1.68	1.60
20	2.68	2.72	2.77	9.00	9.02	n.d.	1.33E+13	1.59E+13	1.92	2.31
21	3.62	3.57	3.54	6.67	6.58	n.d.	2.51E+14	2.28E+14	14.74	12.82
22	3.15	3.00	2.88	12.98	12.31	n.d.	3.56E+13	2.35E+13	15.41	8.66
23	2.82	2.76	2.72	9.76	9.90	n.d.	1.55E+13	1.36E+13	2.85	2.61
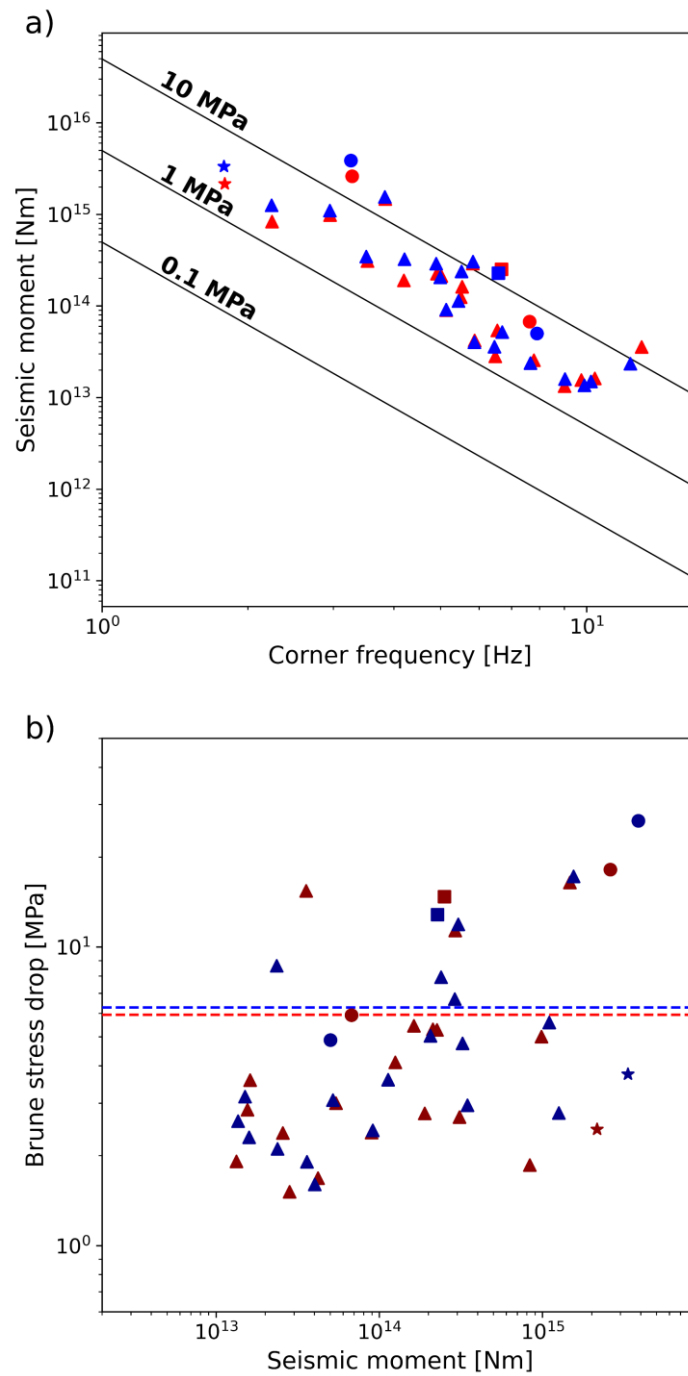
Figure 6-2: a) Comparison of inverted seismic moment versus corner frequency; red and blue markers are results obtained from the parametric model including (ε) and excluding (no ε) uncertainty terms, respectively. Lines of constant Brune stress drop are also shown (black lines). b) Comparison of inverted Brune stress drop versus seismic moment (coloured markers) and related average value (dashed lines). Colour coding is the same as above. Circles, triangles and squares represent events occurred in *ZS904*, *ZS905* and *ZS906,* respectively. The star marks event #13 in Table 5-2, occurred outside of *ZS9* zonation.
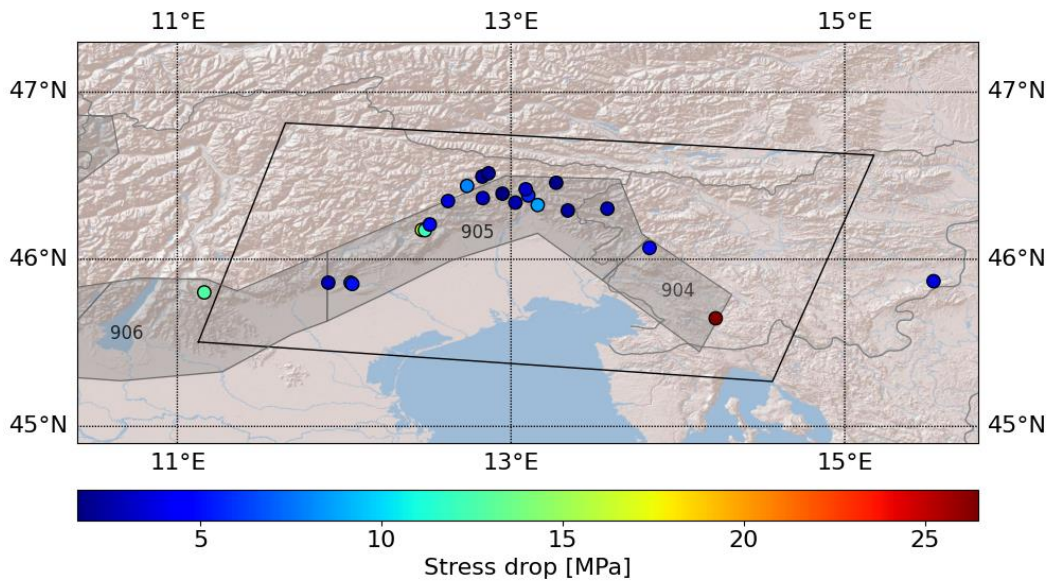
Figure 6-3: Spatial distribution of Brune stress drop estimated from source parameters obtained for the case without uncertainty modelling. *ZS9* seismogenic zones are plotted for reference (grey polygons). Even if the maximum value is higher than the one obtained with uncertainty modelling, all remaining values are overall comparable (cf. Figure 5-10).

A visual comparison of attenuation-independent elastic site response functions is presented in Figure 6-4. Single frequency-dependent amplification curves (left column in Figure 6-4) are in good agreement, especially for rock sites, while for soil sites the shape and position of the peaks is analogous but the amplification values are slightly different. This is probably due to the accommodation of residuals performed by the algorithm, as no dedicated uncertainty term is used. Even so, when the total attenuation-independent site response functions are considered (right column in Figure 6-4), results obtained with the two parametrizations are fully compatible and show comparable 1σ bounds.

Overall, the comparison between results obtained from models including or excluding uncertainty parametrization proves that they are both equally valid solutions to perform FAS inversion and confirms the validity of the parametrization choice made in Chapter 5 to retrieve relatable source description and site amplification curves.
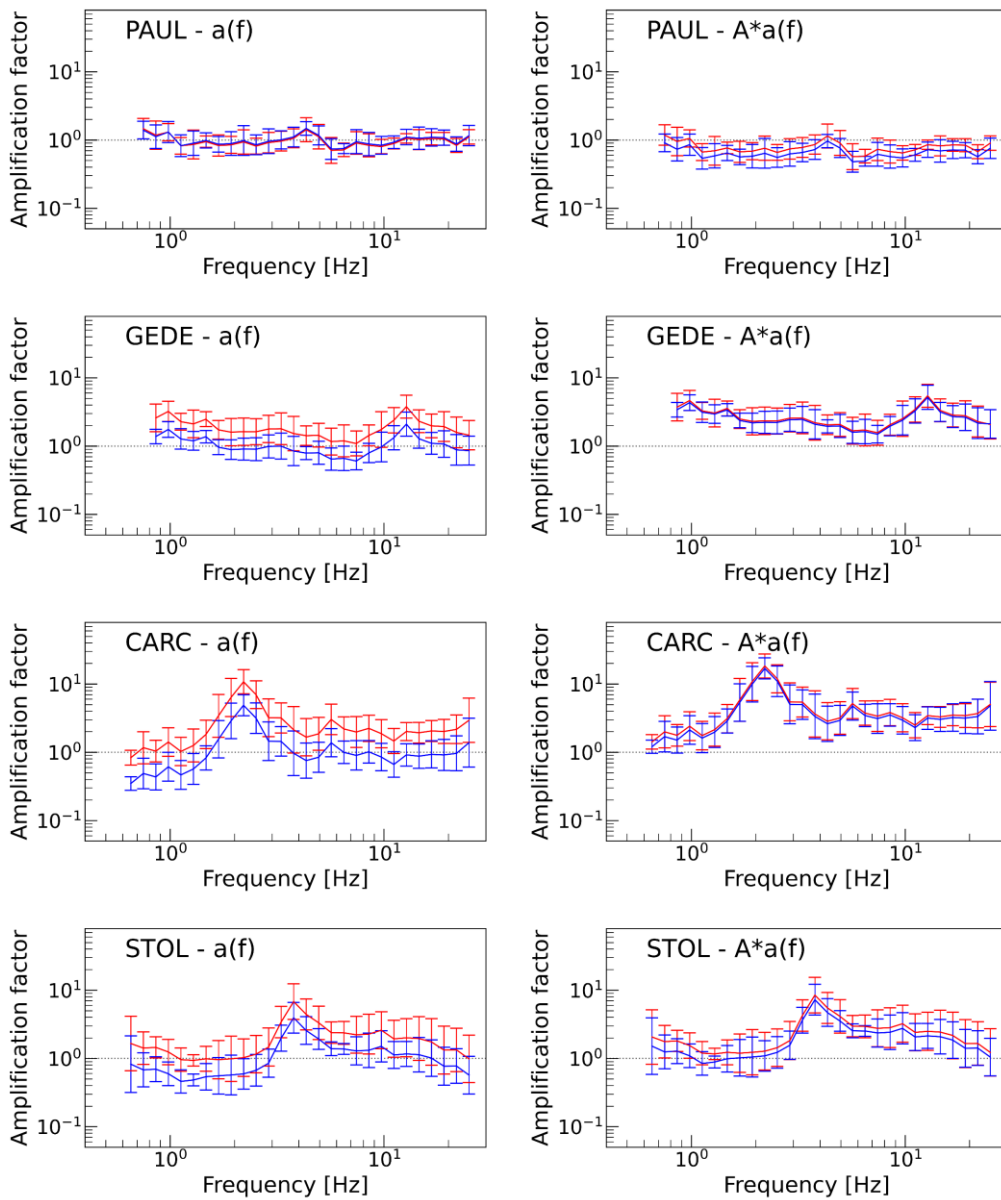
Figure 6-4: Frequency-dependent site amplification curves (a(f), left column) and elastic site response functions (ampl = A × a(f), right column) for stations PAUL (EC8-A), GEDE (EC8-B), CARC and STOL (*EC8-C*). Red and blue lines are results obtained from the parametric model including (ε) and excluding (no ε) uncertainty terms, respectively. Error bars are 1σ confidence levels.

6.1.2   Attenuation parameters

The fall-off of FAS at high frequencies derives from the combined effect of anelastic attenuation Q and site-related attenuation κ_0. Their combined effect can be represented through the whole path anelastic attenuation operator, t^* (Anderson and Hough 1984):

$$t^*_{ij} = \frac{r_{ij}}{v_s Q} + \kappa_{0j} \,, \tag{6-1}$$

where v_s is the average shear wave velocity used to infer Q and r is the hypocentral distance. t^* models spectral decay at high frequencies due to both path and site effects. In this formulation, κ_0 corresponds to the attenuation effects along the portion of ray paths nearest to the station, whereas Q represents the attenuation effects along ray paths from the source to the station proximity. In the inversion procedure described in Chapter 4 and Chapter 5, the simultaneous estimation of both Q and κ_0 was performed with the awareness of the potential trade-off existing between these two parameters, as the focus was mainly on source and site characteristics. The use of fixed parameters taken from literature to describe the geometric spreading function, and the use of a frequency-independent model for Q, were aimed at mitigating trade-off effects, but the results obtained for Q_0 and κ_0 were still possibly weaker and needed further investigation.

For this reason, an alternative approach was investigated, aimed at retrieving more reliable attenuation values without the need of a full inversion procedure to obtain all the spectral parameters. A new parametric inversion was performed based on the same structure used for the preceding results, except that the forward model was changed to include a single t^* term to avoid the possibility of influencing the Q_0 model by using site-dependent κ_0 terms:

$$\log FAS_{ijk}\left(r_{ij}, f_k\right) = \log(2\pi f_k) + \log\left(\frac{\Theta_{\lambda\varphi}F\xi}{4\pi\rho\beta^3 R_0}\right) + \log M_{0i}$$

$$-\log\left(1 + \left(\frac{f_k}{f_{ci}}\right)^2\right) + \log G\left(r_{ij}, f_k\right) \qquad \text{(6-2)}$$

$$+ A_j - \pi f_k t_{ij}^* \, .$$

The regression was focused on retrieving the frequency-dependent terms, and t^* specifically, so the functional form without uncertainties was preferred. To address the problem of two unresolved degrees of freedom, that is, M_0 and A, the inversion was constrained using M_0 values obtained from database M_L using Equation (5-4) and Equation (5-5). A reference rock amplification condition was imposed through Equation (5-3) to limit bias in the determination of t^* due to crustal amplification. This reference model defines the motion at the base of the soil column beneath the station.

A least-squares fit was performed using Equation (6-2) with $Q = Q_0$ over the dataset containing t^* values against hypocentral distances. The dataset common quality factor (obtained as the inverse of the slope divided by v_s) and site attenuation are $\overline{Q_0} = 1446$ and $\overline{\kappa_0} = 0.0314$ s, respectively. The 68% confidence interval of the best-fit slope corresponds to $Q_0 = 1271$ and $Q_0 = 1621$, whereas the 68% confidence interval of the intercept corresponds to $\kappa_0 = 0.0296$ s and $\kappa_0 = 0.0332$ s. This average quality factor is comparable even if slightly higher than the value obtained from directly parametrizing the quality factor in the model, $Q_0 = 1178$. In turn, both values are definitely higher than the ones reported in earlier studies for the Friuli-Venezia Giulia region ($Q_0 = 260$ from Malagnini *et al*. 2002; $Q_0 = 80$ from Console and Rovelli 1981). This difference can be due to the discrepancies in data-selection criteria and in the covered regional area and range of hypocentral distances used in fitting $t^* - r$ data.

Station-specific estimates of κ_0 were calculated to compare them with values obtained from directly parametrizing the site attenuation in the model. Moreover, some engineering applications, like stochastic simulations (Boore 2003), require a prior station-specific measurement of κ_0, so the average value $\overline{\kappa_0}$ cannot be used.

