Site classification of Italian accelerometric stations from cluster analysis of residuals of peak ground motion data regressions

Vincenzo Del Gaudio[1 *], Pierpaolo Pierri[1], Nicola Venisti[2]

([1]) Dipartimento di Scienze della Terra e Geoambientali – Università degli Studi di Bari "Aldo Moro", Campus Universitario - via E. Orabona 4, 70125 Bari – Italy

vincenzo.delgaudio@uniba.it

pierpaolo.pierri@uniba.it

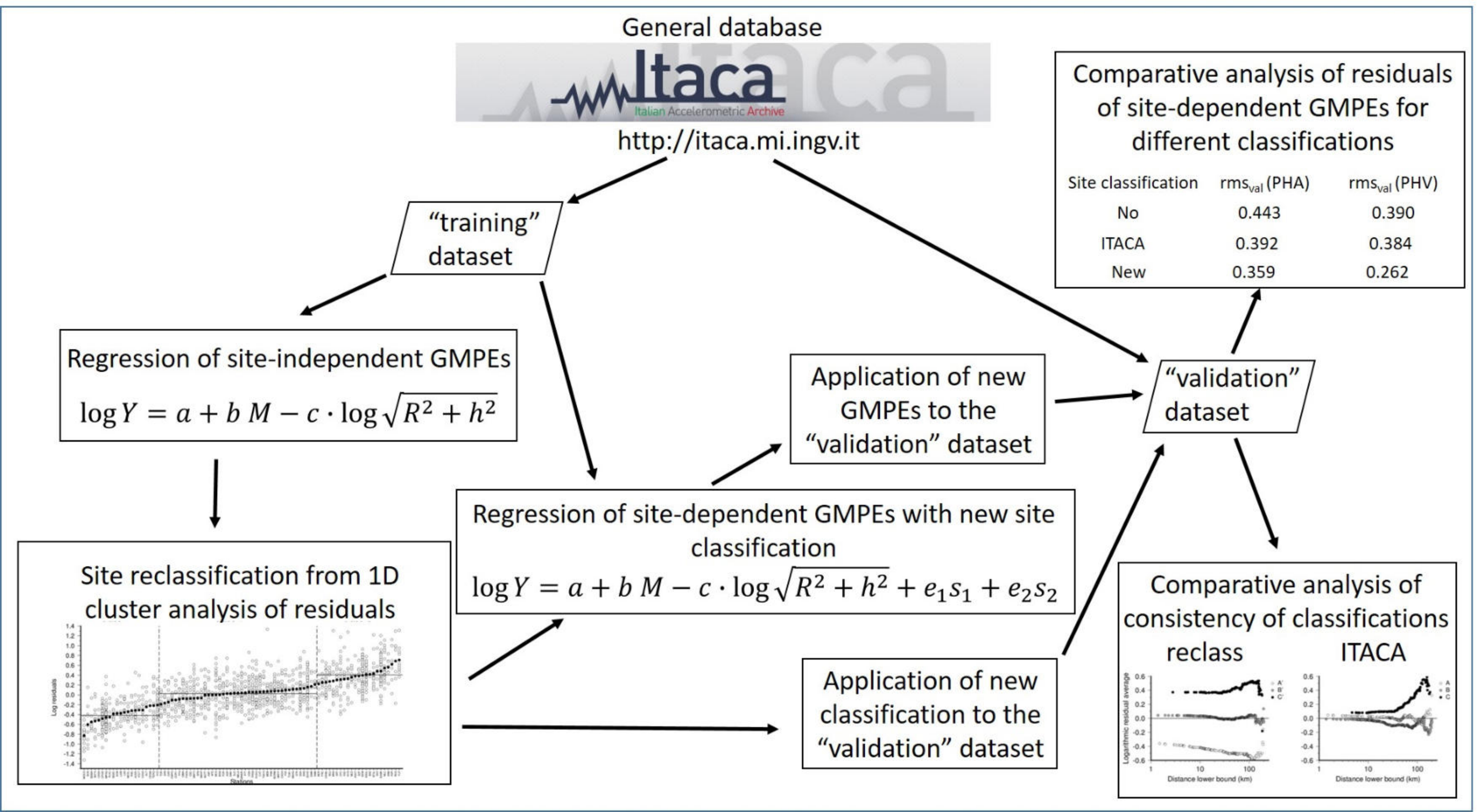
([2]) Osservatorio Sismologico - Università degli Studi di Bari "Aldo Moro", Campus Universitario - via E. Orabona 4, 70125 Bari – Italy

nicola.venisti@uniba.it

(*) Corresponding author

Published on Science of the Total Environment, doi: 10.1016/j.scitotenv.2019.05.073

- A method of accelerometric site reclassification was tested on Italian data.
- Univariate cluster analysis of site-independent GMPE regression residuals was used.
- Application to PHA and PHV observations proved to improve GMPE performances.
- A site can need distinct classifications for different ground motion parameters.
- Conventional classifications can lead to overestimate “base” seismic hazard.


General database
Itaca
Italian Accelerometric Archive
http://itaca.mi.ingv.it
"training" dataset
Regression of site-independent GMPEs
$\log Y = a + b\,M - c \cdot \log\sqrt{R^2 + h^2}$
Site reclassification from 1D cluster analysis of residuals
Regression of site-dependent GMPEs with new site classification
$\log Y = a + b\,M - c \cdot \log\sqrt{R^2 + h^2} + e_1 s_1 + e_2 s_2$
Application of new GMPEs to the "validation" dataset
Application of new classification to the "validation" dataset
"validation" dataset
Comparative analysis of residuals of site-dependent GMPEs for different classifications
Site classification | rms_{val} (PHA) | rms_{val} (PHV)
No | 0.443 | 0.390
ITACA | 0.392 | 0.384
New | 0.359 | 0.262
Comparative analysis of consistency of classifications
reclass
ITACA
Distance lower bound (km)
Distance lower bound (km)

Abstract

One basic element for seismic hazard assessment is the empirical definition of ground motion prediction equations (GMPE) to estimate shaking expected for earthquakes of given magnitude and distance. GMPEs are calibrated from data of accelerometric stations, distinguishing among site categories of different lithological type (e.g. hard rocks, more or less stiff soils) expected to cause different levels of ground motion amplification. Such a site classification is commonly based on geological observations and/or geophysical parameters like the mean propagation velocity of seismic waves through subsoil surficial layers. However, doubts have been raised about the effectiveness of results obtained from these conventional methods. Here we propose a methodology of accelerometric site classification relying on peak ground motion observations, exploiting the large amount of such observations available in the Italian National accelerometric database. The method is based on a cluster analysis of differences between observations and predictions provided by GMPEs whose functional form does not comprise site class among the explanatory variables. The new method was applied to the ITalian ACcelerometric Archive (ITACA), extracting a "training" dataset (used to calibrate some GMPEs through regressions) and a "validation" dataset" (to select the optimal GMPE form). A cluster analysis was then applied to regression residuals, grouping stations into three categories with increasing value of residual average. Checking the reclassification effectiveness through the examination of differences between independent "validation" observations and predictions of GMPEs adopting the new classification, these proved to be more consistent with site response properties than predictions provided by GMPEs using current classification.