Following Bora *et al.* (2017), individual records of t^* were corrected for the slope in the $t^* - r$ straight line fit, corresponding to the regional value $\overline{Q_0}$, and the average value was taken as the station-specific κ_0. Figure 6-5 depicts variation of station κ_0 and associated 16%-84% confidence interval, together as values obtained from direct parametrization, for stations for which an associated v_{S30} value is available, either from topography or from specific assessment. The mean value of all site-specific κ_0 in the sample is $\kappa_0 = 0.042\ s$, in line with the values reported in literature by Gentili *et al.* (2011) ($\kappa_0 = 0.047\ s$) and Malagnini *et al.* (2002) ($\kappa_0 = 0.045\ s$). Station-specific κ_0 values are systematically higher than those obtained through direct parametrization (red squares in Figure 6-5). This is likely the effect of the bias between modelled κ_0 and Q_0 parameters, which in fact results in underestimated values for both. Consequently, results derived from t^* should be preferred.
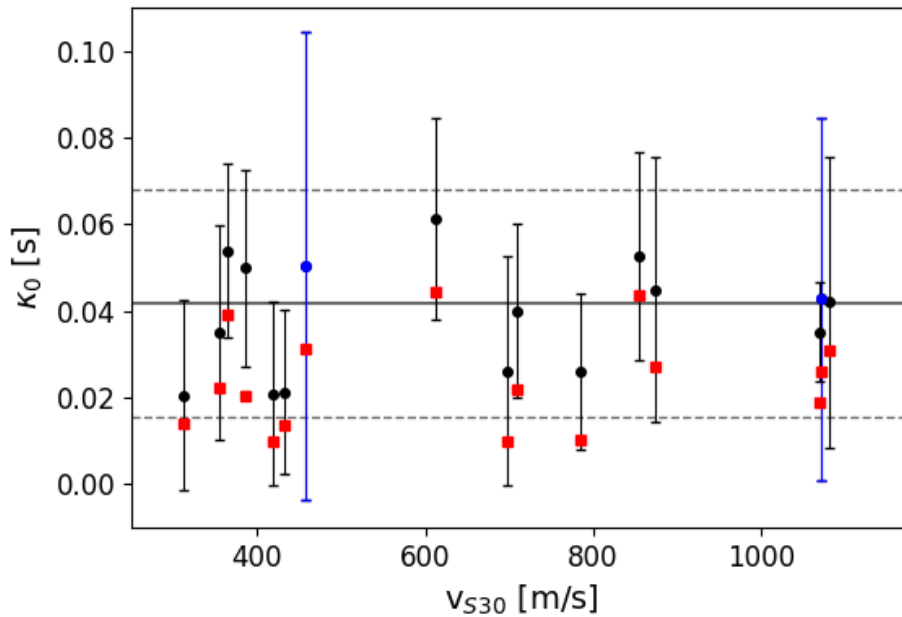
Figure 6-5: Station-specific κ_0 plotted against v_{S30} values. Point markers indicate the median derived from t^*, with associated error bars indicating the values corresponding to 16 and 84 percentiles at each station. Blue colour marks stations for which within-event variability is strongly influenced by inclusion of high hypocentral distances (RST, TLM2). Square markers are κ_0 values obtained from direct parametrization in the inversion model. The horizontal solid grey line indicates the median of all site-specific κ_0 in the sample, the dashed horizontal grey lines indicate the corresponding between-station variability in terms of 16 and 84 percentile values.

The record-to-record (within-station) variability of κ_0 (vertical error bars) is due to the fact that Q_0 is not homogeneous with respect to depth (Edwards *et al.* 2008; Edwards *et al.* 2011). Resultant within-station variability is rather large, which is probably the effect of estimating κ_0 from both near and distant earthquakes while using a homogeneous Q model. Even so, within-station variability is in most cases comparable to the station-to-station (between-station) variability of κ_0 (horizontal dashed-lines), which is mainly affected by regional variations in Q_0. The main exceptions are for stations with highest associated hypocentral distances (RST and TLM2, in blue in Figure 6-5), for which within-station variability is not reliable as it is likely overestimated.

A spatial representation of the station-specific κ_0 is presented in Figure 6-6, together with the corresponding distribution of base site amplification with respect to the reference rock condition. Although no clear correlation emerged between κ_0 and v_{S30}, stronger site attenuation is found with a stronger site amplification. This implies that, on a regional scale, softer soil regions amplify the whole frequency range, while attenuating the high frequencies more strongly. This trend is more evident in Figure 6-7, showing a direct comparison of κ_0 and A. Edwards *et al.* (2008) suggested that this could be explained by the fact that the possible causes of mean amplification (a soft rock site) and of strong local attenuation (partially saturated porous soil) generally occur together. Thus, hard rock sites with little soil coverage lead to mean deamplification and low site attenuation, while soft rock sites with thick soils result in relative amplification, along with strong site attenuation.
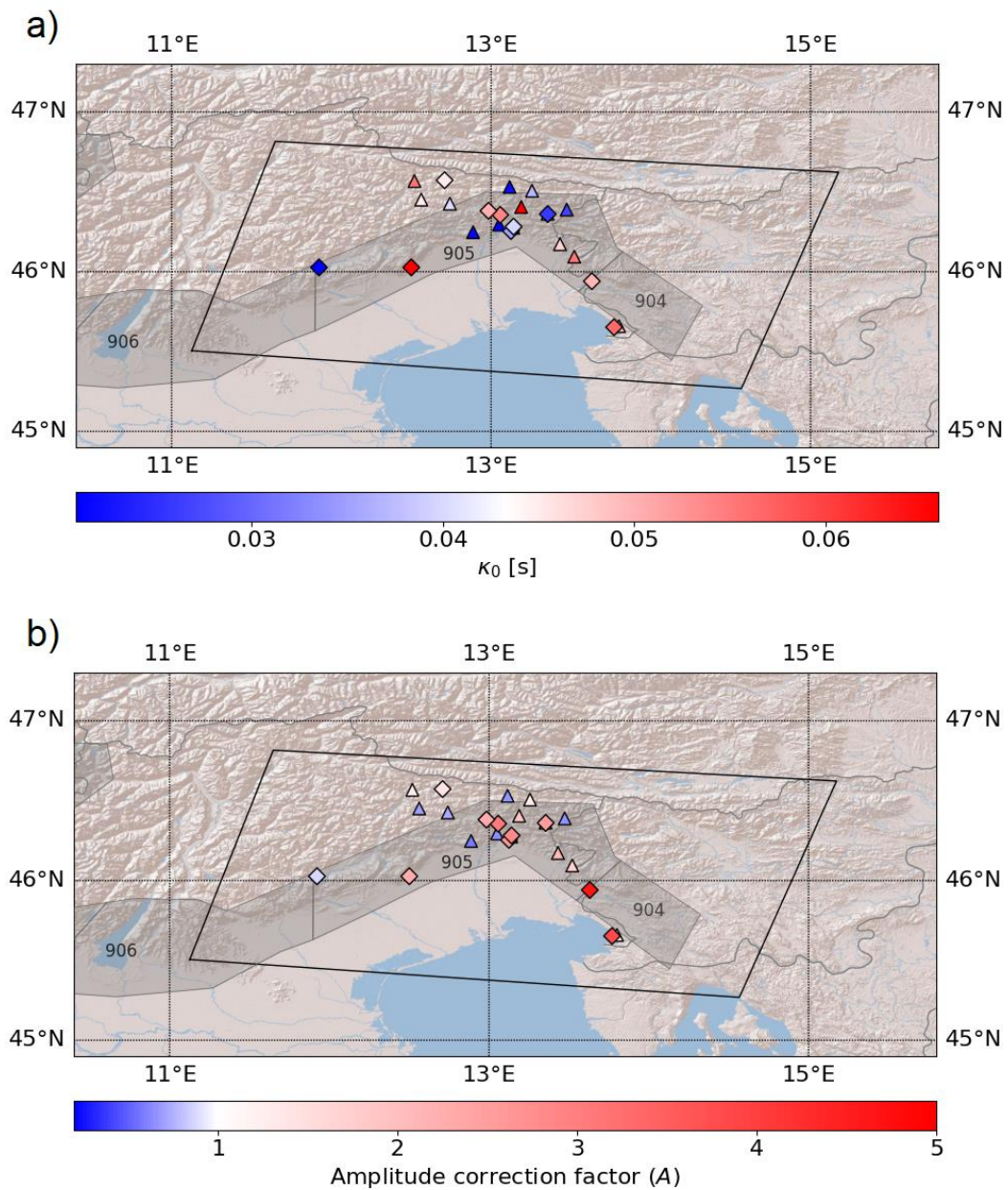
Figure 6-6: a) Spatial distribution of station-specific attenuation factors (κ_0) derived from inversion using t^*. b) Spatial distribution of mean amplitude correction factors (A) obtained from inversion using t^*. Triangles mark stations catalogued as rock sites (*EC8-A*), diamonds mark stations on soil (*EC8-B* and *EC8-C*).
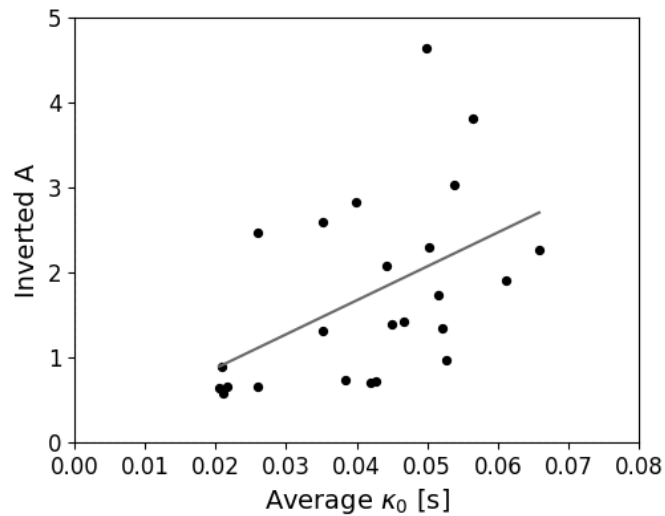
Figure 6-7: Direct comparison of site attenuation κ_0 and site amplification A (black markers). The associated linear regression result is shown for comparison (grey line).

6.2 The 1998 Bovec-Krn scenario

As an example of possible application of the model developed in this work, the case scenario of the 1998 Bovec-Krn earthquake was investigated. This also allows to provide validation of the results obtained for station-specific amplifications on independent data, other than the one used for inversion.

On April 12, 1998, at 10:55 GMT a $M_W = 5.6$ earthquake struck the Krn Mountains area in North-Western Slovenia close to the town of Bovec, near the Italian border (Figure 6-8). This event is the strongest earthquake occurred in Slovenia in the 20th century. The epicentral area itself was characterized by weak seismicity during the last century, although the wider area (Upper Soca valley) is one of the three zones with highest seismic hazard in Slovenia, due to its proximity to the seismically active area of Friuli-Venezia Giulia in Italy. An event re-localization was performed by Bajc *et al*. (2001) by adapting the standard Joint Hypocentre Determination (JHD) method for teleseismic data to local earthquakes. The epicentre is about 40 km east of the destructive Friuli 1976 thrust faulting earthquake (Aoudia *et al*. 2000) and is located near the junction

between Alpine structures and external Dinaric structures (Poljak *et al.* 2000). Bajc *et al.* (2001) obtained a focal mechanism corresponding to an almost pure strike-slip faulting, in agreement with the NW-SE trend of the Dinaric structures. The correspondent seismic moment is $M_0 = 4.5 \times 10^{17}$ Nm, obtained though inversion of strong motion records. A comprehensive review of the seismological investigations related to the earthquake can be found in Gosar (2019).
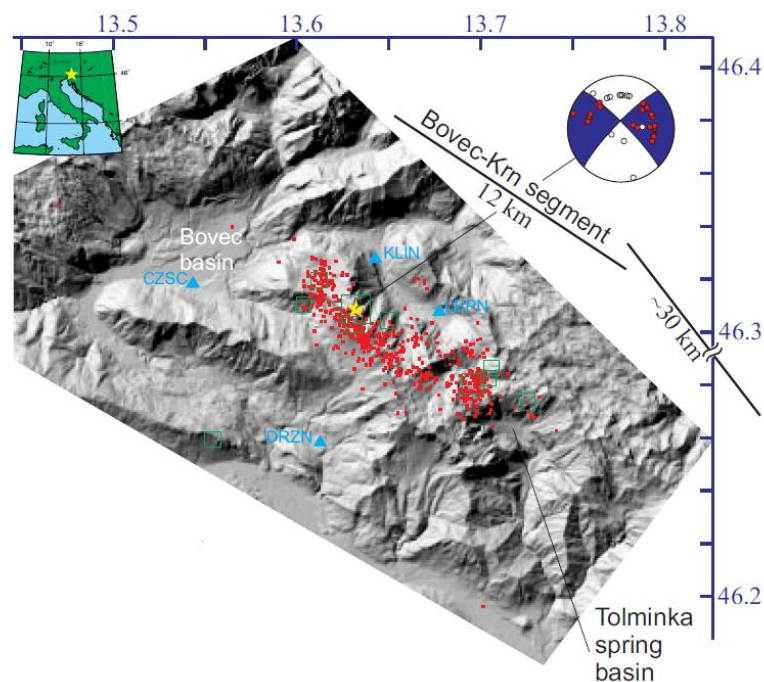
Figure 6-8: The Bovec-Krn earthquake sequence. Yellow star is the main shock; red dots are the relocated events in the seismic sequence; green squares are events with $M_L > 3.0$, with size proportional to the magnitude. Triangles are the closest temporary stations. The P-wave focal mechanism of the main shock is reported. Modified from Bajc *et al.* (2001).

Waveform recordings were retrieved from the revised database provided by Zivcic *et al.* (2000), which combines datasets from RF, OX and SL networks together with temporary local networks. The stations located inside the case study area for which such recordings were available are MOGG, GEPF, GESC, GETM, and PRAD. Waveforms were processed following the same procedure described in Chapter 2 and §5.2. To guarantee the homogeneity of the results, the duration

of the pre-event noise window was kept at ten seconds, even if this meant excluding the triggered records from station GEPF.

For each remaining station, a synthetic scenario for rock condition was built using the model described by Equation (4-12), without the uncertainty terms as in this case the aim is a simple comparison of the results and not the actual data inversion. The used parameter values are the same reported in Chapter 5, namely parameters for geometrical spreading taken from Malagnini *et al.* (2002) and the Q_0 value. In order to create a modelling relative to rock condition, the reference amplification term A was set to one, and the average attenuation factor κ_0 obtained for rock sites was used for all stations. Source parameters M_0 and f_c had to be taken from literature, as the event is not among those used in the inversion dataset. Besides Bajc *et al.* (2001), other authors provided a parametrization for the event. Franceschina *et al.* (2006) obtained $M_0 = 1.14 \times 10^{17}$ Nm and $f_c = 0.46$ Hz from parametric inversion performed of a regional dataset. The ITACA webservice (D'Amico *et al.* 2020) reports a seismic moment of $M_0 = 3.5 \times 10^{17}$ Nm, together with a slightly different localization.

Table 6-2: Literature source parameters for the 1998 Bovec main shock. Fields marked with an asterisk are not reported in the relative original study; missing localizations are assumed from ITACA (D'Amico *et al.* 2020), missing stress drop values are assumed from Franceschina *et al.* (2006).

Author	Latitude (°N)	Longitude (°E)	Depth (km)	M_0 [Nm]	f_c [s]	$\Delta\sigma$ [MPa]
Bajc *et al.* (2001)	46.3090	13.6320	7.6	4.5×10^{17}	n.d.	n.d.*
ITACA (D'Amico *et al.* 2020)	46.2506	13.6661	5.1	3.5×10^{17}	n.d.	n.d.*
Franceschina *et al.* (2006)	n.d.*	n.d.*	n.d.*	1.14×10^{17}	0.46	2.62

Unfortunately, no unique information on both the event localization and the estimated M_0 and f_c (or $\Delta\sigma$) values at the same time was available, so different source parametrizations were considered (cf. Table 6-2) and the results compared. Figure 6-9 displays the visual comparison between the observed spectrum and the modelled one at rock condition using different source configurations. The associated uncertainty ranges are all equivalent and related to the path uncertainty term. A variability related to the selected source parametrization emerges. Curves that use the parametrization from Bajc *et al.* (2001) and ITACA database (green and red curves in Figure 6-9) show only a slight difference compatible with the uncertainty bounds, whereas Franceschina *et al.* (2006) proposed a lower value of M_0 which results in a lower amplitude curve (blue curves). Regardless of the chosen source parametrization, the models simulating the Fourier amplitude spectra at rock condition are in good agreement with the main trend of the observed FAS. This is especially true for stations MOGG and PRAD, which are actually located on rock (*EC8-A* class), while for stations GESC and GETM, which are located on sediments, the actual amplitude is reasonably higher than the one estimated at rock condition. A GMPE-like representation of the rock scenarios for different magnitudes is presented in Figure 6-10, in terms of modelled spectral amplitude at selected frequencies (1 Hz, 5 Hz and 10 Hz), together with the three models considered for the Bovec main shock (purple lines). Individual observed data at *EC8-A* stations for events in the range $M_L \pm 0.3$ are plotted for comparison. In the case of lower magnitudes (yellow, orange, and red lines in Figure 6-10) they coincide with values used to calibrate the model. Data observed for the Bovec event mostly fall inside the uncertainty range of the modelled spectra.
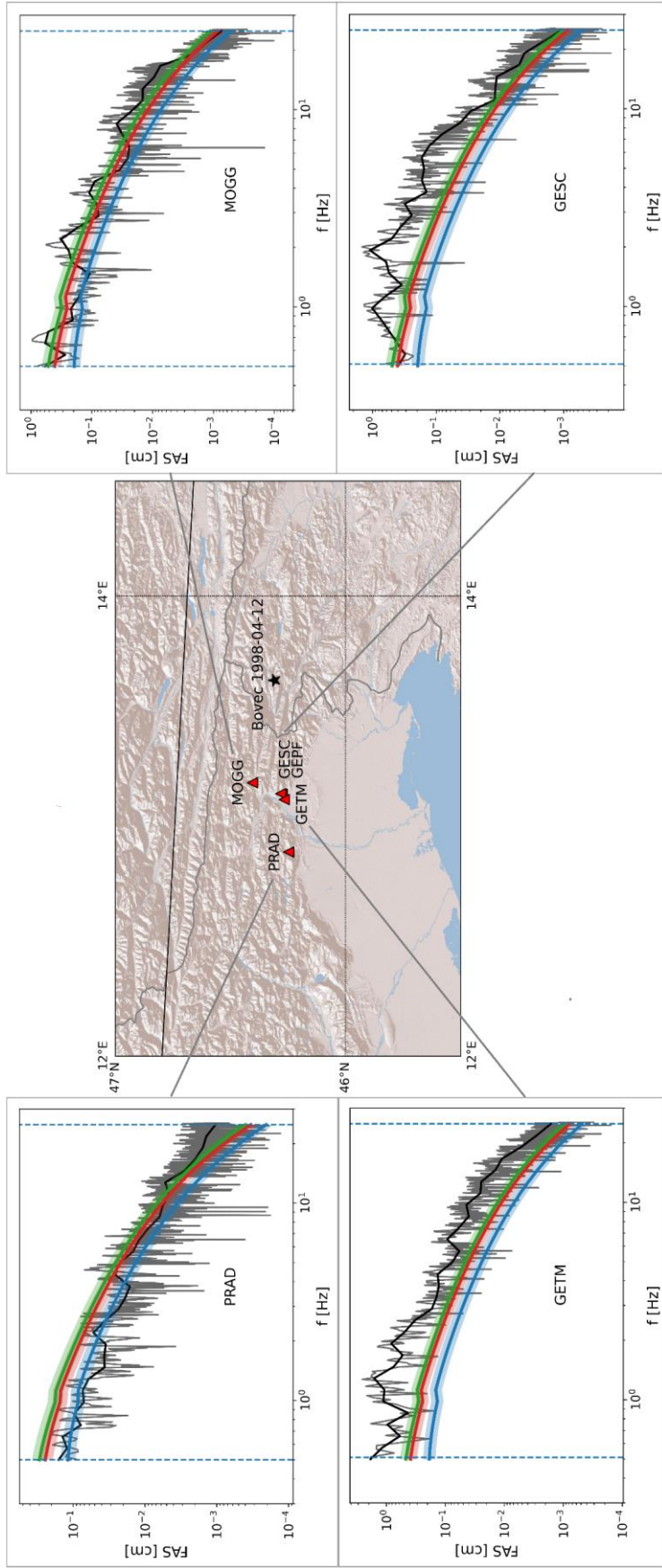
Figure 6-9: Localization of stations inside the case study area for which recordings of the main shock were available, with associated visual comparison between the observed spectrum and the spectrum modelled at rock condition using different source configurations. Green, red, and blue lines are models obtained using Bajc *et al.* (2001), ITACA (D'Amico *et al.* 2020) and Franceschina *et al.* (2006) source parameters, respectively.
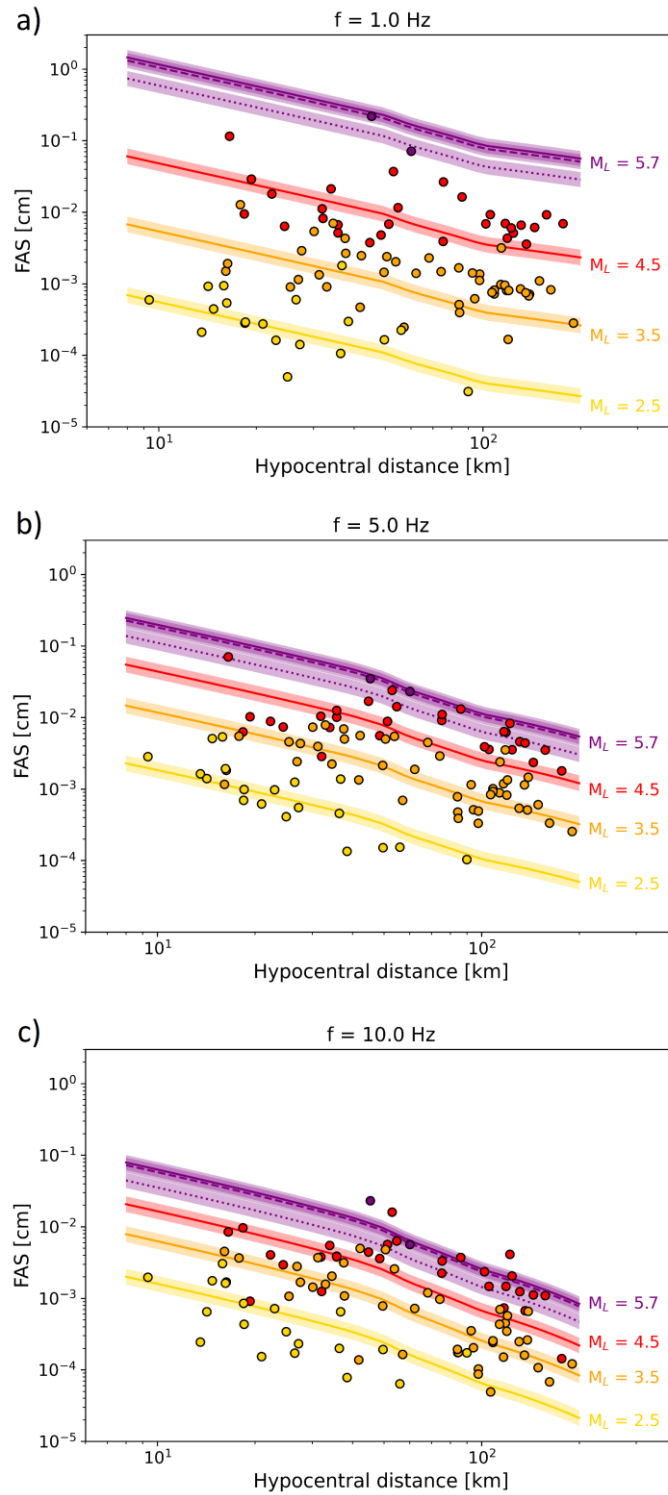
Figure 6-10: Rock FAS scenarios from inverted parameters as a function of hypocentral distance, at 1 Hz (a), 5 Hz (b) and 10 Hz (c). Three sample magnitudes are represented ($M_L = 2.5$ in yellow, $M_L = 3.5$ in orange, $M_L = 4.5$ in red), together with the models considered for the Bovec main shock (purple lines). Individual data observed at *EC8-A* stations for events in the range $M_L \pm 0.3$ are plotted for comparison (filled circles) using the same colour code.

Among the available stations, GESC, MOGG and PRAD were also part of the dataset used for parametric inversion in the Northeast Italy case study. As more information is available, for these stations also the full spectra are modelled, including terms of frequency-independent amplitude and attenuation. It is interesting to compare the residuals between observed and fully modelled spectra to the frequency-dependent amplification curves obtained from inversion (Figure 6-11). For stations located on rock (MOGG, PRAD), residuals mainly fall inside the 1σ uncertainty range independently of the source parametrization. In the case of MOGG station, models obtained from the parametrization by Bajc *et al*. (2001) and ITACA database fall inside the 2σ range, while all models differ from the inverted curves at higher frequencies ($f > 10$ Hz). This is likely the effect of the lower resolution of high-frequency amplitude peaks for stations located on rock. Interestingly, as a peak at a frequency slightly over 2 Hz is noticeable for all stations, that amplification could be due to a source feature rather than a site effect. As previously noted in §5.5, GESC station is located on alluvial fan sediments near Gemona del Friuli (Friuli-Venezia Giulia region) and has been known for showing high amplification peaks for some major events only, while the average site response is quite flat. In fact, the 1998 Bovec main shock is among these cases, and amplification recorded at GESC has been analysed in literature. Costa *et al*. (2006) found that the same event produces different resonant frequencies at GESC, and that different events cause different resonance frequencies observed at the same site. They proposed that this behaviour could be explained with complex subsurface structures under the sites that produce interference of seismic waves. They used gravimetric measurements to derive a 3D model of the alluvial fan under the city of Gemona, confirming that it cannot be modelled appropriately with a simple 1D structure and that observed site effects can be explained in terms of wave propagation in laterally heterogeneous media. For the 1998 Bovec earthquake, Costa *et al*. (2006) evidenced different amplification peaks in the ranges $3 - 4$ Hz and $5.5 - 7$ Hz using SSR and H/V analyses. Results obtained from spectral modelling are in good agreement with these findings, with two peaks centred around 3.2 and 6 Hz (dashed vertical lines in Figure 6-11).
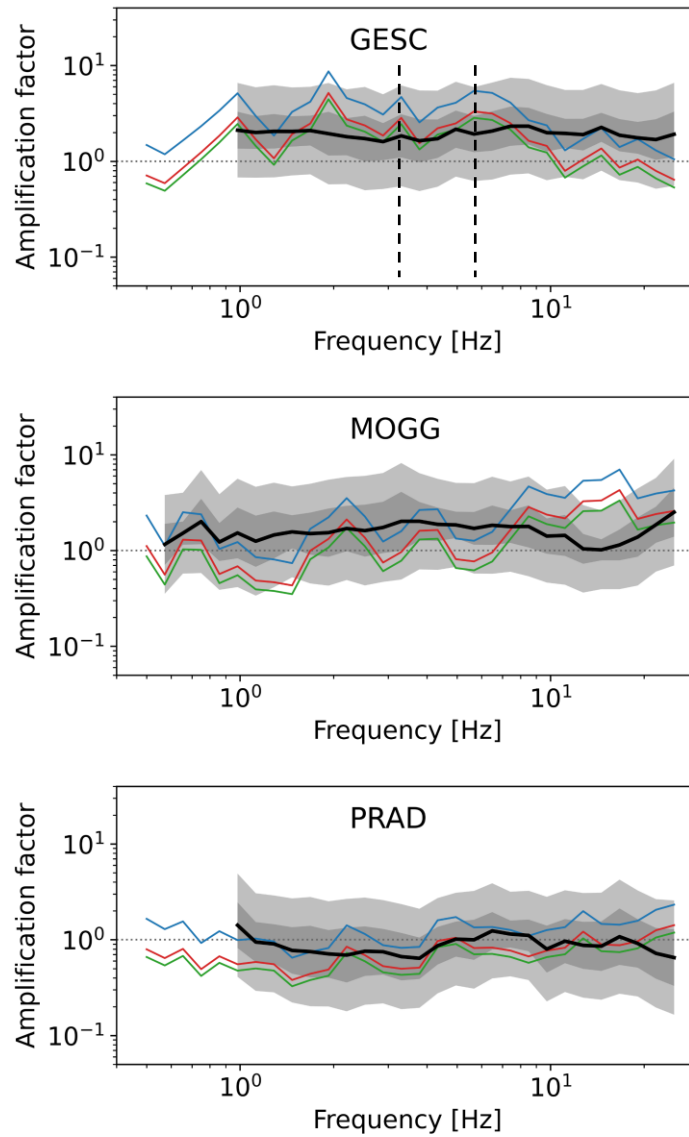
Figure 6-11: Comparison of the residuals between observed and modelled spectra (coloured lines) with the frequency-dependent amplification curves obtained from inversion (thick black lines). The grey areas represent 1σ and 2σ uncertainty bounds respectively, for each curve deriving from inversion. Green, red, and blue lines are residuals from models that use Bajc *et al.* (2001), ITACA (D'Amico e*t al.* 2020) and Franceschina *et al.* (2006) source parameters, respectively. Dashed vertical lines for station GESC mark the station-specific main peaks (3.2 and 6 Hz).

6.3 Discussion

A flexible parametric inversion strategy was developed to solve the inversion problem related to FAS modelling and applied to a case study area located in Northeast Italy. Even if it was oriented towards a specific application, the software used to perform the inversion was purposely built in a flexible, modular, and customizable way to ensure the possibility of different parametrization and methodology choices. These features combined represent an improvement with respect to currently available software for spectral inversion. Uncertainty parameters (epsilons) were added to the standard formulation of the FAS forward model that relies on source, path, and site terms. ε^{SO}, ε^{P} and ε^{SI} act as a collector for all the uncertainties related to the correspondent contribution, as a mean value over all frequencies. The software versatility allows to fix selected parameters, either to values taken from literature or to null values to "turn them off".

A case study area was selected in the Northeast Italy region, at the boundary with Slovenia and Austria. The area has significant seismic activity and suffered from destructive historical earthquakes. The case study area was defined as a simple polygon encasing all the main seismogenic features present in the region and including as many seismic stations as possible. Parametric inversion was performed on a dataset built from 234 velocity Fourier amplitude spectra, corresponding to 23 events recorded by 24 stations. Data was carefully deconvolved with seismograph response to ensure that the instrument response function term would not affect the inversion. Amplitudes were estimated on a set of $N_k = 30$ log-distributed frequency points in the range $0.5 - 25.0$ Hz, selected based on signal to noise ratio. Constant amplitude correction terms *A* are defined as the amplification relative to the network average rock (*EC8-A*) site. The apparent geometrical spreading parametric form was fixed a priori using trusted reference from literature to limit the trade-off between parameters.

Results obtained for source parameters are comparable with estimates from literature, proving the robustness of the inversion algorithm and its capability of realistically estimating information on the events used in the dataset. Elastic site

response functions are used to suggest a selection of sites to be used as regional reference. They are obtained by multiplying constant amplitude corrections for the frequency-dependent amplification curves, which are calculated as the geometric mean of the factorial residuals between observed and modelled spectra for each station, at each frequency. When a comparison is available, frequency-dependent amplification curves are in good agreement with literature. A set of stations with a flat response close to 1 and located with no ambiguity on rock according to geological maps was identified (AUP, AVS, DST2, FDS, GEPF, MASA, MOGG, PRAD, PURA, and RST), to serve as a tool in imposing reference site conditions for the region.

The robustness of the inversion methodology and of the underlying modelling choices was evaluated by running separate tests on the handling of uncertainty and path terms. The comparison between results obtained from models including or excluding uncertainty parametrization proves that they are both equally valid solutions to perform FAS inversion to obtain information on source and site amplification. On the other side, path contributions are known to be the weaker result, as a consequence of trade-off between quality factor and site attenuation. A control parametric inversion was performed in which M_0 values were fixed from bulletin and the forward model was changed to include a single t^* term including both contributions from Q_0 and κ_0. Both the average sample quality factor and the station-specific attenuation values, obtained from subsequent decomposition of t^*, are systematically higher than their counterparts obtained through direct inversion. Moreover, a positive correlation is observed between site attenuation and site amplification, in agreement with literature (e.g., Edwards *et al*. 2008). Results indicate that the inversion algorithm is capable of reliably retrieving source description and site amplification curves, whereas path parameters are only indicative and should be more precisely defined using other methodologies.