Keywords: Italian accelerometric stations; site classification; ground motion prediction equations; cluster analysis; regression residuals.

1. Introduction

The main strategy followed at present for seismic risk mitigation relies on the adoption of design criteria aimed at providing constructions with the capacity of resisting to seismic shakings (defined “design seismic actions”). In order to optimize such a capacity, standard procedures of seismic hazard assessment are employed to evaluate the level of shaking expected not to be exceeded at a site with a probability representative of a targeted safety level (e.g. 90% in 50 years).

In the commonly used procedures of hazard assessment, after having estimated the number of earthquakes of different magnitudes expected to be generated in a given span of time by seismogenic structures located at different distances from a selected site, a fundamental stage consists of evaluating the ground motion expected at this site for such events. The most practical tools for these evaluations are empirical relations, named “Ground Motion Prediction Equations” (GMPE). They model with a simplified functional form the relation between some shaking parameters and a set of explanatory variables related to the length of the wave path and to properties both of the seismic source and of the wave propagation medium. With regard to the latter properties, a special emphasis is laid on local geological conditions at the site where ground motion needs to be predicted, in that such conditions are commonly observed to considerably modify local ground motion, sometimes increasing the level of ground shaking. In particular, ground motion amplitude is typically higher at sites where the upper few tens of meters of subsoil lithology present relatively lower stiffness. For this reason, GMPEs generally include in their functional form some terms accounting for such amplification effects, containing explanatory variables related with the propensity of site to amplify ground motion.

Such a propensity is also taken into account in the rules for the calculation of design seismic actions, typically through a categorization of sites into a set of classes differing according to a description of the stratigraphic profile and to geophysical/technical parameters measuring subsoil stiffness. In some cases, the belonging to these site classes is used as explanatory variable in GMPEs. For instance, the GMPEs obtained for Italy by Bindi et al. (2011) adopt the site categorization proposed within the technical rules recommended by the European Community for the design of earthquake-resistant structures in the Eurocode no. 8 (EC8 – see EN 1998-1). These rules define five ground types named A, B, C, D and E, identified according to the criteria reported in Table 1, and for each of them a different factor of increase of seismic actions is defined in comparison to class A, consisting of hard rock and assumed as reference site type.

As geophysical criterion supporting the site classification, the EC8 proposes to consider the average value of propagation velocity of S waves along a vertical path in the upper 30 m of the soil, commonly named VS,30. After being first introduced in the United States within the provisions for seismic regulations recommended by the National Earthquake Hazards Reduction Program (NEHRP), this parameter has had a widespread employment in site classification, even to model site effect in GMPE functional form. Indeed, VS,30 is commonly used either to categorize site classes represented in GMPE expression through dummy binary variables (e.g. Danciu and Tselentis, 2007; Bindi et al., 2011; Chousianitis et al., 2018), or is used directly as explanatory variable (e.g. Abrahamson et al., 2014; Boore et al., 2014; Campbell and Bozorgnia, 2014; Chiou and Youngs, 2014; Akkar et al., 2014; Lanzano et al., 2019a).

However, some doubts have been raised about the effectiveness of VS,30 in characterizing site dynamic response to shaking (cf. Castellaro et al., 2008), considering that it does not account for an important factor of site amplification, i.e. the stiffness contrast between the overburden soil and the underlying bedrock, whose interface could be located at depths different from 30 m. Furthermore, some studies have observed that, especially for shaking parameters depending on ground motion acceleration, the introduction of common site classes among the explanatory variables of GMPEs does not seem to improve their predictive performance (cf. Chousianitis et al., 2018). Therefore, we are experimenting a new approach to site classification to be used in GMPE calibration, based on the analysis of ground motion observations reported in large-scale accelerometric databases. The new method analyzes the residuals of regressions carried out on such databases, using GMPE functional forms not including site effect terms. The basic idea is that these residuals should present, within a variability related to factors not considered by GMPEs, a significant difference in their mean values for sites affected by different levels of amplification. A similar approach has been adopted in recent studies for site classifications relying on observed spectral accelerations (Puglia et al., 2015) and pseudo spectral accelerations (PSA) (Kotha et al., 2018). Here we apply this principle in a simplified way to the analysis of peak ground motion parameters, i. e. peak horizontal acceleration (*PHA*) and velocity (*PHV*).

For the purposes of this approach to site classifications, one can exploit the large amount of data currently available in national accelerometric databases, thanks to the increased number of stations of modern networks and their improved sensitivity, which allows the recording of events of different magnitudes over a large range of epicentral distances. A preliminary test of the new approach was conducted in a previous study, within an investigation on the performances of GMPEs calibrated for the prediction of *PHA* in the Greek area (Del Gaudio et al., under review). In the present study, the method is applied by using GMPEs aimed at estimating *PHA* and *PHV* values, calibrated for the Italian area on data of the ITtalian ACcelerometric Archive (ITACA – Pacor et al. 2011; Luzi et al., 2016). The extension of the experimentation of this method to a new area, with a larger dataset, provided new insight into potential and limits of the new approach to site classification.

2. Methodology

Several different functional forms have been proposed for GMPE calibration through regressions carried out on global or regional databases. While the first formulae were very simple, including just a couple of explanatory variables to represent the seismic source energy (typically through the earthquake magnitude) and the effect of ground motion reduction with distance, in the last decades new functional forms have been introduced, including an increasing number of terms and explanatory variables. For instance, separate terms were introduced to represent shaking attenuation with distance, one representing the effect of wave geometrical spreading and the other accounting for inelastic attenuation of the propagation medium, whereas the influence of site response and of the earthquake fault style were taken into account through additional variables.

However, the employment of more articulated functional forms does not always clearly demonstrate to provide a real improvement of the GMPE predictive performance. The reduction of misfit between predicted and observed ground motion sometimes is just a statistical artifact due to the better capacity of a function depending on more variables to adapt itself to a regression dataset. To certify that a certain functional form actually improves the GMPE prediction capacity, capturing the most significant sources of observation variability, one should test this capacity on a “validation” dataset different from that employed for regression.