As an example of possible application of the developed inversion framework, the case scenario of the 1998 Bovec-Krn earthquake was investigated. Observed FAS were extracted from available waveforms for stations inside the case study area (MOGG, GESC, GETM, and PRAD) and compared with the corresponding

synthetic scenario for rock conditions. Regardless of the used source parametrization, the simulated rock models are in good agreement with the main trend of the observed spectra, in particular for stations actually located on rock. Full synthetic models were calculated by including information on frequency-independent amplitude and attenuation for stations belonging to the dataset used in the case study inversion. When compared with frequency-dependent amplification curves obtained from inversion, residuals between observed and modelled spectra mostly fall inside the uncertainty bounds of the curves. Additional observed peaks for the station located on soil are in good agreement with specific literature for the event at that station.

These results suggest that the developed inversion methodology can serve as a tool for building realistic rock scenarios for sites inside its application. Moreover, they provide reasonable validation of the results obtained for station-specific amplifications on independent data. Site response curves, especially in the cases with evident peaks, can be used as input in further applications that require a prior measurement of amplification at a given site. In particular, the possible inclusion of empirical site transfer functions into seismic hazard assessment as a substitute for proxy GMPE formulations has been investigated (e.g., Aristizábal *et al*. 2017). In this perspective, future refinement of the inversion procedure is advisable to obtain more precise information on attenuation directly form the inversion. Other possible improvements include the use of more refined source and quality factor description, other than the standard Brune source model and the frequency-independent Q, as well as the use of multi-step weighted inversion to improve the estimation of specific parameters. An essential requirement for such further implementation is the availability of more abundant data to ensure the statistic stability of the results, both in terms of waveforms number and of hypocentral, azimuth and magnitude range coverage. With this reasoning in mind, the processing algorithm was purposely built in a flexible way that ensures the possibility of adding new data as soon as it becomes available. In the same way, the parametric form itself can be easily modified with future and more precise information on regional characterization, such as the Q factor and the geometric spreading models.

7 Conclusions

This thesis work was aimed at improving the quality of tools available to seismic monitoring, which have a role also in the context of seismic hazard assessment. A novel definition of predicted macroseismic intensity with associated uncertainty was developed for Italy. A smaller region in Northeast Italy was also used as case study area to develop a solid spectral inversion algorithm for the characterization of the expected site-specific amplification at selected sites. Careful data collection and processing were applied to increase the quality and the stability of both methodologies.

The Gaussian Naïve Bayes technique was implemented for application to the Italian case, in discontinuity with the custom of using linear GMICEs. The strength of the implementation lies in the care employed in data processing and in correctly handling macroseismic data in particular. Most macroseismic databases include half-integer intensity measures as a form of uncertainty assignment; if used as-is, such data often leads to a false belief of having achieved an improvement in forecasts, whereas the actual uncertainty is in fact hidden. A pre-processing procedure was selected to transform the initial database in a more compliant version according to the definition of the MCS scale. Half integer values were reassigned to the nearby integer classes, with the use of weights to include the information on uncertainty, so that only integer classes would be used in the calculations. The complete database counted 90 events occurred from 1972 to 2016, with 376 associated GMP-intensity pairs and intensity values between $I_{MCS} = $ II and $I_{MCS} = $ X. Eight different GMPs were investigated.

A definition of instrumental intensity was provided as the class with highest associated probability based on input PGA or PGV value. The forecast values have by default an associated uncertainty measure in terms of the posterior probability calculated by GNB for that class. A set of GMICEs was also estimated

on the same dataset for comparison. They were developed according to their most used formulation, but the use of additional pre- and post-processing ensured that the knowledge on issues related to the use of discrete ordinal values would be taken into consideration. Results from both GNB and GMICEs were interpreted and compared. Both Cross-Entropy loss scores and confusion matrices indicated that overall GNB models perform better than GMICEs on the whole considered intensity range, both in terms of average performance on unseen data and of capability to capture data uncertainty. Among the considered parameter classes, forecasts based on PGV appeared to provide the best scores, while more data would be needed to obtain stable results based on spectral parameters. The obtained GNB classification models were also converted to GMICE-like scales so that they can find direct application in seismic monitoring products such as ShakeMaps.

A flexible parametric inversion strategy was developed to solve the inversion problem related to Fourier amplitude spectra modelling. A modular and customizable algorithm was implemented to ensure the possibility of different parametrization and inversion method choices. Uncertainty estimators were added to the standard formulation of the FAS forward model that relies on source, path, and site terms. The software versatility allows to fix selected parameters, making it easy to use it in performing multi-step regressions with different settings in each step. These features combined represent an improvement with respect to other available software for spectral inversion.

A case study area was selected in the Northeast Italy region, at the boundary with Slovenia and Austria. This region has a significant seismic activity and is well monitored by many seismic networks. Parametric inversion was performed on a dataset built from 234 velocity Fourier amplitude spectra, corresponding to 23 events recorded by 24 stations. Data was carefully deconvolved with seismograph response to ensure that the instrument response function term did not affect the inversion. A set 30 of log-distributed frequency points in the range $0.5 - 25.0\,\mathrm{Hz}$ was extracted from each spectrum, selected based on signal to noise ratio. Spectral amplification was defined with respect to the network average rock site. The

apparent geometrical spreading function was fixed a priori using trusted reference from literature to limit the trade-off between parameters. Results obtained for source parameters were comparable with estimates from literature, proving the robustness of the inversion algorithm. Inverted seismic moments ranged between 1.3×10^{13} and 2.6×10^{15} Nm; Brune stress drop values ranged between 1.5 and 18 MPa, comparably with reference values for the region (e.g., $0.1 - 10$ MPa, Franceschina *et al*. 2006). The average rate of attenuation was lower than the ones reported in earlier studies for the Northeast Italy region. This could be due either to the intrinsic variability in the seismic Q factor, or in the used regional area and range of hypocentral distances. Elastic site response functions were calculated using the results of both the parametric inversion and the probabilistic analysis of inversion residuals, and used to suggest a selection of sites to be used as regional reference. Site-specific amplification factors as high as 20 were found, with typical resonance frequencies in the range $2 - 9$ Hz. Resulting frequency-dependent amplification curves were in good agreement with site amplifications available in literature. For one station in particular (CARC), both the frequency and the amplitude of the main amplification peak at 2 Hz were correctly retrieved with respect to references obtained through other techniques, such as SSR and HVSR. The uncertainty associated to each frequency point of the site response curves was assessed statistically. A set of stations with a flat response close to one and located with no ambiguity on rock according to geological maps was identified to serve as a tool in imposing reference site conditions for the region.

The robustness of different aspects of the inversion methodology was evaluated by running separate tests on the handling of uncertainty and on estimation of path terms. The comparison between results obtained from models including or excluding uncertainty parametrization confirmed the stability of the results obtained for source and site parameters, and proved that different parametrizations could be used depending on the desired application. A control test was performed using yet another parametrization choice to avoid inverting for the site attenuation and the quality factor at the same time. Results for path parameters obtained from statistical analysis of the inversion products were more satisfying than their counterpart obtained directly through inversion, proving that inverted path

parameters are only indicative and should be more precisely defined using other methodologies. These findings were in agreement with observed trade-off effects between inversion parameters described in literature, specifically between the site attenuation, the quality factor and the geometrical spreading parameters. A case scenario was built as an example of possible application of the proposed algorithm. Observed FAS from available waveforms recorded by stations inside the case study area were compared with the corresponding synthetic scenario for rock conditions. The simulated rock models were in good agreement with the main trend of the observed spectra, in particular for stations actually located on rock. Residuals from full synthetic models, calculated for stations with available information on site-specific behaviour, were in agreement with the uncertainty range of amplification curves derived from inversion.

Both the proposed definition of instrumental intensity and the obtained information on site-specific spectral response contribute to the improvement of seismic monitoring products, both for usage in seismic hazard assessment and in civil protection and engineering applications. Results obtained from GNB classification can be directly used in the creation of shakemaps, with the additional advantage of providing class-specific uncertainty estimations. Analogously, site response curves with evident amplification peaks could be included into seismic hazard assessment in support of proxy GMPE formulations. Both methodologies developed for this work are flexible and can be easily updated using new data, allowing to extend their possible application to contexts other than Italy.

In this perspective, future refinement based on newly collected data should be easy to achieve as soon as more data becomes available. GNB classification could be calibrated at regional level to support the real-time generation of shakemaps operated by local seismological institutions. With more strong seismicity data available, more specific investigations could be carried out on damages observed in near-fault areas. Additional data would also ensure the stability of results obtained from more refined Fourier spectra models, while the addition of other independent modelling from novel literature could improve the regional

characterization. Some present aspects might also benefit from such improvements, namely the stability of GNB forecasts obtained from spectral parameters and the residual trade-off between path parameters. Nevertheless, the proposed procedures provide both results and stable, easily implementable tools that can be directly applied in the context of seismic monitoring and hazard assessment.

Appendix

Tables

Table A-1: All the events used for the ODR GMICEs estimations. [1] = partial/cut recording, for which spectral GMPs were discarded.

Date (M_W or M_L*)	Time	Epicentral Area (lat lon)	Station Name (lat lon)	Epicentral dist. [km]	I_{MCS}
1972-06-14 (4.8)	18:55:46	Ancona (43.65 13.60)	ANP (43.60 13.47) [1]	11.4	VII
			ANR (43.62 13.51) [1]	7.7	VIII
1976-05-06 (6.4)	20:00:12	Friuli (46.26 13.30)	CDR (45.96 12.98) [1]	48.4	VI
			CNG (45.88 12.29) [1]	91.1	VI
			CST (45.66 11.90) [1]	130	V
			FLT (46.02 11.91) [1]	110.1	V-VI
			TLM1 (46.38 12.98) [1]	21.7	VII
1977-07-24 (4.0*)	09:55:30	Irpinia (41.16 14.96)	ARI (41.15 15.09) [1]	11	V-VI
1978-03-11 (5.2)	19:20:43	Ferruzzano (37.98 16.18)	FRR (38.05 16.13) [1]	9.2	V
			PLL (38.02 15.65) [1]	46.7	V
			VLS2 (38.25 15.65) [1]	53.9	VI
1978-04-15 (6.0)	23:33:47	Golfo di Patti (38.27 15.11)	MLZ (38.23 15.24) [1]	12.2	VII
			NAS (38.19 14.79) [1]	33	VII-VIII
			PTT1(38.15 14.97) [1]	18.3	VIII
1979-09-19 (5.8)	21:35:37	Val Nerina (42.80 13.04)	ARQ (42.77 13.29) [1]	21	VI
			BVG (42.93 12.61) [1]	37.9	V
			CSC (42.72 13.01) [1]	9.3	VIII
			NCR (43.11 12.78) [1]	40.4	VI
			SPL (42.74 12.74) [1]	25.7	VI-VII
1980-11-23 (6.9)	18:34:53	Irpinia (40.76 15.31)	ALT (40.56 15.40) [1]	23.8	VIII
			BGI (40.83 15.07) [1]	21.8	VII
			BNV (41.12 14.80) [1]	58.5	VII
			BRN (40.47 15.63)	42.2	VIII
			BVN (41.25 15.34) [1]	54.4	VII
			MRT (40.79 14.76) [1]	46.1	VII-VIII
			STR (41.02 15.12) [1]	33.3	VII
			TDG (40.80 14.38) [1]	78.1	VI
			TRR (40.62 16.16) [1]	73.1	VI-VII

1981-06-07 (4.9)	13:01:00	Sicily (37.63 12.72)	MZR (37.65 12.61) [1]	9.7	VI
1982-03-21 (5.0*)	09:44:00	Calabria (39.70 15.64)	LRS (40.05 15.84) [1]	41.6	VII
1982-10-17 (4.1*)	04:50:12	Marche (43.13 12.60)	NCR (43.11 12.78) [1]	14.9	V-VI
1984-04-29 (5.6)	05:03:00	Gubbio (43.21 12.57)	CTC (43.46 12.25) [1]	38.1	VI
1984-05-07 (5.9)	17:49:43	Val Comino (41.70 13.86)	BRS (42.32 13.59) [1]	72.9	V
			BSS (42.19 13.84) [1]	54.5	VI
			ORT (41.96 13.65) [1]	33.7	VI
			PNT (41.50 13.68) [1]	26.9	VI
			RCC (41.29 13.98) [1]	46.9	VII
			SCF (42.26 14.00) [1]	63.8	VI-VII
			STG (41.57 14.23) [1]	34.1	VI-VII
			ATN (41.62 13.80) [1]	10.2	VII
			CSN0 (41.52 13.86) [1]	19.7	VII
1984-05-11 (5.5)	10:41:48	Val Comino (41.78 13.89)	PSC00 (41.81 13.79) [1]	9.1	VI-VII
			SCF (42.26 14.00) [1]	54.6	VI
			STG (41.57 14.23) [1]	37.1	VI
			VLB (41.76 13.99) [1]	8.5	VI-VII
			ATN (41.62 13.80) [1]	19.2	VI
			CSN0 (41.52 13.86) [1]	28.7	VI
1985-01-23 (4.7)	10:10:18	Garfagnana (44.14 10.57)	BRG (42.32 13.59)	3.7	V-VI
			SST (44.23 10.77) [1]	33.8	IV
			VGL (44.11 10.29) [1]	11.2	V
1985-05-20 (4.2*)	10:00:30	Aquilano (42.23 13.32)	BRS (42.32 13.59) [1]	19.1	IV-V
1987-05-02 (4.7)	20:43:54	Parma (44.81 10.72)	NVL (44.84 10.73) [1]	6.8	VI
1987-07-05 (4.4)	13:12:37	Casentino (43.74 12.22)	PGL (43.70 12.50) [1]	24.3	V
1988-02-01 (4.6)	14:21:40	Friuli (46.31 13.15)	TLM2 (46.38 12.98) [1]	7.6	V
1988-02-08 (4.6)	11:24:46	Garfagnana (44.17 10.54)	VGL (44.11 10.29) [1]	14	V-VI
1989-03-11 (3.7*)	21:06:00	Gargano (41.76 15.33)	SNN (41.83 15.57) [1]	18.1	V
1990-05-05 (5.8)	07:21:19	Potenza (40.68 15.85)	BRN (40.47 15.63) [1]	26.7	V-VI
			CLT (40.90 15.44) [1]	45.7	VI
			RNR (40.93 15.67) [1]	35.8	VI
1990-12-13 (5.6)	00:24:26	East Sicily (37.20 15.47)	GRR (37.73 15.16) [1]	52.6	V-VI
			NOT (36.9 15.07) [1]	46.8	VI
			PCH (36.71 15.09) [1]	65.4	VI-VII
			SRT (37.16 15.03) [1]	28.3	VI-VII
			VTT (36.95 14.52) [1]	79.6	V-VI
			VZZ (37.16 14.75) [1]	51.4	VI
1991-01-14 (4.0*)	07:38:37	Casentino (43.82 11.92)	CVT (44.01 11.94) [1]	20.1	IV-V
			STS (43.94 11.91) [1]	12.8	IV-V
1991-05-26 (5.1)	12:26:01	Potentino (40.68 15.81)	GRM (40.31 15.89) [1]	42.3	V
			RNV (40.93 15.67) [1]	28.9	VI
			TRR (40.62 16.16) [1]	31.6	V-VI
1992-04-06 (4.7)	13:08:33	Monti Nebrodi (37.80 14.64)	NCS (37.75 14.40) [1]	16.4	V
1993-06-04 (3.9*)	21:36:52	Perugia (43.14 12.67)	NCR (43.11 12.78) [1]	5.9	V-VI