Applying this approach to look for the functional forms best predicting different shaking parameters in the Greek region, we found that generally the best performance were not provided by the equations with the largest number of variables (Chousianitis et al., 2014; 2018; Del Gaudio et al., under review). For instance, the earthquake fault style turned out to be irrelevant for the accuracy of prediction of all the shaking parameters examined, in agreement with what found recently by other authors (Kotha et al., 2018; Lanzano et al., 2019a) about the absence of a clear correlation between differences of fault mechanism and of ground motion parameters.

With regard to the influence of site response, we found that the inclusion of terms depending on soil category was ineffective for the prediction of several acceleration-based parameters, including PHA, whereas improved the predictive performance of GMPEs for velocity-base parameters like PHV (Chousianitis et al., 2018). This, however, could depend on an incorrect classification of accelerometric station sites: one cannot exclude that currently used classification criteria, while are sufficiently adequate for PHV predictions, do not work properly for PHA. Therefore, we have experimented a new approach of site classification for PHA and PHV predictions, directly based on ground motion observations, rather than on parameters expected to be correlated to ground motion.

For this purpose, we propose to analyze the residuals derived from the regression of a GMPE not including site class as explanatory variable, optimized starting from a basic equation with a simple form, i.e.

$$\log Y = a + bM + c \cdot \log\sqrt{R^2 + h^2} + \, d \cdot \sqrt{R^2 + h^2} \pm \sigma \qquad [1],$$

where Y is a ground motion parameter (e.g. *PHA* or *PHV*), M is the moment magnitude, R is the epicentral distance and a, b, c, d, h are coefficients to be determined through regression carried out on a properly selected dataset (defined as “training” dataset).

In equation [1], the third and fourth terms are representative of the effects of geometrical spreading and inelastic attenuation, respectively, and the coefficient h is the so-called effective-depth parameter, introduced to account for the saturation effect constraining Y to finite values as R tends to zero. The coefficients of [1] are obtained from a two-step regression (Joyner and Boore, 1993). At a first stage, the coefficients of terms representing the effect of wave propagation (c, d, h) are calculated together with generic “event coefficients”, and then, with a second regression, coefficients depending on source properties (a, b) are determined from the “event coefficients” obtained at the previous stage (see Chousianitis et al., 2014 for more details). In this way it is possible to distinguish better the effect of ground motion variation with distance (within-event variability), from that depending on magnitude (between-event variability). The total standard deviation σ of regression is calculated as quadratic average of those of each regression step.

For the selection of the optimal functional form, we carry out preliminary tests comparing equations derived from [1] by removing the inelastic term or the coefficient h, or fixing the coefficient c to the theoretical value −1 (instead of calculating it from regression), or variously combining these modifications. Among all functional forms, we select the one providing the best predictive performance when applied to a distinct “validation” dataset (completely different from the “training” dataset used for regression). In particular, as optimal equation we choose the one minimizing the root mean square *rmsl* of prediction deviations from the actual observations.

The residuals of regression of the best performing equation are then used to classify the sites of stations whose recordings constitute the regression dataset. For this purpose, starting from the set of residual values found for different stations (see example in Fig. 1a), the averages of residuals relative to each station are calculated (black dots in Fig. 1) and the stations are ordered by increasing value of such averages (Fig. 1b). Stations are then grouped into a number l of classes, through a univariate cluster analysis, finding the l-1 limits (marked by dashed vertical lines in Fig. 1b) separating l classes so to minimize the quantity:

$$q = \sum_{k=1,l} \sum_{j=m_1(k),m_2(k)} \sum_{i=1,n(j,k)} \left(\varepsilon_{ijk} - \overline{\varepsilon_k}\right)^2 \quad [2]$$

where ε_{ijk} is the residual of the i-th recording of the j-th station belonging to the k-th class, $m_1(k)$ and $m_2(k)$ are the order numbers of the first and last station belonging to the k-th class, $n(j,k)$ is the number of recordings of the j-th station belonging to the k-th class and $\overline{\varepsilon_k}$ is the average of the values ε_{ijk} relative to the recordings of the stations belonging to the k-th class. Thus, in equation [2], q represents a measurement of observation scattering within a class, given by the quadratic sum of the deviations of residuals of recordings from the average relative to all the stations belonging to the same class (marked by solid horizontal lines in Fig. 1b).

The class limits satisfying the minimization of [2] are found through a procedure calculating iteratively the quantity q for all the possible combinations of the l-1 class limits, obtained by moving such limits along the series of stations ordered by residual averages. For the first class, $m_1(1) = 1$ and $m_2(1)$ varies from 2 to $M-l+1$, with M equal to the total number of stations, whereas for each of the following classes, $m_1(k)$ varies from $m_2(k-1)+1$ to $M-l+k$, and $m_2(k)$ varies from $m_1(k)$ to $M-l+k$.

Once a new site classification is obtained, its effectiveness is evaluated using it to calibrate a new GMPE, in this case including a function of site class expressed as linear function of l-1 dummy variables s_k according to the expression

$$S = \sum_{k=1,l-1} e_k s_k \quad [3].$$

The site class with the minimum negative residuals, corresponding to stations having the lowest values of ground motion, is represented by setting to 0 all the variables s_k, whereas each of the other classes is represented setting just one of the l-1 variables to 1 and all the others to 0.

The expression [3] is added to that of the best predicting equation previously obtained and the coefficients e_k are calculated from the first of a new two-step regression (i.e. the stage analyzing the factors of within-event variability), carried out on the training dataset. The resulting equation is then tested on the validation dataset to evaluate if the inclusion of site class terms based on the new classification improves the GMPE predictive performance in comparison to an equation with the same functional form, which relies on a conventional site classification.

3. Data

In order to apply the previously described methodology to the stations of the Italian National Accelerometric Network, *PHA* and *PHV* data were extracted from the database ITACA - release 2.1 (Luzi et al., 2016), now more easily available from the database ESM (Engineering Strong Motion database – Lanzano et al., 2019b: see https://esm.mi.ingv.it) . The ITACA database includes 25222 three-component accelerometric waveforms of 1365 earthquakes with magnitude M ≥ 3.0, recorded from 1972 to 2015 by 1210 stations (see Fig. 2a for their location).

Data selection was conditioned by two requirements of the techniques adopted for analysis i.e.: i) in the regression dataset, more than one recording for each seismic event must be included and ii) for each station, recordings should be distributed over a range of event magnitude and epicentral distances as wide as possible.

The first requirement is related to the need of excluding events for which just the recording at a single station is available, because such events cannot provide information on within-event variability (cf. Cotton et al., 2006; Bommer et al., 2010). The requirement ii) was suggested by the results of a first implementation of the method carried out on the Greek accelerometric database (Del Gaudio et al., under review). These results showed that the improvement of GMPE predictions provided by the new classification might be rather poor if data from a single station cover a narrow range of magnitude/distance combinations.