1993-06-05 (4.7)	19:16:18	Perugia (43.15 12.67)	CSA (43.02 12.59) [1]	14.9	V
			NCR (43.11 12.78) [1]	15.2	V-VI
1995-09-30 (5.2)	10:14:34	Gargano (41.80 15.90)	LSN (41.85 15.36)	46.1	V-VI
			SNN (41.83 15.57) [1]	28.4	V-VI
			VSD (41.88 16.17) [1]	22.5	V
1996-10-15 (5.4)	09:56:00	Parma (44.80 10.66)	NVL (44.84 10.73)	13.2	VI
			SRB (44.87 10.46)	16.5	V
1997-09-03 (4.5)	22:07:30	App. Umbro-Marchigiano (43.02 12.88)	ASS (43.07 12.60)	22.8	IV-V
			NCR (43.11 12.78) [1]	12.1	V
1997-09-07 (3.8*)	23:28:06	App. Umbro-Marchigiano (43.03 12.80)	ASS (43.07 12.60)	23.1	III
			NCR (43.11 12.78) [1]	11.6	V
1997-09-26 (5.7)	00:33:11	Umbria Marche 1st shock (43.02 12.89)	AQK (42.34 13.40)	86.1	IV-V
			ASS (43.07 12.60)	24	V-VI
			BVG (42.93 12.61) [1]	24.9	VI-VII
			CSA (43.02 12.59)	24.5	VI
			CSC (42.72 13.01) [1]	35.2	VI
			NCR (43.11 12.78) [1]	13.1	VI-VII
1997-09-26 (6.0)	09:40:24	Umbria Marche 2nd shock (43.02 12.86)	AQI (42.34 13.40)	86.9	V-VI
			AQK (42.34 13.40)	86.9	V-VI
			ASS (43.07 12.60)	21.4	VI-VII
			BVG (42.93 12.61) [1]	21.8	VI
			CSA (43.02 12.59) [1]	21.4	VI-VII
			CSC (42.72 13.01) [1]	35.4	V-VI
			GBB (43.36 12.60) [1]	43.2	VI-VII
			NCR (43.11 12.78) [1]	12.1	VII-VIII
			PGL (43.70 12.50) [1]	81	V
1997-10-03 (5.2)	08:55:20	App. Umbro-Marchigiano (43.04 12.85)	ASS (43.07 12.60)	18.3	VI
1997-10-06 (5.4)	23:24:51	App. Umbro-Marchigiano (43.03 12.85)	ASS (43.07 12.60)	20.4	VI
			BVG (42.93 12.61) [1]	21.9	V-VI
			CSA (43.02 12.59) [1]	20.9	V-VI
			GBB (43.36 12.60) [1]	41.7	V-VI
			NRC (42.79 13.10) [1]	33.2	V-VI
1997-10-14 (5.6)	15:23:09	Umbria Marche 3rd shock (42.93 12.93)	AQI (42.34 13.40)	74	V
			AQK (42.34 13.40)	74	V
			ASS (43.07 12.60)	31	V
			BVG (42.93 12.61) [1]	23.7	VI
			CSC (42.72 13.01) [1]	22	VI
			NCR (43.11 12.78) [1]	20	VI
1997-11-09 (4.5*)	19:07:32	App. Umbro-Marchigiano (42.85 13.06)	CSA (43.02 12.59) [1]	37	IV-V
			NRC (42.79 13.10) [1]	10.7	V
			SELW (42.89 12.92)	7	V-VI
1998-02-07 (4.3*)	00:59:44	App. Umbro-Marchigiano (43.00 12.86)	NCB (43.10 12.81)	11	V
			NCR2 (43.11 12.78)	12.3	V

Date (Mag)	Time	Location	Station (coords)	Dist	Int
			NOCE (43.12 12.79)	13.1	V
			SELE (42.89 12.93)	15.4	V
			SELW (42.89 12.92)	15.5	V
1998-03-21 (5.0)	16:45:09	App. Umbro-Marchigiano (42.98 12.91)	NCB (43.10 12.81)	19.3	IV
			NCR2 (43.11 12.78) [1]	21	IV
			NOCE (43.12 12.79)	21.5	IV
1998-04-05 (4.8)	15:52:20	App. Umbro-Marchigiano (43.18 12.77)	ASS (43.07 12.60)	18.4	V
			NCB (43.10 12.81)	10.1	V
			NCR (43.11 12.78) [1]	8.8	V
			NCR2 (43.11 12.78)	8.8	V
			NOCE (43.12 12.79)	8	V
			SELE (42.89 12.93) [1]	35.9	IV-V
			SELW (42.89 12.92)	36	IV-V
1998-06-01 (3.8*)	13:57:09	App. Umbro-Marchigiano (43.18 12.80)	ASS (43.07 12.60)	18.9	IV
1998-06-02 (4.3)	23:11:22	App. Umbro-Marchigiano (43.17 12.82)	ASS (43.07 12.60)	19.2	III
1998-08-15 (4.4)	05:18:09	Monti Reatini (42.45 13.06)	CONT (42.42 12.77)	24.3	III
1998-09-23 (3.7*)	18:44:02	App. Lucano (39.97 16.03)	VGG (39.97 16.05)	6.5	IV-V
1999-07-07 (4.7)	17:16:11	Lunigiana (44.25 10.82)	ZCC (44.32 10.97) [1]	10.1	V
1999-11-29 (3.8*)	03:20:34	App. Umbro-Marchigiano (42.82 13.07)	NOR (42.79 13.09)	8.1	IV-V
1999-12-29 (4.8)	20:42:35	Massiccio Ortles (46.61 10.31)	SEM (46.49 10.27) [1]	14.4	V
2000-04-01 (4.5)	08:08:05	Monte Amiata (42.94 11.73)	PNC (42.85 11.71) [1]	2.3	VI
			PNS (42.85 11.69) [1]	1.6	VI
2000-05-08 (4.6)	12:29:56	Zona Ravenna (44.27 11.95)	FAZ (44.30 11.89)	3.7	V-VI
2000-05-10 (4.8)	16:52:11	App. Forlivese (44.25 11.94)	FNZ (44.30 11.89)	7.6	V-VI
2000-08-21 (4.8)	17:14:28	Alessandria (44.91 8.39)	NZZ (44.78 8.36) [1]	6.2	V-VI
2000-12-16 (4.2)	07:31:06	Terni (42.53 12.62)	NRN (42.52 12.52) [1]	5.5	V-VI
2001-07-17 (4.8)	15:06:14	Massiccio Ortles (46.63 11.05)	SLP (46.81 11.25) [1]	18.6	V
2002-02-14 (4.5)	03:18:02	Carnia (46.39 13.12)	MOGG (46.41 13.19)	6.3	IV-V
			CESC (46.36 13.06)	10	V-VI
			GESC (46.28 13.14)	17.9	IV-V
2002-04-05 (4.4)	04:52:21	Isole Eolie (38.35 15.10)	MFG (38.19 15.54)	50.4	III-IV
2002-04-17 (4.9)	06:42:54	Calabria (39.70 16.84)	ACR (39.49 16.38)	47.8	IV-V
			RSN (39.57 16.63)	24.7	IV-V
			SGV (39.26 16.69)	49.3	III-IV
2002-04-18 (4.4)	20:56:47	Vallo di Diano (40.61 15.59)	MRV (40.36 15.83) [1]	34.3	III
2002-09-06 (5.9)	01:21:28	Southern Tyrrenian (38.38 13.70)	CDI (37.49 14.64) [1]	131.3	IV
			CLG (37.21 14.52) [1]	150.7	IV
			PTT (38.134 14.975)	115.3	IV-V
2002-11-01 (5.7)	15:09:02	Molise 2nd shock (41.72 14.84)	CHT (42.37 14.15)	90.4	IV-V
			CMM (41.87 14.45)	35.5	IV
			ORC (41.95 13.64)	102.2	IV-V

			SCV (41.30 14.88) [1]	48.5	IV
			SSV (41.68 15.39) [1]	45.6	III-IV
2002-11-12 (4.6)	09:27:48	Subapp. Dauno (41.68 14.80)	CAMO (41.62 15.10)	27.4	III-IV
			CASA (41.74 14.85)	7.5	IV-V
			CAST (41.70 14.73)	4.7	V-VI
			CMM (41.87 14.45) [1]	34.3	IV
			SEP (41.63 14.88)	10.6	V
2003-01-26 (4.7)	19:57:03	App. Forlivese (43.88 11.96)	CVT (44.01 11.94) [1]	12.5	V-VI
			STS (43.94 11.91)	6	VI
2003-04-11 (4.8)	09:26:57	OltrePo Pavese (44.76 8.87)	TRT (44.89 8.88) [1]	17.2	V
2003-04-28 (3.4*)	20:12:36	SubApp. Dauno (41.69 14.85)	SEP (41.63 14.88)	7.8	IV
2003-06-01 (4.8)	15:45:18	SubApp. Dauno (41.66 14.82)	CHT (42.37 14.15)	96.4	II
			LDP (42.04 14.18)	67.5	III
			SEP (41.63 14.88)	6.3	IV
2003-09-14 (5.3)	21:42:53	Monghidoro (44.26 11.38)	FAZ (44.30 11.89)	41.0	IV-V
			FRE1 (44.13 11.40)	15.3	V-VI
2003-12-07 (4.2)	10:20:33	App. Forlivese (44.16 12.18)	CSN (44.14 12.24)	5.6	IV-V
			FOR (44.20 12.04)	11.8	IV-V
			MRS (43.94 12.18)	24.2	V
2003-12-30 (4.5)	05:31:38	SubApp. Dauno (41.64 14.85)	ARI (41.15 15.09)	57.6	II
			CMM (41.87 14.45) [1]	41.7	II
			CNM (41.62 15.10)	21.4	V
			PSC (41.81 13.79)	88.4	IV
			SCM (41.71 14.98)	13.7	V-VI
2004-02-23 (3.6*)	19:48:45	Vallo di Diano (40.70 15.43)	LVN (40.78 15.30)	14.4	IV-V
2004-07-12 (5.2)	13:04:04	Alpi Carniche (46.32 13.66)	CNV (45.97 12.45)	97.8	III
			MAJ (46.18 13.07)	44.7	IV-V
			PSR (45.95 13.01)	55.9	IV-V
			MOGG (46.18 13.07)	35.7	V
			GEPF (46.18 13.07)	38.5	IV-V
			CESC (46.18 13.07)	44.5	IV-V
2004-09-03 (4.1*)	00:04:12	Potentino (40.70 15.68)	MRV (40.36 15.83) [1]	39.7	II
2004-11-24 (5.3)	22:59:38	Garda Lake (45.69 10.52)	GVD (45.61 10.38) [1]	14.4	V-VI
2004-12-09 (4.0*)	02:44:25	Zona Ascolana (42.79 13.79)	AMT (42.63 13.29)	44.7	II
			CHT (42.37 14.15)	55.0	II
			CPS (42.27 13.76)	57.6	II
			NOR (42.79 13.09)	56.9	II
2004-12-30 (3.9*)	04:04:50	Zona Ragusa (36.88 14.68)	CDI (37.49 14.64)	68.7	II
			ISI (36.80 14.89)	20.8	IV-V
			LNT (37.29 15.00)	53.9	II
			PLZ (37.07 14.91)	29.5	IV
			RGS (36.92 14.70)	5.7	IV-V
			SCR (36.83 14.53)	14.6	IV-V
2008-12-23 (5.1)	15:24:22	Parma (44.54 10.35)	PNM (44.38 9.88)	41.9	V

2009-04-06 (5.8)	01:32:40	L'Aquila (42.34 13.38)	AQA (42.38 13.34)	4.7	VI
			AQG (42.37 13.34)	4.0	VI
			AQK (42.34 13.4)	2.0	VII-IX
			AQM (42.38 13.35)	4.1	VI
			AQV (42.38 13.34)	4.7	VI
			CLN (42.09 13.52)	31.1	V-VI
			MTR (42.52 13.25)	21.7	V-VI
2009-12-15(4.2)	13:11:59	Valle del Tevere (43.01 12.27)	ARO (43.47 11.88)	61.1	III
2009-12-19(4.4)	09:01:15	Etna (37.78 14.97)	BNT (37.78 14.84)	4.3	IV-V
			NVR1 (38.01 15.13)	31.9	III
			TAO (37.85 15.29)	37.2	III-IV
			TOR (38.04 14.81)	26.1	III
2010-08-16(4.7)	12:54:48	Isole Eolie (38.35 14.89)	LPR (38.47 14.96)	10.0	V-VI
			MZZ (38.23 15.25)	29.6	V
2011-06-23(4.7)	22:02:47	Monti Nebrodi (38.06 14.78)	GSM (38.17 14.9)	16.1	IV
			MZZ (38.23 15.25)	45.3	III-IV
			NSA (37.77 14.41)	45.8	IV
			PTT (38.13 14.98)	19.2	IV
			SVN (37.68 15.13)	52.2	III-IV
			TOR (38.04 14.81)	3.4	V-VI
2011-07-17(4.8)	18:30:28	Pianura lombardo-veneta (45.01 11.37)	MDN(44.65 10.89)	56.2	III
			MRN(44.88 11.06)	29.8	IV
2011-07-25(4.7)	12:31.21	Torinese (45.02 7.37)	PNR (44.88 7.34)	14.6	V
			SLZ (44.65 7.5)	41.3	IV
			SUS (45.14 7.04)	29.7	IV
2012-01-25(4.9)	08:06:37	Pianura Emiliana (44.87 10.51)	BRR (44.51 9.99)	57.4	IV-V
			NVL (44.84 10.73)	17.7	IV-V
			SRP (44.85 10.45)	5.2	VI
2012-10-25(5.3)	23:05:25	Pollino (39.87 16.02)	0FRA (39.83 16.27)	22.7	IV
			0LAI (39.95 15.97)	9.5	V-VI
			0MOR (39.86 16.09)	6.9	VI
			0ORS (39.8 15.91)	11.6	V
			0PAP (39.88 15.91)	8.6	V
			CLS7 (40.01 15.97)	15.9	V
			CVL (39.82 16.19)	16.4	V
			LRS (40.05 15.83)	25.2	IV-V
			LTR (40.09 16.01)	24.4	IV
			MRM (39.88 15.99)	2.0	VI
			ORS (39.8 15.91)	11.6	V
			ROT7 (39.95 16.04)	9.2	VI
			SDN (39.71 16.05)	18.1	IV
			VGG (39.97 16.05)	11.6	V-VI
			VGG7 (39.97 16.09)	13.0	V
2013-01-04(4.4)	07:50:06	Monti Nebrodi (37.88 14.72)	BNT (37.78 14.84)	14.7	V
			TOR (38.04 14.81)	19.1	IV

2013-06-21(5.4)	10:33:56	Lunigiana (44.13 10.14)	FVZ (44.24 10.13)	11.2	V-VI
			PZS (44.19 10.29)	15.4	V
2013-08-15(4.7)	23:06:51	Monti Peloritani (38.11 14.91)	BNT (37.78 14.84)	44.4	IV
			MZZ (38.23 15.25)	37.2	IV
			NAS (38.12 14.79)	7.5	V-VI
			PTT (38.13 14.98)	14.3	IV
2013-12-29(5.1)	17:08:43	Matese (41.40 14.43)	BOI (41.48 14.47)	10.1	VI
			BOJ (41.48 14.47)	10.1	VI
			GSN (41.31 14.45)	8.9	VI
			PDM (41.36 14.39)	6.0	VI
			SPT1 (41.46 14.49)	8.5	V
2015-08-29(3.9)	18:47:04	Slovenia (46.25 13.64)	GORI (45.94 13.63)	42.3	IV
2016-02-08(4.6)	15:35:44	Ibleo (37.00 14.80)	CLG1 (37.22 14.52)	40.3	III
			ISI (36.8 14.89)	20.2	IV
			LNT (37.29 15)	36.6	IV
			NSC (37.15 14.39)	45.9	III
			NTE (36.91 15.07)	20.3	IV
			PLZ (37.07 14.91)	10.9	IV-V
			SCR (36.83 14.53)	33.8	IV
			SRC (37.09 15.29)	40.2	IV
			SVN (37.68 15.13)	81.3	III
			VTT (36.95 14.52)	30.5	IV
			VZZ (37.16 14.76)	21.9	IV
2016-08-24(6.3)	01:36:32	Amatrice (42.70 13.23)	NOR (42.79 13.09)	15.8	V-VI
			MTR (42.52 13.25)	20.0	V-VI
			SNO (43.04 13.3)	38.1	V-VI
			PCB (42.56 13.34)	17.6	V
			LSS (42.56 12.97)	27.1	V
			PZI1 (42.44 13.33)	29.8	V
			AMT (42.63 13.29)	8.8	X

Table A-2: Weighted arithmetic mean values (μ_k^*) of the logarithm in base 10 of each GMP, for each intensity class from II to X. Standard deviations associated with the logarithm in base 10 of each GMP are also reported, common to all intensity classes (σ_{CSD}).