Based on these considerations, we selected a dataset including not less than 10 recordings for each accelerometric station and, for the training dataset, events with at least two recordings at differently distant stations. As a result of this selection, we extracted, from the ITACA database, recordings acquired at 87 stations, listed in Table 2 and whose geographical distribution is shown in Fig. 2b. For each station, Table 2 reports, together with its coordinates, the site class assigned by ITACA (according to the EC8 criteria), the VS,30, when available, the number of recordings used for our analysis and the ranges of event magnitudes and epicentral distances covered by these recordings. This dataset was subdivided into two subsets consisting of approximately 2/3 and 1/3 of the total, respectively, the former to be used as training dataset, with only events for which more than one recordings are available, and the latter forming a validation dataset with the remaining recordings. The distribution of recordings between the two subsets was made so to obtain a relatively good coverage through the total range of magnitudes and distances (see Fig. 3). The training dataset includes 1389 recordings of 204 events with magnitude between 3.1 and 6.4 (see Fig. 2c for their location), whereas, the validation dataset consists of 740 recordings of 378 events with magnitude between 3.1 and 6.9 (Fig. 2d). Both dataset includes only recordings of crustal events (depth up to 30 km) acquired at distances from about 1 to 200 km.

The selected recordings stations were classified by the ITACA database following the EC8 nomenclature (Luzi et al., 2015): they belong almost only to three of the existing classes and, in particular, 21 of them are of type A (rock), 40 of type B (very stiff soil) and 25 of type C (stiff soil). Just one station (CMPO – Campotto Po) is classified in class D but, for the purpose of this study, was associated to class C.

From Fig. 3, one can notice that the distribution of recordings over the total range of magnitudes and distances is not homogeneous and, in particular, only a very small number of recordings acquired at short distances (< 10 km) are relative to events of magnitude higher than 5.0. This depends on the fact that, while low energy seismicity is very diffuse, making more probable the detection of events even very close to an accelerometric station, the more infrequent strong events are more likely to be recorded at longer distances. This inhomogeneity, commonly observed in accelerometric dataset, is the reason for the adoption of a two-step regression aimed at better separating the modelling of ground motion reduction with distance, from that depending on magnitude decrease. Indeed, the simultaneous increase of distance of observations and of event magnitude tends to partially compensated each other, thus possibly causing an underestimate of the attenuation rate resulting from regression (cf. Fukushima and Tanaka, 1990).

4. Results of reclassification

The preliminary optimization of GMPE functional forms provided the following site-independent equations for *PHA* and *PHV*:

$$\log PHA = 0.620 + 0.689\,M - 1.975 \cdot \log\sqrt{R^2 + 10.69^2} \pm 0.359 \qquad [4],$$

$$\log PHV = -1.934 + 0.834\,M - 1.654 \cdot \log\sqrt{R^2 + 8.74^2} \pm 0.306 \qquad [5],$$

where *PHA* is the median value estimated for the geometric mean of the peak horizontal accelerations along east-west and north-south directions, measured in Gal (cm/s^2) and *PHV* is the analogous geometric mean for peak velocities measured in cm/s. Applied to the validation dataset, these equations provided errors, in the estimate of log *PHA* and log *PHV*, whose root mean square *rmsl* is 0.443 and 0.390, respectively. These *rmsl* values are slightly lower than those obtained with functional forms including the term representative of inelastic attenuation (0.447 and 0.393 for *PHA* and *PHV*, respectively). Such a result indicates that, at least within the examined range of distances (< 200 km), the GMPE predictive performance is not improved by separating geometrical spreading and inelastic attenuation instead of incorporating both of them into a single attenuation term. Figure 4 shows the distribution of regression residuals for both equations, together with the mean values of residuals and of magnitude of the recorded events, both calculated as running average over 21 recordings ordered by increasing epicentral distances. Residuals are largely scattered around an average close to 0 (0.02 for PHA and 0.04 for PHV) and their variations with distance show random oscillations without any clear trend. A possible asymmetry in residual distribution can be recognized for recordings at epicentral distances less than 4 km, which show prevailingly positive residuals, and, possibly, for recordings at distances greater than 100 km, which show a slightly decreasing trend. The first asymmetry concerns very few observations for events of low magnitude (from 3.1 to 4.5). The descending trend at distances greater than 100 km recalls what observed for high frequency PSAs by Kotha et al. (2018) about a major sensitivity of more distant observations to the inelastic attenuation, which motivated a bilinear modelling of attenuation with distance. This trend appears also inversely correlated with the obvious tendency to an increase of mean magnitude of events recorded at longer distances (see red lines in Fig.4). Thus, overall, it is possible that the adoption of a simplified GMPE implies an underestimation of predictions for small very close events and an overestimation for stronger events at distances > 100 km. However the underestimation affects only very few events, whereas the amount of mean residual decrease between 100 and 200 km is very low in comparison to the residual scattering, so that these possible biases should have minor influence on site classification based on residual analysis.

A cluster analysis according to the procedure described in section 2 was then applied to the regression residuals, tentatively assigning different values to the number l of classes. The resulting quadratic sums of residual deviations from class average (q in equation [2]) are plotted in Fig. 5 for both PHA and PHV residuals as function of the number l of classes. The value of q decreases quickly as l increases from 1 to 3 (by about the 90% of the total variation) and then undergoes only minor further reductions. This result is consistent with the fact that almost all the stations (save one), were assigned by ITACA to three classes, thus, also to maintain a comparability between the new and the standard classification, the value of l was finally set to 3.

The 87 stations examined in this study were therefore reclassified by grouping them into three classes named A′, B′ and C′, for the *PHA*-based reclassification, and A″, B″ and C″, for the *PHV*-

based one. These classes are intended as site typologies characterized by an increasing amplification factors in comparison to a reference category identified as that with the lower amplification level (classes A′ and A″).

Figure 6 illustrates the results of the cluster analyses based on residuals of *PHA* and *PHV* estimates. With regard to the *PHA*-based classification, 21 stations are assigned to the reference class A′, characterized by logarithmic residuals of −0.419±0.287 (where −0.419 is the average and 0.287 the standard deviation); 43 stations are classified as of category B′, with residuals of 0.027±0.279; the remaining 23 stations are in category C′, with residuals of 0.406±0.313. With regard to the *PHV*-based classification, 22 stations are in class A″, with logarithmic residuals of −0.302±243; 36 stations are classified into category B″, with residuals of 0.030±0.265; 29 stations are in category C″, with residuals of 0.302±0.277. In numerical terms, the distribution of stations among the three categories is quite similar to that of the conventional classification, with a prevalence of sites of intermediate classes (40 B, 43 B′, 36 B″), in comparison to sites with weaker (21 A and A′, 22 A″) or stronger amplification (26 C, 23 C′, 29 C″).