|		μ_{II}^*	μ_{III}^*	μ_{IV}^*	μ_V^*	μ_{VI}^*	μ_{VII}^*	μ_{VIII}^*	μ_{IX}^*	μ_X^*	σ_{CSD}
GMP	PGD (cm)	-2.29	-1.70	-1.41	-0.76	-0.23	0.16	0.38	1.06	0.90	0.61
	PGV (cm/s)	-1.33	-0.71	-0.42	0.20	0.62	0.88	1.12	1.55	1.64	0.50
	PGA (cm/s^2)	-0.01	0.59	0.94	1.56	1.88	2.01	2.31	2.56	2.93	0.52
	I_A (cm/s)	-2.83	-1.92	-1.23	-0.25	0.77	1.40	2.06	2.06	2.06	0.96
	I_H (cm)	-0.77	-0.26	0.08	0.62	1.32	1.37	2.18	2.18	2.15	0.57
	PSA03 (cm/s^2)	0.28	0.89	1.14	1.70	1.70	2.77	2.77	2.77	3.25	0.56
	PSA10 (cm/s^2)	-0.38	0.07	0.39	0.95	1.71	1.69	2.58	2.58	2.59	0.62
	PSA30 (cm/s^2)	-1.25	-0.80	-0.48	0.16	0.83	0.71	1.91	1.91	1.75	0.67

Figures
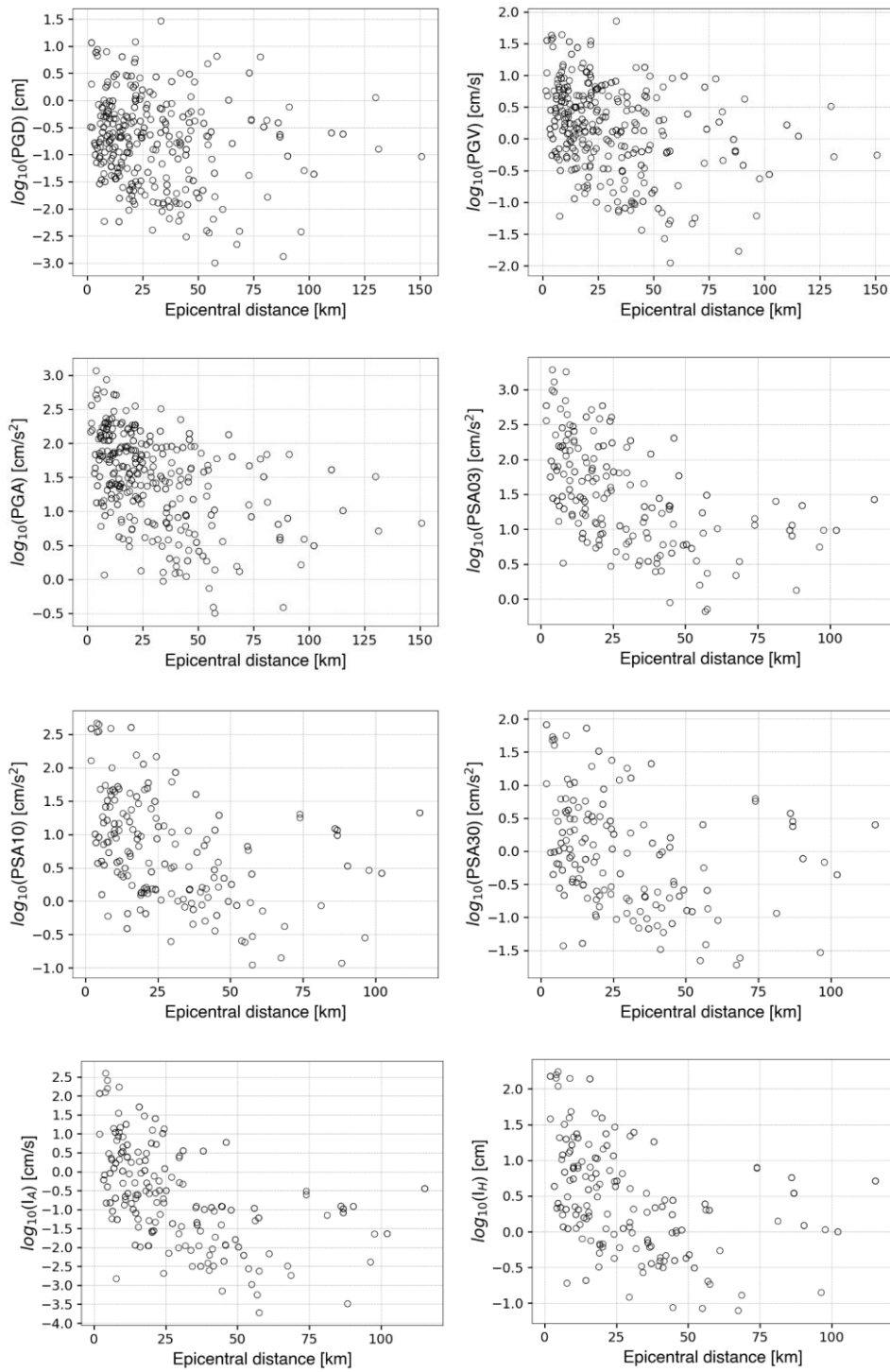
Figure A-1: Epicentral distance coverage of the GMP-intensity couples.

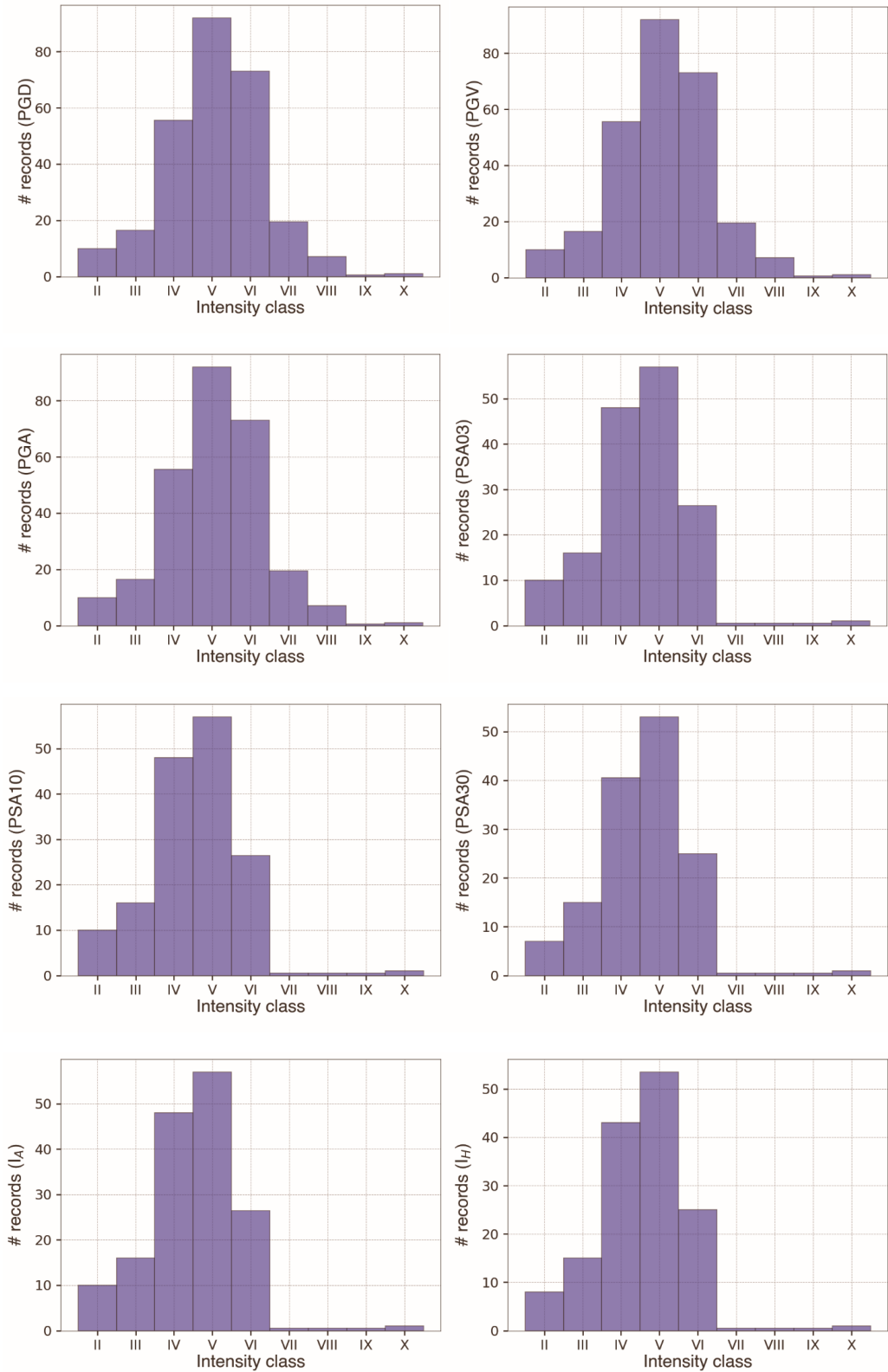Figure A-2: Distribution of the different GMP values binned into classes at integer intensity intervals.
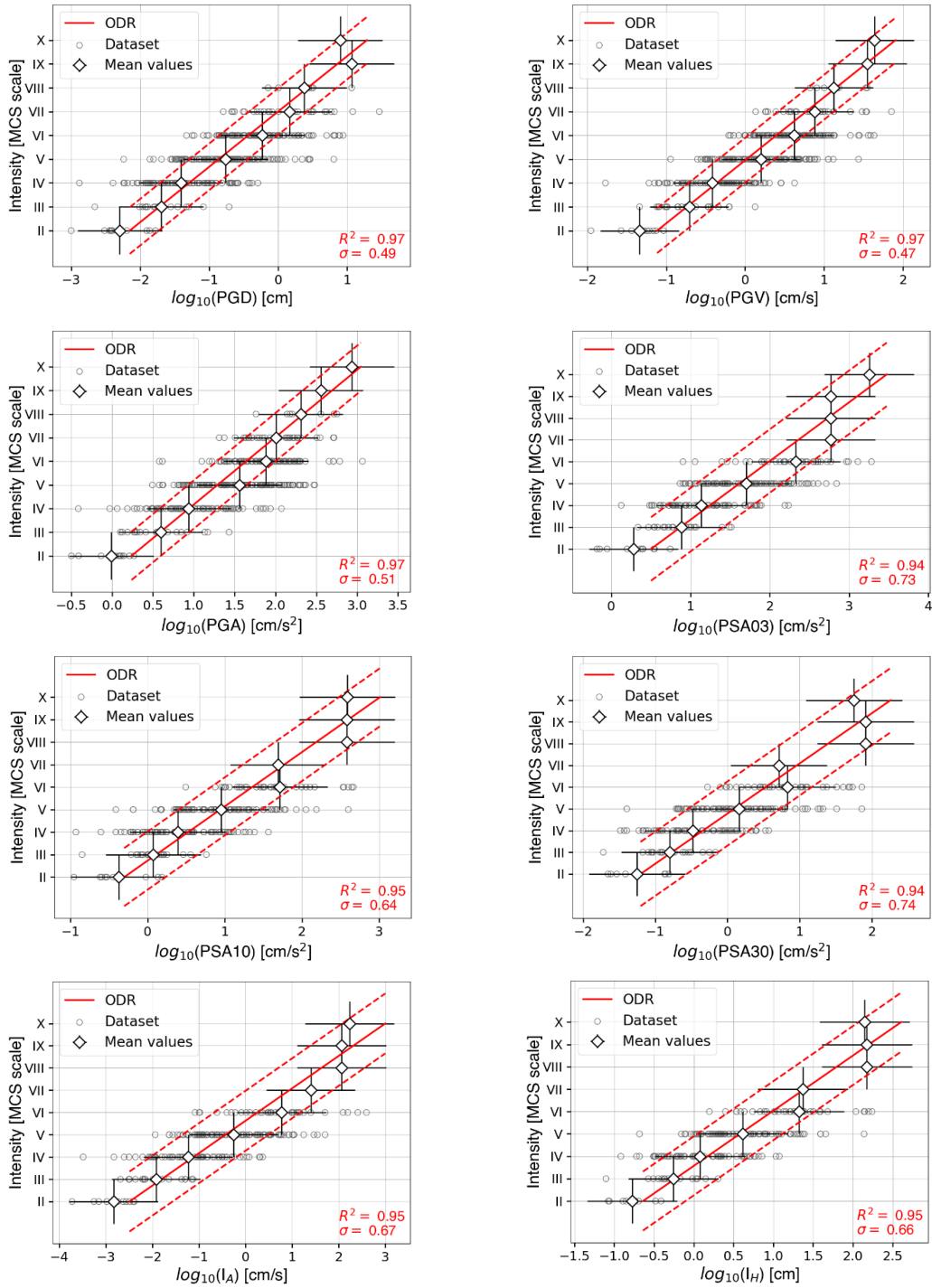
Figure A-3: GMICE estimations for each considered GMP class. The grey circles represent the underlying dataset, the white diamonds are the weighted arithmetic means for each class of values with the associated standard deviations; the red line is the ODR GMICE, with associated ±2σ error (dotted red line).

References

Agresti, A. (2013). *Categorical Data Analysis* (3rd ed.). Hoboken, NJ: John Wiley & Sons.

Aki, K. (1993). Local site effects on weak and strong ground motion. *Tectonophysics*, *218*, 93-111. https://doi.org/10.1016/0040-1951(93)90262-I

Anderson, J.G., and Hough, S.E. (1984). A model for the shape of the fourier amplitude spectrum of acceleration at high frequencies. *B. Seismol. Soc. Am.*, *74 (5)*, 1969–1993.

Andrews, D.J. (1986). Objective determination of source parameters and similarity of earthquakes of different size. *In: Das, S., Boatwright, J., and Scholz, C.H., Eds., Earthquake Source Mechanics* (American Geophysical Union, Washington DC).

Aoudia, A., Saraò, A., Bukchin, B., and Suhadolc, P. (2000). The 1976 Friuli (NE Italy) thrust faulting earthquake: A reappraisal 23 years later. *Geophys. Res. Lett., 27*, 573-576. https://doi.org/10.1029/1999GL011071

Aristizabal, C., Bard, P.Y., and Beauval, C. (2017). *Site-specific PSHA: combined effects of single station sigma, host-o-target adjustments and non-linear behavior* [Paper presentation]. 16th World Conference on Earthquake Engineering (WCEE) Conference 2017, Santiago, Chile.

Azzaro, R., D'Amico, S., Mostaccio, A., Scarfì, L., and Tuvè, T. (2016). *RILIEVO MACROSISMICO DEL TERREMOTO IBLEO DELL'8 FEBBRAIO 2016*. INGV dell'Osservatorio Etneo – Sezione di Catania. http://www.questingv.it/index.php/rilievi-macrosismici/28-ibleo-08-02-2016-ml-4-6/file

Bajc, J., Aoudia, A., Saraò, A., and Suhadolc, P. (2001). The 1998 Bovec-Krn Mountain (Slovenia) Earthquake Sequence. *Geophys. Res. Lett., 28*, 1839-1842. https://doi.org/10.1029/2000GL011973

Bianchi, I., Behm, M., Rumpfhuber, E.M., and Bokelmann, G. (2014). A New Seismic Data Set on the Depth of the Moho in the Alps. *Pure Appl. Geophys., 172,* 295–308. https://doi.org/10.1007/s00024-014-0953-1

Bindi, D., Cotton, F., Kotha, S.R., Bosse, C., Stromeyer, D., and Grünthal, G. (2017). Application-driven ground motion prediction equation for seismic hazard assessments in non-cratonic moderate-seismicity areas. *J. Seismol.*, *21*, 1201–1218. https://doi.org/10.1007/s10950-017-9661-5

Boatwright, J. (1978). Detailed spectral analysis of two small New York State earthquakes. *B. Seismol. Soc. Am.*, *68*, 1117–1131.