The mean residual at the boundary between the new classes could be used to classify other accelerometric stations, different from those listed in Table 1, whose data were not included in the training dataset. In particular, with regard to the *PHA*-based classification, the average between the mean residuals of stations AVZ and TOR, and that between the stations AMN and MCR, i.e. −0.200 and 0.218, respectively, can be set as limit separating classes A′ and B′ and classes B′ and C′, respectively. Thus, applying the equations [4] and [5] to calculate residuals of observations acquired at other stations, these can be assigned to A′, B′ or C′ according that the residual average is below −0.200, between −0.200 and 0.218 or over 0.218, respectively. Similarly, with regard to the *PHV*-based classification, mean residuals equal to −0.135 and 0.175, derived as averages over the station couples ACER-BOTT and BRZ-VSD, respectively, can be assumed as limits between A″-B″ and between B″-C″, respectively.

Tables 3 and 4 compare the classification reported by ITACA and the new ones. In both cases, old and new classifications agree for less than half of the classified sites (40 and 39 for *PHA*- and *PHV*-based classification, respectively) and in 10-13% of cases the ranking, in terms of expected amplification, differs by two levels (11 and 9 stations for *PHA*- and *PHV*-based classification, respectively). Stations classified by ITACA as of class A, are almost evenly distributed among the three categories of the new classifications (6 in A′ and A″, 7 in B′ and C″, 8 in C′ and B″) and sites of class B are more than half of those classified as of class A′ and A″. This can in part derive from an uncertain definition of boundaries between categories, so that moving such boundaries even by a small amount can transfer some sites across the limit between contiguous classes. However, in several cases the difference of classification reflects a poor correlation of observed residuals with conventional class assignments, which suggest the need of a reclassification. This result is consistent with the outcome of a recent study by Felicetta et al. (2018), which, adopting a combination of 6 geological and geophysical proxies to reclassify 47 stations of the Italian Accelerometric Network, found that a considerable number of stations previously assigned to class A needs to be reclassified as not adequate to be used as reference sites. From our analysis, a high number of class A sites (rock sites expected not to be amplified) are among those with the highest level of peak ground motion (8 in C′ and 7 in C″) and some of class C stations expected to be considerably amplified, are characterized by negative residual of ground motion prediction (3 in A′ and 2 in A″).

A measurement of the discrepancy between the ITACA classification of a station and that provided by the residual analysis can be obtained by calculating the minimum difference between the station mean residual and the values of residuals within the class having the same ranking as that assigned by ITACA to the station. For instance, for a class A station assigned by the PHA residual analysis to class B′ or C′, the minimum difference of its mean residual from a value compatible with the assignment of the station to the first class (as in ITACA) is that at the boundary between classes A′ and B′ (i.e. the average between AVZ and TOR). Normalizing these differences by the standard deviation of the same station residuals, these differences result larger than one standard deviations for a considerable number of stations (26% and 15% of *PHA*- and *PHV*-based classifications, respectively), and larger than two standard deviations for 8 of *PHA*-classified and 5 of *PHV*-classified stations (see values of *difn* in Tables 3 and 4).

Comparatively, a much better agreement is present between the two new classifications. No site is assigned by the PHA-based classification to a category differing by more than one level from that derived from *PHV* residuals (see Table 5). Discrepancies are mostly concentrated on groups of stations whose residual average is close to the limit separating two contiguous classes and only for one station (AVZ – Avezzano) this average differs from the limit of the class corresponding to that assigned using *PHA* residual by more than 1 standard deviation (but just for a tiny amount).

5. Validation tests

In order to evaluate the effectiveness of the new classifications, we first examined its consistency with ground motion parameters observed for recordings different from those used in GMPE regressions. For this purpose, the equations [4] and [5], estimating *PHA* and *PHV*, respectively, without consideration of site class differences, were applied to the validation dataset and differences of observations from GMPE predictions were examined. In order to verify whether these residuals show some dependence on seismic event magnitude and epicentral distance, they were plotted as function of such variables, after having averaged over some ranges of magnitude and distance, in order to reduce the influence of random fluctuations due to other variability factors not considered by GMPEs. Figures 7 and 8 shows the results obtained for *PHA* and *PHV*, respectively, averaging logarithmic residuals of predictions relative to magnitude and distance range comprised between an increasing lower bound and the maximum. Such averages were calculated separately for groups of stations classified as belonging to the same category, following both the conventional ITACA classification and the new ones based on *PHA* (Fig. 7) and *PHV* (Fig. 8) residuals. It is apparent that, adopting the new classifications, the residuals distribution points out clearly a systematic increase of peak ground motion values passing from the first to the third ground category for almost the entire range of magnitude and distances examined. Only when the averages are based on a small number of data (as it occurs towards the right end of the data series), separation among classes appears less pronounced, likely being more influenced by random fluctuations. The same clear separation among classes cannot be recognized when the classification reported by ITACA is adopted. In this case, sites of class A and B appear practical indistinguishable in terms of mean level of ground motion, whereas only the higher level of amplifications affecting class C sites can be in part recognized. Thus, the new classifications show a much better consistency with the expected influence of site response also for observations different from those used in the classification procedure.

As further test, we verified if, the adoption of the new classifications for the calibration of new GMPEs accounting for site effects is able to improve the GMPE predictive performance. Therefore, we calibrated new GMPEs including site terms in the form of equation [3]. Since all the examined stations are grouped into three categories, two dummy binary variables were introduced, i.e. $s_{B'}$ and $s_{C'}$ for the *PHA*-based classification, and $s_{B''}$ and $s_{C''}$ for the *PHV*-based classification. For sites belonging to the reference class (A′ and A″), both variables are set to 0, whereas, for each of the other classes, only the variable having the class name as subscript is set to 1. Using the same training dataset for regressions, the following equations were obtained:

$$\log PHA = 0.318 + 0.677\,M - 1.999 \cdot \log\sqrt{R^2 + 10.26^2} + 0.410\,s_{B'} + 0.777\,s_{C'} \pm 0.249 \quad [6]$$

$$\log PHV = -2.282 + 0.792\,M - 1.542 \cdot \log\sqrt{R^2 + 6.10^2} + 0.321\,s_{B''} + 0.597\,s_{C''} \pm 0.227 \quad [7].$$

For comparison homogeneity, regressions according to the same functional form, but assigning to stations the ITACA classification, were also carried out obtaining

$$\log PHA = 0.588 + 0.686\,M - 1.936 \cdot \log\sqrt{R^2 + 9.87^2} - 0.097 s_B + 0.072\,s_C \pm 0.351 \quad [8]$$

$$\log PHV = -2.018 + 0.826\,M - 1.598 \cdot \log\sqrt{R^2 + 7.27^2} - 0.060\,s_B + 0.190\,s_C \pm 0.286 \quad [9],$$

where s_B and s_C are set to 1 for class B and C, respectively, and to 0 otherwise.

Figure 9 shows the curve of predictions of these equations as function of distance, for different magnitudes and site classes, compared to the observations acquired at stations of the corresponding classes.

As first consideration, standard deviations of regressions are smaller for equations [6] and [7] adopting the new classifications. This could be expected, since such classifications are obviously better correlated to residuals of equations [4] and [5] (not including site terms) than the classifications used in equations [8] and [9], so that a larger amount of the observed ground motion variability is explained in terms of site response by the equations adopting the new classifications. More interestingly, on the one hand, the result of regressions adopting the ITACA classification, implies rather small amplification factors for the class of maximum site effect (1.2 for *PHA*, against the factor 1.5 proposed by EC8 - see Table 1 - and 1.5 for *PHV*). On the other hand the coefficients obtained for the second categories (B′ and B″) appear anomalous, in that they are negative, thus predicting for class B sites a slight de-amplification in comparison to rock sites (by a factor of 0.8 and 0.9, for *PHA* and *PHV*, respectively). These anomalous results can be explained considering that, according to the observed ground motion, most (about 70%) of class A sites should be assigned to classes of higher amplification level, about 1/3 of the class B sites to the reference class and a percentage from 44% (for *PHV*) to 60% (for *PHA*) of class C to less amplified classes. This leads to a reduction of differences among the mean amplification factors of the three ITACA classes, also exchanging the role of less amplified category between the first two classes.

Comparatively, equations [6] and [7], adopting the new classifications, are consistent with the expected increase of amplification factor passing from the first to the third class. The resulting factors are much larger than those proposed by EC8 (2.6 for class B′ and 6.0 for class C′), consistently, however, with what commonly observed on soil sites, where amplification factors larger than 2 are anything but unusual.

The better performance of the new classifications could reflect a better adaptation of the regression results just to the training dataset. To compare their performance on independent data, we applied all the obtained equations (from [4] to [8]) to predict the observations comprised in the validation dataset. Table 6 summarizes the results obtained. It is apparent that the adoption of the new classification outperforms not only the equations not including site effect among the explanatory variable, but also those obtained following the ITACA classification, by reducing estimate errors by an amount from 10% to 30%, on average.

6. Discussion

6.1 Comparison with other classification methods

An approach to site classification based on ideas similar to those inspiring the method we present in this study was recently proposed by Kotha et al. (2018). Their method applies a multidimensional cluster analysis to residuals of GMPE predicting pseudo spectral acceleration (PSA) for several periods between 0.01 to 2 s. For this purpose, these authors had the possibility to exploit a very large and high quality database acquired in Japan by the Kiban-Kyoshin network (Okada et al. 2004; Dawood et al., 2016), from which they extract a dataset including 15896 records of 850 events with magnitude between 3.4 and 7.3 recorded at distances up to 543 km. The functional form they adopted for GMPE has a term accounting for attenuation with distance differentiated for distances greater or shorter than 100 km, also including an effective-depth h depending on magnitude, and a magnitude scaling differentiated for different magnitude ranges. Comparatively, our method differs from that by Kotha et al. (2018) for the use of simpler functional forms for GMPE, justified by the availability of a smaller dataset, covering a smaller range of distances, to constrain the model parameters. Furthermore, our approach applies a unidimensional clustering approach separately to PHA and PHV, with the aim of exploring the hypothesis that different ground motion parameters requires distinct criteria of site classification.

Despite the simplifications introduced by our approach in comparison to more sophisticated ones, it proved to be able to considerably improve the GMPE predictions. This is particularly important for the predictions relative to the reference site class used for the assessment of "base" seismic hazard.

Indeed, it should be reminded that the main employment of GMPE in the framework of hazard assessment is for prediction of ground motion parameter under reference site conditions not affected by amplifications. The employment of ground type classification to account for the influence of local conditions on hazard and, consequently, on the definition of design seismic actions for building codes, is recommended only for few simple situations (flat horizontal layering with Vs increasing with depth according a few types of stratigraphic profiles). Otherwise, the evaluation of expected ground motion should be obtained by using site response numerical modelling, starting from ground shaking estimated for reference conditions, rather than using a GMPE accounting for different site conditions.

Our validation tests showed that an incorrect classification of reference site, including sites affected by amplification effects, can lead to a considerable overestimation of ground motion prediction (see dashed lines relative to class A sites in Fig. 9). Indications of the same type were also derived by Felicetta et al. (2018) in their exam of reference rock sites in Italy, conducted through a multiparametric criterion of reclassification.

The results of our tests show additionally that no “natural” separation among site classes emerges from ground motion observations (see Fig. 6): the averages of residuals relative to different stations present a gradual variation without no jump that could be related to a sharp change of site response properties. Thus, the subdivision into categories appears just as an artifice, which is functional to the practical convenience of discretizing the modelling of site effect influence on ground motion. This could be an argument in favor of a direct use of technical/geophysical parameters as continuous variable in a GMPE rather than as basis of site categorization, but the comparison of regression residuals with VS,30 values (when available) does not support the effectiveness of its use for this purpose. In Figure 10 the averages of residuals of site-independent *PHA* and *PHV* regressions, relative to stations for which VS,30 is available, are plotted as function of the VS,30 estimated values. While a decrease of residuals should be expected for stiffer, higher velocity site conditions, no correlation with VS,30 is observed for *PHA* residuals (the coefficient of determination R^2 of a linear regression being 0.001) and only a weak descending trend results from *PHV* data (but with a R^2 of only 0.22).

6.2. Possible causes of discrepancies from previous classifications

We also examined possible reasons of major discrepancies between conventional and new classifications. In general, such discrepancies could depend on an incorrect site classification for lack of data on subsoil properties. Actually, for the majority of the accelerometer stations selected for this study (52 out of 87), the site classification reported by ITACA relies on geological observations alone, not being available the VS,30 values (see Table 2).

On the other hand, the results of the classification criteria proposed here could be biased by data scarcity. It can be observed that the amplification factor of peak ground motion presents a certain dependence on magnitude and distance (cf. Del Gaudio and Wasowski, 2011), likely in relation to the closeness of the resonance frequency of site response to the frequency of wave maximum amplitude. Since low-magnitude, short-distance events present maxima of wave amplitude at relatively higher frequencies, for such events, rock sites, although not affected by significant amplification, can appear relatively more amplified because of a stronger response to higher frequency in comparison to sites on soil where such waves are more attenuated. Thus, if the training dataset includes, for a rock site, recordings comprised within a limited range of small magnitudes and short distances, this site would be classified as belonging to a site class affected by high amplification.