Boggs, P.T., Spiegelman, C.H., Donaldson, J.R., and Schnabel, R.B. (1988). A computational examination of orthogonal distance regression. *J. Econometrics*, *38*, 169–201. https://doi.org/10.1016/0304-4076(88)90032-2

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

Boore, D.M. (1983). Stochastic Simulation of high frequency ground motions based on seismological models of the radiated spectra. *B. Seismol. Soc. Am.*, *73(6)*, 1865–1894.

Boore, D.M. (2003). Simulation of ground motion using the stochastic method. *Pure Appl. Geophys.*, *160(3),* 635–676. https://doi.org/10.1007/PL00012553

Boore, D.M., and Boatwright, J. (1984). Average body-wave radiation coefficients. *B. Seismol. Soc. Am.*, *74*, 1615–1621.

Bora, S.S., Cotton, F., Scherbaum, F., Edwards, B., and Traversa, P. (2017). Stochastic source, path and site attenuation parameters and associated variabilities for shallow crustal European earthquakes. *B. Earthq. Eng.*, *15*, 4531–4561. https://doi.org/10.1007/s10518-017-0167-x

Borcherdt, R.D., and Gibbs, J.F. (1976). Effects of local geological conditions in the San Francisco Bay region on ground motions and the intensities of the 1906 earthquake. *B. Seismol. Soc. Am.*, *66 (2),* 467–500.

Bragato, P.L., Sugan, M., Augliera, P., Massa, M., Vuan, A., and Saraò, A. (2011). Moho reflection effects in the Po plain (northern Italy) observed from instrumental and intensity data. *B. Seismol. Soc. Am.*, *101(5),* 2142–2152. https://doi.org/10.1785/0120100257

Bragato, P.L., *et al.* (2014). The Central and Eastern European Earthquake Research Network-CE3RN. *Geophys. Res. Abstracts*, *16*, 13911. https://doi.org/10.13140/RG.2.1.3507.7843

Bressan, G., Barnaba, C., Bragato, P.L., Peresan, A., Rossi, G., and Urban, S. (2019). Seismic districts of Friuli Venezia Giulia. *B. Geofis. Teor. Appl.*, *60*, S1-S74.

Bressan, G., Bragato, P.L., and Venturini, C. (2003). Stress and strain tensors based on focal mechanisms in the seismotectonic framework of the Friuli-Venezia Giulia region (northeastern Italy). *B. Seismol. Soc. Am.*, *93*, 1280-1290. https://doi.org/10.1785/0120020058

Bressan, G., Ponton, M., Rossi, G., and Urban, S. (2016). Spatial organization of seismicity and fracture pattern in NE Italy and W Slovenia. *J. Seismol., 20*, 511–534. https://doi.org/10.1007/s10950-015-9541-9

Bressan, G., Snidarcig, A., and Venturini, C. (1998). Present state of tectonic stress of the Friuli area (eastern Southern Alps). *Tectonophysics*, *292(3–4),* 211-227. https://doi.org/10.1016/S0040-1951(98)00065-1

Brune, J.N. (1970). Tectonic stress and the spectra of seismic shear waves from earthquakes. *J. Geophys. Res*., *75*, 4997–5010. https://doi.org/10.1029/JB075i026p04997

Brune, J.N. (1971). Correction: Tectonic stress and the spectra of seismic shear waves from earthquakes. *J. Geophys. Res*., *76*, 5002. https://doi.org/10.1029/JB076i020p05002

Caprio, M., Tarigan, B., Worden, C.B., Wiemer, S., and Wald, D.J. (2015). Ground Motion to Intensity Conversion Equations (GMICEs): A Global Relationship and Evaluation of Regional Dependency. *B. Seismol. Soc. Am.*, 105 (3), 1476–1490. https://doi.org/10.1785/0120140286

Castro, R.R., Anderson, J.G., and Singh, S. K. (1990). Site response, attenuation and source spectra of S waves along the Guerrero, Mexico, subduction zone. *B. Seismol. Soc. Am.*, 80, 1481–1503.

Castro, R.R., Pacor, F., and Petrungaro, C. (1997). Determination of S-wave energy release of earthquakes in the region of Friuli, Italy. *Geophys. J. Int*., 128, 399–408. https://doi.org/10.1111/j.1365-246X.1997.tb01563.x

Castro, R.R., Pacor, F., Sala, A., and Petrungaro, C. (1996). *S* wave attenuation and site effects in the region of Friuli, Italy. *J. Geophys. Res.*, 101(B10), 22355–22369. https://doi.org/10.1029/96JB02295

Cataldi, L., Tiberi, L., and Costa, G. Estimation of MCS intensity for Italy from high quality accelerometric data, using GMICEs and Gaussian Naïve Bayes Classifiers. *B. Earthq. Eng.* (2021). https://doi.org/10.1007/s10518-021-01064-6

Chiaruttini, C., and Siro, L. (1981). The correlation of peak ground horizontal acceleration with magnitude, distance, and seismic intensity for Friuli and Ancona, Italy, and the Alpide Belt. *B. Seismol. Soc. Am.*, *71(6),* 1993–2009.

Console, R., and Rovelli, A. (1981). Attenuation parameters for Friuli region from strong-motion accelerogram spectra. *B. Seismol. Soc. Am*., *71(6)*, 1981–1991.

Costa, G., *et al.* (2014). *The Italian Strong Motion Network (RAN), near-real time data acquisition and data analysis: a useful tool for seismic risk mitigation* [paper presentation]. 2nd European Conference on Earthquake Engineering and Seismology (2ECCES) Conference 2014, Istanbul, Turkey. http://www.eaee.org/Media/Default/2ECCES/2ecces_ss/1858.pdf

Costa, G., Moratto, L., and Suhadolc, P. (2010). The Friuli Venezia Giulia Accelerometric Network: RAF. *B. Earthq. Eng.*, *8*, 1141–1157. https://doi.org/10.1007/s10518-009-9157-y

Costa, G., Suhadolc, P, Delise, A., Moratto, L., Furlanetto, E., and Fitzko, F. (2006). *Estimation of site effects at some stations of the Friuli (NE Italy) Accelerometric Network (RAF)* [paper presentation]. 3rd International Symposium on the Effects of Surface Geology on Seismic Motion 2006, Grenoble, France.

Cuffaro, M., *et al.* (2010). On the geodynamics of the northern Adriatic plate. *Rend. Fis. Acc. Lincei*, *21(1).* https://doi.org/10.1007/s12210-010-0098-9

D'Amico, M., et al. (2020). Italian Accelerometric Archive v 3.1 - Istituto Nazionale di Geofisica e Vulcanologia, Dipartimento della Protezione Civile Nazionale. https://doi.org/10.13127/itaca.3.1

DHA/UNDRO - DMTP - UNDP (1994). Disaster Preparedness (2nd Ed.), 66 p.

Di Bona, M., and Rovelli, A. (1988). Effects of the bandwidth limitation of stress drops estimated from integrals of the ground motion. *B. Seismol. Soc. Am.*, *78(5),* 1818–1825.

DISS Working Group (2018). *Database of Individual Seismogenic Sources (DISS), Version 3.2.1: A compilation of potential sources for earthquakes larger than M 5.5 in Italy and surrounding areas. http://diss.rm.ingv.it/diss/,* Istituto Nazionale di Geofisica e Vulcanologia. https://doi.org/10.6092/INGV.IT-DISS3.2.1

Drouet, S., Chevrot, S., Cotton, F., and Souriau, A. (2008). Simultaneous inversion of source spectra, attenuation parameters, and site responses: application to the data of the French Accelerometric Network. *B. Seismol. Soc. Am.*, *98,* 198-219. https://doi.org/10.1785/0120060215

Drouet, S., Triantafyllidis, P., Savvaidis, A., and Theodoulidis, N. (2008). Comparison of Site-Effects Estimation Methods Using the Lefkas, Greece, 2003 Earthquake Aftershocks. *B. Seismol. Soc. Am.*, *98,* 2349-2363. https://doi.org/10.1785/0120080004

Duval, A.-M., Bard, P.Y., Lebrun, B., Lacave-Lachet, C., Riepl., J., and Hatzfeld, D. (2001). H/V technique for site response analysis. Synthesis of data from various surveys. *B. Geofis. Teor. Appl.*, *42(3–4),* 267–280.

EC8 (2004). *Eurocode 8: Design of structures for earthquake resistance. Part 1: General rules, seismic actions and rules for buildings*. Brussels, Belgium: European Norm, European Committee for Standardisation, European Committee for Standardisation Central Secretariat.

Edwards, B., Allmann, B., Fäh, D. and Clinton, J. (2010). Automatic computation of moment magnitudes for small earthquakes and the scaling of local to moment magnitude. *Geophys. J. Int., 183(1),* 407–420. https://doi.org/10.1111/j.1365-246X.2010.04743.x

Edwards, B., and Fäh, D. (2013). Measurements of stress parameter and site attenuation from recordings of moderate to large earthquakes in Europe and the Middle East. *Geophys. J. Int., 194(2),* 1190–1202. https://doi.org/10.1093/gji/ggt158

Edwards, B., Fäh, D., and Giardini, D. (2011). Attenuation of seismic shear wave energy in Switzerland. *Geophys. J. Int., 185(2),* 967–984. https://doi.org/10.1111/j.1365-246X.2011.04987.x

Edwards, B., Michel, C., Poggi, V., and Fäh, D. (2013). Determination of site amplification from regional seismicity: Application to the Swiss National Seismic Networks. *Seismol. Res. Lett*., *84(4),* 611–621. https://doi.org/10.1785/0220120176

Edwards, B., Rietbrock, A., Bommer, J.J., and Baptie, B. (2008). The Acquisition of Source, Path, and Site Effects from Microearthquake Recordings Using *Q* Tomography: Application to the United Kingdom. *B. Seismol. Soc. Am., 98(4)*, 1915–1935. https://doi.org/10.1785/0120070127

Ertuncay, D., Cataldi, L., and Costa, G. (2021). Web-based macroseismic intensity study in Turkey: entries in Ekşi Sözlük. *Geosci. Commun., 4,* 69–81. https://doi.org/10.5194/gc-4-69-2021

Faccioli, E., and Cauzzi, C. (2006). Macroseismic intensities for seismic scenarios estimated from instrumentally based correlations. In: (s.n.), *First European Conference on Earthquake Engineering and Seismology 2006 (1st ECEES): Joint Event of the 13th European Conference on Earthquake Engineering and the 30th General Assembly of the European Seismological Commission*. SGEB Press.

Faccioli, E., Vanini, M., and Frassine, L. (2002). *"Complex" Site Effects in Earthquake Ground Motion, including Topography* [paper presentation]. 12th European Conference on Earthquake Engineering 2002, London, UK.

Faenza, L., and Michelini, A. (2010). Regression analysis of MCS intensity and ground motion parameters in Italy and its application in ShakeMap. *Geophys. J. Int., 180(3)*, 1138–1152. https://doi.org/10.1111/j.1365-246X.2009.04467.x

Faenza, L., and Michelini, A. (2011). Regression analysis of MCS intensity and ground motion spectral accelerations (SAs) in Italy. *Geophys. J. Int*., 186, 1415–1430. https://doi.org/10.1111/j.1365-246X.2011.05125.x

Ferrari, G., and Guidoboni, E. (2009). Seismic scenarios and assessment of intensity: Some criteria for the use of the MCS scale. *Ann. Geophys-Italy, 43*. https://doi.org/10.4401/ag-3664

Field, E.H., and Jacob, K.H. (1995). A Comparison and Test of Various Site-Response Estimation Techniques, including Three That Are Not Reference-Site Dependent. *B. Seismol. Soc. Am.*, *85*, 1127-1143.

Fitzko, F., Costa, G. Delise, A., and Suhadolc, P. (2007). Site Effects Analyses in the Old City Center of Trieste (NE Italy) Using Accelerometric Data. *J. Earthq. Eng.*, *11*, 33-48. https://doi.org/10.1080/13632460601123123

Franceschina, G., Kravanja, S., and Bressan, G. (2006). Source parameters and scaling relationships in the Friuli-Venezia Giulia (Northeastern Italy) region. *Phys. Earth Planet. In.*, *154(2),* 48-167. https://doi.org/10.1016/j.pepi.2005.09.004

Gail, M., Atkinson, R., and Mereu, F. (1992). The shape of ground motion attenuation curves in southeastern Canada. *B. Seismol. Soc. Am.*, *82(5),* 2014–2031.

Galli, P., Peronace, E., and Tertulliani, A. (2016). *Rapporto sugli effetti macrosismici del terremoto del 24 Agosto 2016 di Amatrice in scala MCS*. DPC, CNR-IGAG, INGV. https://doi.org/10.5281/zenodo.161323

Gallo, A., Costa, G., and Suhadolc, P. (2014). Near real-time automatic moment magnitude estimation. *B. Earthq. Eng.*, *12*, 185–202. https://doi.org/10.1007/s10518-013-9565-x

Gentili, S., and Franceschina, G. (2011). High frequency attenuation of shear waves in the southeastern Alps and northern Dinarides. *Geophys. J. Int., 185(3),* 1393–1416. https://doi.org/10.1111/j.1365-246X.2011.05016.x

Gómez Capera, A.A., D'Amico, M., Lanzano, G., Locati, M., and Santulin, M. (2020). Relationships between ground motion parameters and macroseismic intensity for Italy. *B. Earthq. Eng.*, *18*, 5143–5164. https://doi.org/10.1007/s10518-020-00905-0

Gorini, A., *et al.* (2010). The Italian strong motion network. *B. Earthq. Eng.*, *8*, 1075–1090. https://doi.org/10.1007/s10518-009-9141-6

Gosar, A. (2014). Analysis of the Impact of Fault Mechanism Radiation Patterns on Macroseismic Fields in the Epicentral Area of 1998 and 2004 Krn Mountains Earthquakes (NW Slovenia). *The Scientific World Journal,* 2014, 206843. https://doi.org/10.1155/2014/206843

Graziani, L., Del Mese, S., Tertulliani, A., Arcoraci, L., Maramai, A., and Rossi, A. (2019). Investigation on damage progression during the 2016–2017 seismic sequence in Central Italy using the European Macroseismic Scale (EMS-98). *B. Earthq. Eng.*, *17*, 5535–5558. https://doi.org/10.1007/s10518-019-00645-w

Grimaz, S., and Malisan, P. (2017). How could cumulative damage affect the macroseismic assessment?. *B. Earthq. Eng.*, *15*, 2465–2481. https://doi.org/10.1007/s10518-016-0016-3

Grünthal, G. (ed.) (1998). *European Macroseismic Scale 1998 (EMS-98)*. Luxembourg: Cahiers du Centre Européen de Géodynamique et de Séismologie 15, Centre Européen de Géodynamique et de Séismologie.

Grünthal, G., Wahlström, R., and Stromeyer, D. (2009). The unified catalogue of earthquakes in central, northern, and northwestern Europe (CENEC) - Updated and expanded to the last millennium. *J. Seismol., 13*, 517-541. https://doi.org/10.1007/s10950-008-9144-9

Guidoboni, E., and Stucchi, M. (1993). The contribution of historical records of earthquakes to the evaluation of seismic hazard. *Ann. Geophys-Italy, 36*. https://doi.org/10.4401/ag-4264

Gutenberg, B., and Richter, C.F. (1942). Earthquake magnitude, intensity, energy and acceleration. *B. Seismol. Soc. Am.*, *32*, 163–191.

Hanks, T.C., and Kanamori, H. (1979). A moment magnitude scale. *J. Geophys. Res.*, *84(B5),* 2348– 2350. https://doi.org/10.1029/JB084iB05p02348

International Seismological Centre (2020). *On-line Bulletin.* https://doi.org/10.31905/D808B830

Konno, K., and Ohmachi, T. (1998). Ground-motion characteristics estimated from spectral ratio between horizontal and vertical components of microtremor. *B. Seismol. Soc. Am.*, *88(1),* 228–241.