The need of considering, for residual-based reclassification, observations spanning, for each station, over a wide range of magnitudes and distances is also supported by the results of the analysis of residual scattering around GMPE predictions, shown in Fig. 4. The bias due to a possible underestimation of GMPE predictions at very short distances (< 4 km) and to overestimation at very long distances (> 100 km) can be countered avoiding the use of data of stations for which only recordings acquired within one of these two ranges are available. The validation test conducted on a dataset independent form that employed for regressions confirmed that, despite the possible presence of bias in GMPE estimates of ground motion at very short and very long distances, the reclassification proved to be able to correctly predict the distinct behaviour of reference sites in terms of expected ground motion (see Figs. 7-8).

6.3 Analysis of cases of major differences between previous and new classifications

To shed more light on causes of classification discrepancies, we analyzed in more detail the cases of differences larger than two standard deviations between the residual average of a station and the limit of the class corresponding to that assigned by ITACA to the same station (see parameter *difn* in Tables 3 and 4). For 5 stations (ASOL, ASO7, BCN, SNN, SRT) this condition is encountered both in *PHA*- and *PHV*-based classifications, whereas, for three more stations (ASR, BRZ, MOCO), only for the classification using *PHA* residuals.

In most of such cases (5 out of 8: ASOL, ASO7, ASR, BCN, MOCO) the ITACA classification actually is based only on geological surface observations, since VS,30 was not measured.

For stations ASOL and ASO7, both located in the municipality of Asolo (Veneto region), in different location (the local cemetery and a fortress, respectively) and for different spans of time, the attribution to class A is likely related to the local outcrop of a Miocene sandstone formation (Dal Piaz et al., 1946) in the station area. However, the mean difference of logarithm of peak ground motion observed from the median of GMPE prediction are rather high (0.413 – 0.381, for *PHA*, 0.379 – 0.377, for *PHV*), implying a mean increase by a factor larger than 2 in comparison to the expected median, which justifies its assignment to class C′ and C″. This assignment is based on a large number of data comprised in the training dataset (27 for ASOL and 6 for ASO7), covering a wide range of distances (60-200 km) and magnitudes (3.1-6.1). Thus, it appears quite well constrained and the association of these stations to class A could depend on lack of consideration of mechanical conditions of outcropping rocks and/or possible unrecognized shallow lithological variations. As alternative, considering that both stations are inside a building, the considerable amplification of ground motion could be due to the building response.

In case of station ASR (Ascoli Satriano, Puglia region), its attribution to class A appear weakly founded, since it is located on a surface where a Pleistocene conglomerate outcrops. This conglomerate is generally poorly compact, consisting of cobbles included in a sandy matrix, which only locally appear strongly cemented to form a pudding-stone (Malatesta et al., 1967). Thus, the attribution of station ASR to category C′ or C″, resulting from *PHA* and *PHV* residual analysis appears plausible, although relying on a relatively small number of recordings belonging to the training dataset (7), which covers a limited range of distances (30-100 km) and magnitude (3.2-4.5). The presence of significant amplification conditions is also supported by historical records reporting a high level of damages (with 4000 victims) in the Ascoli Satriano zone for a magnitude 6.0 earthquake occurred in 1361 (Boschi et al., 2000).

An opposite situation is found at station BCN (Buccino, Campania region), which ITACA classifies as C category, possibly in relation to the local presence of incoherent Holocene debris covering a substratum consisting of Cretaceous limestone (Cestari, 1971). *PHA* and *PHV* data includes this station in classes A′ and A″, on the basis of highly negative residuals (-0.605 and -0.409, for *PHA* and *PHV*, respectively) averaged over 11 recordings covering a range of magnitudes from 3.5 to 5.0 and distances between 15 to 190 km. Thus, ground motion appear here much weaker than the expected median, and the classification of BCN as an amplified site, reported by ITACA, could depend on the lacking evaluation of the local thickness of overburden soil, hiding a much more compact substratum which could have a major influence on site response.

The last case of classification not based on VS,30 (MOCO – Biccari Monte Cornacchia, Puglia region) is particular. Although ITACA classification differs only by one level from those resulting by the new approach (B against A′ and A″), station residual averages are very distant from the limits separating the first from the second class. Indeed, MOCO residuals are characterized by strongly negative values (on average -0.828 and -0.472, for *PHA* and *PHV*, respectively), which are the lowest among all the examined stations, thus indicating ground motion levels far from those expected for an even weakly amplified sites. Ground motion dataset, consisting of 12 recordings of events of magnitude between 3.6 and 5.2, acquired at distances from 30 to 180 km, provides a quite large range of observations. Thus, the assignment of MOCO to class B can be related to the local presence of a Miocene flysch formation, which however includes terms consisting of more compact carbonate rocks (Jacobacci et al., 1967). A more detailed analysis of local rock typology and mechanical conditions could be necessary for a clarification of site response properties of this station, which does not show evidence of significant amplifications.

Among the eight stations presenting major discrepancies between conventional and new classifications, three (BRZ, SNN, SRT) were classified by ITACA on the basis of VS,30 measurement. Despite resulting values larger than 800 m/s for all three stations (1030, 865 and 871 m/s, respectively), however, the residual averages of ground motion predictions are largely positive, particularly high for *PHA* (0.418, 0.561, 0.303, respectively) and a bit lower for *PHV* (0.165, 0.332, 0.236, respectively). Thus, while ITACA classifies the sites as of class A, the new classification assigns them to the third class (C′ and C″), with the only exception of *PHV* residuals of BRZ, which places this station in class B″.

The case of BRZ (Bersezio, Piemonte region), located on a relief consisting of Cretaceous carbonate rock (Crema et al., 1971), is particular, because, although 11 recordings are available in the complete dataset, only 2 of them are relative to events satisfying the requirements for two-step regression. Thus, its classification does not appear well constrained, also considering that the two recordings were acquired at similar epicentral distances (around 30 km) from two events of magnitude 3.4 and 4.9. Taking additionally into account that the *PHA*- and *PHV*-based classifications are not completely in agreement and that data from validation dataset show lower residuals, this reclassification based on residual analysis might be unreliable.

On the contrary, the classifications of SNN (Sannicandro Garganico, Puglia region) and SRT (Sortino, Sicilia region) are based on a relatively more consistent set of data (7 and 9 recordings, respectively) distributed over a diversified range of epicentral distances (20-120 and 20-135 km, respectively) and of magnitudes, at least in case of SNN (3.6-5.7 and 3.1-4.2, respectively). For these stations, ITACA provides also a Vs vertical profile, which, in case of SRT, does not extend below 30 m, whereas, for SNN, presents a complex pattern of variations, including alternation of slower and faster layers, with a deeper substratum characterized by velocities significantly higher than the VS,30 value (1500-2000 m/s at depths larger than 50 m). Overall, it is possible that VS,30 is not able to capture deeper velocity contrasts, responsible for amplification effects resulting in the high values of peak ground motion observed, which leads to their classification in classes C′ and C″.