Kotha, S.R., Weatherill, G., Bindi, D., and Cotton, F. (2020). A regionally-adaptable ground-motion model for shallow crustal earthquakes in Europe. *B. Earthq. Eng.*, *18*, 4091–4125. https://doi.org/10.1007/s10518-020-00869-1

Kraft, D. (1988). *A software package for sequential quadratic programming* (DFVLR-FB 88-28). DLR German Aerospace Center – Institute for Flight Mechanics.

Kuehn, N.M., and Scherbaum, F. (2010). A Naive Bayes Classifier for Intensities Using Peak Ground Velocity and Acceleration. *B. Seismol. Soc. Am.*, *6,* 3278–3283. http://doi.org/10.1785/0120100082

Lagomarsino, S., and Giovinazzi, S. (2006). Macroseismic and mechanical models for the vulnerability assessment of current buildings. *B. Earthq. Eng.*, *4*, 415-443. https://doi.org/10.1007/s10518-006-9024-z

Lancieri, M., Renault, M., Berge-Thierry, C., Gueguen, P., Baumont, D., and Perrault, M. (2015). Strategy for the selection of input ground motion for inelastic structural response analysis based on naïve Bayesian classifier. *B. Earthq. Eng.*, *13*. https://doi.org/10.1007/s10518-015-9728-z

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

Lesueur, C., Cara, M., Scotti, O., Schlupp, A., and Sira, C. (2013). Linking ground motion measurements and macroseismic observations in France: a case study based on accelerometric and macroseismic databases. *J. Seismol., 17(2)*, 313-333. https://doi.org/10.1007/s10950-012-9319-2

Locati, M., *et al.* (2016). *DBMI15, the 2015 version of the Italian Macroseismic Database.* Istituto Nazionale di Geofisica e Vulcanologia. http://doi.org/10.6092/INGV.IT-DBMI15

Locati, M., *et al.* (2019). *Database Macrosismico Italiano (DBMI15), versione 2.0.* Istituto Nazionale di Geofisica e Vulcanologia (INGV). https://doi.org/10.13127 /DBMI/DBMI1 5.2

Luzi, L., Hailemikael, S., Bindi, D., Pacor, F., Mele, F., and Sabetta, F. (2008). Itaca (Italian Accelerometric Archive): A web portal for the dissemination of Italian strong-motion data. *Seism. Res. Lett.*, *79(5),* 716–722. https://doi.org/10.1785/gssrl.79.5.716

Malagnini, L., Akinci, A., Herrmann, R., Pino, N., and Scognamiglio, L. (2002). Characteristics of the Ground Motion in Northeastern Italy. *B. Seismol. Soc. Am.*, *92*, 2186-2204. https://doi.org/10.1785/0120010219

Mantovani, E., *et al.* (2006). Post-Late Miocene Kinematics of the Adria Microplate: Inferences from Geological, Geophysical and Geodetic data. In: Pinter, N., Grenerczy, G., Weber, J., Stein, S., and Medak, D. (Eds.), *The Adria Microplate: GPS Geodesy, Tectonics and Hazard*, *61(NATO Science Series IV-Earth and Environmental Sciences)*. Springer.

Margottini, C., Molin, D., and Serva, L. (1992). Intensity versus ground motion: A new approach using Italian data. *Eng. Geol.*, *33(1),* 45-58. https://doi.org/10.1016/0013-7952(92)90034-V

MedNet Project Partner Institutions (1990). *Mediterranean Very Broadband Seismographic Network (MedNet)*. Istituto Nazionale di Geofisica e Vulcanologia (INGV). https://doi.org/10.13127/SD/fBBBtDtd6q

Medvedev, S., Sponheuer, W., and Karník, V. (1964). Neue seismische Skala—Intensity scale of earthquakes, *Proceedings 7 Tagung der Europäischen Seismologischen Kommission*, Jena, GDR, 1962, Berlin: Veröff. Inst. Bodendyn. Erdbebenforsch., 1964, vol. 77, pp. 69–76.

Meletti, C., and Valensise, G. (2004). *Zonazione sismogenetica ZS9–App. 2 al rapporto conclusivo.* Gruppo di Lavoro MPS (2004). Redazione della mappa di pericolosità sismica prevista dall'Ordinanza PCM 3274 (2004). http://zonesismiche.mi.ingv.it/documenti/App2.pdf

Michel, C., *et al.* (2014). Assessment of site effects in Alpine regions through systematic site characterization of seismic stations. *B. Seismol. Soc. Am.*, *104(6),* 2809–2826. https://doi.org/10.1785/0120140097

Michelini, A., Faenza, L., Luciani, V., and Malagnini, L. (2008). ShakeMap implementation in Italy. *Seismol. Res. Lett.*, 79(5), 688–697. https://doi.org/10.1785/gssrl.79.5.688

Moratalla, J.M., Goded, T., Rhoades, D., Canessa, S., and Gerstenberger, M. (2021). New Ground Motion to Intensity Conversion Equations (GMICEs) for

New Zealand. *Seismol. Res. Lett*., *92*, 448-459. https://doi.org/10.1785/0220200156

Moratto, L., Costa, G., and Suhadolc, P. (2009). Real-time generation of Shake Maps in the Southeastern Alps. *B. Seismol. Soc. Am., 99(4),* 2489–2501. https://doi.org/10.1785/0120080283

Morozov, I.B. (2008). Geometrical attenuation, frequency dependence of Q, and the absorption band problem. *Geophys. J. Int., 175*, 239–252. https://doi.org/10.1111/j.1365-246X.2008.03888.x

Morozov, I.B. (2009). Thirty years of confusion around ''scattering Q''?. *Seismol. Res. Lett*., *80*, 5–7. https://doi.org/10.1785/gssrl.80.1.5

Munafò, I., Malagnini, L., and Chiaraluce, L. (2016). On the Relationship between Mw and ML for Small Earthquakes. *B. Seismol. Soc. Am.*, *106*. https://doi.org/10.1785/0120160130

Murphy, J.R., and O'Brien, L.J. (1977). The correlation of peak ground acceleration amplitude with seismic intensity and other physical variables. *B. Seismol. Soc. Am.*, *67(3),* 877–915.

Musson, R.M.W., Grünthal, G., and Stucchi, M. (2010). The comparison of macroseismic intensity scales. *J. Seismol., 14*, 413–428. https://doi.org/10.1007/s10950-009-9172-0

Nakamura, Y. (1989). A Method for Dynamic Characteristics Estimation of Subsurface using Microtremor on the Ground Surface. In: (s.n.), *Railway Technical Research Institute, Quarterly Reports* (30(1), pp. 25-33). Railway Technical Research Institute Press.

Nocedal, J., and Wright, S. (2006). *Numerical Optimization*. Springer. https://doi.org/10.1007/978-0-387-40065-5

OGS (Istituto Nazionale Di Oceanografia E Di Geofisica Sperimentale) (2016). *North-East Italy Seismic Network*. International Federation of Digital Seismograph Networks. https://doi.org/10.7914/SN/OX

OGS (Istituto Nazionale Di Oceanografia E Di Geofisica Sperimentale) and University Of Trieste (2002). *North-East Italy Broadband Network*. International Federation of Digital Seismograph Networks. https://doi.org/10.7914/SN/NI

Oth, A., Bindi, D., Parolai, S., and 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. *B. Seismol. Soc. Am.*, *101*. https://doi.org/10.1785/0120100135

Oth, A., Bindi, D., Parolai, S., and Wenzel, F. (2008). S-wave attenuation characteristics beneath the Vrancea region in Romania: new insights from the inversion of groundmotion spectra. *B. Seismol. Soc. Am.*, *98*, 2482-497. https://doi.org/10.1785/0120080106

Pacor, F., *et al.* (2011). Overview of the Italian strong motion database ITACA 1.0. *B. Earthq. Eng.*, *9(6),* 1723–1739. https://doi.org/10.1007/s10518-011-9327-6

Pacor, F., *et al.* (2016). Spectral models for ground motion prediction in the L'Aquila region (central Italy): evidence for stress-drop dependence on magnitude and depth. *Geophys. J. Int*., 204(2), 697–718. https://doi.org/10.1093/gji/ggv448

Panza, G., Cazzaro, R., and Vaccari, F. (1997). Correlation between macroseismic intensities seismic ground motion parameters. *Ann. Geophys-Italy*, *40.* https://doi.org/10.4401/ag-3872

Parolai, S., Bindi, D., and Augliera, P. (2000). Application of the generalized inversion technique (GIT) to a microzonation study: numerical simulations and comparison with different site-estimation techniques. *B. Seismol. Soc. Am.*, *90*, 286-297. https://doi.org/10.1785/0119990041

Parolai, S., *et al.* (2004). Comparison of different site response estimation techniques using aftershocks of the 1999 Izmit earthquake. *B. Seismol. Soc. Am.*, *94*, 1096. https://doi.org/10.1785/0120030086

Virtanen, P., *et al.* (2020). SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nat. Methods*, *17(3),* 261-272. https://doi.org/10.1038/s41592-019-0686-2

Pedregosa, F., *et al.* (2011). Machine Learning in Python. *J. Mach. Learn. Res.*, *12*, 2825–2830.

Poggi, V., Edwards, B., and Fäh, D. (2011). Derivation of a Reference Shear-Wave Velocity Model from Empirical Site Amplification. *B. Seismol. Soc. Am.*, *101(1),* 258–274. https://doi.org/10.1785/0120100060

Poljak, M., Živčić, M., and Zupančič, P. (2000). The Seismotectonic Characteristics of Slovenia. *Pure Appl. Geophys.,* *157.* https://doi.org/10.1007/PL00001099

Pondrelli, M., *et al.* (2020). Geological map and stratigraphic evolution of the central sector of the Carnic Alps (Austria-Italy). *Ital. J. Geosci., 139(3),* 469–484. https://doi.org/10.3301/IJG.2020.16

Rice, J.R., and White, J. (1964). Norms for Smoothing and Estimation. *Siam Review*, *6*, 243-256.

Salazar, W., Sardina, V., and de Cortina, J. (2007). A hybrid inversion technique for the evaluation of source, path, and site effects employing S-wave spectra for subduction and upper-crustal earthquakes in El Salvador. *B. Seismol. Soc. Am.*, *97*, 208-221. https://doi.org/10.1785/0120060076

Scherbaum, F. (1990). Combined inversion for the three-dimensional Q structure and source parameters using microearthquake spectra. *J. Geophys. Res*., *95(B8)*, 12423–12438. https://doi.org/10.1029/JB095iB08p12423

Schreiber, J. (2018). Pomegranate: fast and flexible probabilistic modeling in python. *J. Mach. Learn. Res.*, *18(164),* 1–6.

Sieberg A. (1930). Geologie der Erdbeben. *Handbuch der Geophysik*, 2(4), 552-555.

Slovenian Environment Agency (2001). *Seismic Network of the Republic of Slovenia*. International Federation of Digital Seismograph Networks. https://doi.org/10.7914/SN/SL

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

Stover, C.W., and Coffman, J.L. (1993). *Seismicity of the United States, 1568-1989 (Revised)* (U.S. Geological Survey Professional Paper 1527, 417 pp.). Washington, D.C.: U.S. Government Printing Office.

Sugan, M., and Vuan, A. (2014). On the ability of Moho reflections to affect the ground motion in northeastern Italy: a case study of the 2012 Emilia seismic sequence. *B. Earthq. Eng.*, *12*, 2179–2194. https://doi.org/10.1007/s10518-013-9564-y

Tertulliani, A., and Azzaro, R. (eds.) (2016). *QUEST - Rilievo macrosismico in EMS98 per il terremoto di Amatrice del 24 agosto 2016.* INGV. http://www.questingv.it/index.php/rilievi-macrosismici/29-amatrice-24-08-2016-ml-6-0-rapporto-sul-rilievo-macrosismico-ems98/file

Tiberi, L., Costa, G., Jamšek Rupnik, P., Cecić, I., and Suhadolc, P. (2018). The 1895 Ljubljana earthquake: can the intensity data points discriminate which one of the nearby faults was the causative one?. *J. Seismol., 22*, 927–941. https://doi.org/10.1007/s10950-018-9743-z

Trifunac, M.D., and Brady, A.G. (1975). On the correlation of seismic intensity with peaks of recorded ground motion. *B. Seismol. Soc. Am.*, *65(1),* 139–162.

University Of Zagreb (2001). *Croatian Seismograph Network* [Data set]. International Federation of Digital Seismograph Networks. https://doi.org/10.7914/SN/CR

Wald, D.J., Lin, K.W., Quitoriano, V., and Worden, C. (2007). *USGS ShakeMap Developments, Implementation, and Derivative Tools* [paper presentation]. American Geophysical Union (AGU) Fall Meeting 1999, San Francisco, United States.

Wald, D.J., Quitoriano, V., Dengler, L., and Dewey, J. (1999a). Utilization of the Internet for Rapid Community Intensity Maps. *Seismol. Res. Lett*., *70*, 680-697. https://doi.org/10.1785/gssrl.70.6.680

Wald, D.J., Quitoriano, V., Heaton, T.H., and Kanamori, H. (1999b). Relationships between peak ground acceleration, peak ground velocity and

modified Mercalli intensity in California. *Earthq. Spectra*, *15*, 557–564. https://doi.org/10.1193/1.1586058

Wald, D.J., Quitoriano, V., Heaton, T., Kanamori, H., Scrivner, C.W., and Worden, C.B. (1999c). TriNet "ShakeMaps": Rapid Generation of Peak Ground-motion and Intensity Maps for Earthquakes in Southern California. *Earthq. Spectra, 15(3),* 537-556. https://doi.org/10.1193/1.1586057

Wang, Z. (2011). Seismic Hazard Assessment: Issues and Alternatives. *Pure Appl. Geophys.,168*, 11-25. https://doi.org/10.1007/s00024-010-0148-3

Wasserman, L. (2004). *All of statistics. A coincise course in statistical inference*. Springer.

Weginger, S., Jia, Y., Papi Isaba, M., and Horn, N. (2017). Real-time Shakemap implementation in Austria. In: (s.n.), *Geophysical Research Abstracts Vol. 19* (p.6624). EGU General Assembly 2017.

Wood, H.O., and Neumann, F. (1931). Modified Mercalli intensity scale of 1931. *B. Seismol. Soc. Am.*, 21 (4), 277–283.

Worden, C.B., Gerstenberger, M.C., Rhoades, D.A., and Wald, D.J. (2012). Probabilistic relationships between ground-motion parameters and modified Mercalli intensity in California. *B. Seismol. Soc. Am.*, 102(1), 204–221. https://doi.org/10.1785/0120110156

Yamamoto, Y., and Baker, J. (2011). *Stochastic model for earthquake ground motion using wavelet packets* [paper presentation]. 11th International Conference On Applications Of Statistics And Probability In Civil Engineering 2011, Zurich, Switzerland.

ZAMG-Zentralanstalt Für Meterologie Und Geodynamik (1987). *Austrian Seismic Network*. International Federation of Digital Seismograph Networks. https://doi.org/10.7914/SN/OE

Zanini, M., Hofer, L., and Faleschini, F. (2018). Reversible ground motion-to-intensity conversion equations based on the EMS-98 scale. *Eng. Struct., 180*, 310-320. https://doi.org/10.1016/j.engstruct.2018.11.032

Zhang, H. (2004). The Optimality of Naive Bayes. In: Barr, V., and Markov, Z. (Eds.), *Proceedings of the Seventeenth International Florida Artificial Intelligence Research Society Conference*. AAAI Press.

Živčić, M., Costa, G., and Govoni, A. (2000). *The 1998 Krn mountains earthquake sequence: the combined data set of the URSG, OGS and DST temporary and permanent seismic network (release 0.98)* [CD-ROM]. Uprava RS za Geofiziko, Ljubljana, Slovenia, Dipartimento di Scienze della Terra dell'Università di Trieste, Trieste, Italy, and Istituto Nazionale di Oceanografia e di Geofisica Sperimentale, Dipartimento Centro Ricerche Sismologiche, Udine, Italy.

Zollo, A., Orefice, A., and Convertito, V. (2014). Source Parameter Scaling and Radiation Efficiency of Microearthquakes Along the Irpinia Fault Zone in Southern Apennines, Italy. *J. Geophys. Res-Sol. Ea., 119*. https://doi.org/10.1002/2013JB010116