Combining the observation relatives to stations SNN and SRT with what shown by Figure 7, it appears that situations where VS,30 is poorly representative of site propensity to amplification are quite recurrent. Thus, unless to find other parameters proving to work better than VS,30, the practice to take into account site effects for ground motion predictions through a discretized categorization, rather than using a single physical parameter, still appears a reasonable solution. The results of tests presented in this study and in a previous one relative to the Greek area (Del Gaudio et al., under review) indicate that categorizations based on GMPE regression residual analysis, can produce results not completely homogeneous when different ground motion parameters are employed. Although, until now, this has been found comparing *PHA* and *PHV* only, it is likely that similar conclusions would derive from the application of the proposed classification method to other shaking parameters, e.g. spectral acceleration response at different periods, which are more explicitly related to site resonance frequencies. However, the proposed classification approach offers the possibility of associating a “multi-categorization” to accelerometric sites, possibly assigning different classes to each of them for shaking parameters reflecting ground motion at different frequency intervals.

7. Conclusion

Considering uncertainties and errors that can affect conventional methods of site classifications relying on qualitative observation of surface geology and/or measurement of parameters not enough representative of all possible site amplification conditions, the direct use of ground motion observations for the classification of accelerometric stations offers interesting perspectives. In particular, a univariate cluster analysis of residuals of regressions of GMPEs that adopts site-independent functional forms, proved to be able to improve predictive performance of new GMPEs including site terms defined according to classifications that use the outcome of cluster analysis. Although the proposed approach for site classification cannot be of general use, being applicable only at sites for which a large amount of accelerometric observations are available, it can contribute to improve the definition of GMPEs suitable for "base" hazard assessment, which requires the prediction of ground motion expected for site conditions not affected by amplification. Indeed, the use of accelerometric database including recordings of stations classified as sites of class A on the basis of conventional methods can introduce ground motion data resulting from unrecognized amplification conditions. Furthermore, accelerometric stations not included in the dataset used for class limit definitions could be also re-classified by comparing these limits with the average of residuals resulting from the application of site-independent GMPE to recordings available for such stations. Finally, the proposed method of accelerometer site classification offers a tool for a cross-check of the reliability of station classifications obtained through different methods. In presence of strong inconsistencies of classifications, a more thorough investigation of the causes of such discrepancies can reveal inaccuracies in the results of the conventional or the new classification method, thus suggesting appropriate revisions.

Acknowledgements

This study was carried out with the financial support of Italian Ministry of Education, University and Research. Constructive comments and suggestions of two anonymous reviewers greatly contribute to improve the final version of this paper.

References

Abrahamson, N. A., Silva, W. J., Kamai, R., 2014. Summary of the ASK14 Ground Motion Relation for Active Crustal Regions. Earthquake Spectra, 30,. 1025-1055, doi: 10.1193/070913EQS198M.

Akkar, S., Sandıkkaya, M. A., Bommer, J. J., 2014. Empirical ground-motion models for point- and extended-source crustal earthquake scenarios in Europe and the Middle East. Bull. Earthq. Eng. 12, 359–387, doi 10.1007/s10518-013-9461-4.

Bindi, D., Pacor, F., Luzi, L., Puglia, R., Massa, M., Ameri, G., Paolucci, R., 2011. Ground motion prediction equations derived from the Italian strong motion database. Bull. Earthquake Eng., 9, 1899–1920, doi:10.1007/s10518-011-9313-z.

Bommer, J. J., Douglas, J., Scherbaum, F., Cotton, F., Bungum, H., Fäh, D., 2010. On the Selection of Ground-Motion Prediction Equations for Seismic Hazard Analysis. Seismol. Res. Lett., 81, 783-793, doi: 10.1785/gssrl.81.5.783.

Boore, D. M., Stewart, J. P., Seyhan, E., Atkinson, G. M., 2014. NGA-West2 Equations for Predicting PGA, PGV, and 5% Damped PSA for Shallow Crustal Earthquakes. Earthquake Spectra, 30, 1057-1085, doi: 10.1193/070113EQS184M.

Boschi, E., Guidoboni, E., Ferrari, G., Mariotti, D., Valensise, G., Gasperini, P., 2000. Catalogue of strong Italian earthquakes from 461 B.C. to 1997. Ann. Geofis., 43, 609-868.

Campbell, K. W., Bozorgnia, Y., 2014. NGA-West2 Ground Motion Model for the Average Horizontal Components of PGA, PGV, and 5% Damped Linear Acceleration Response Spectra. Earthquake Spectra, 30, 1087-1115, doi: 10.1193/062913EQS175M.

Castellaro, S., Mulargia, F., Rossi, P. L., 2008. VS,30: Proxy for Seismic Amplification? Seismol. Res. Lett., 79, 540-543.

Cestari, G., 1971. Note illustrative della Carta Geologica d'Italia. Foglio 198 – Eboli. Servizio Geologico d'Italia, Nuova Tecnica Grafica, Roma.

Chiou, B. S.–J., Youngs, R. R., 2014. Update of the Chiou and Youngs NGA Model for the Average Horizontal Component of Peak Ground Motion and Response Spectra. Earthquake Spectra, 30, 1117-1153, doi:10.1193/072813EQS219M.

Chousianitis, K., Del Gaudio, V., Kalogeras, I., Ganas, A., 2014. Predictive model of Arias intensity and Newmark displacement for regional scale evaluation of earthquake-induced landslide hazard in Greece. Soil Dyn. Earthq. Eng., 65, 11-29, doi: 10.1016/j.soildyn.2014.05.009.

Chousianitis, K, Del Gaudio, V, Pierri, P., Tselentis, GA., 2018. Regional Ground-Motion Prediction Equations for amplitude−, frequency response−, and duration−based parameters for Greece. Earthquake Engng Struct Dyn., 47, 2252–2274, doi: 10.1002/eqe.3067.

Cotton, F., Scherbaum, F., Bommer, J. J., Bungum, H., 2006. Criteria for selecting and adjusting ground-motion models for specific target regions: Application to Central Europe and rock sites. J. Seismol., 10, 137-156, doi: 10.1007/s10950-005-9006-7.

Crema, G., Dal Piaz, G.V., Merlo C., Zanella, E., 1971. Note illustrative della Carta Geologica d'Italia. Fogli 78-79-90 – Argentera – Dronero Demonte. Servizio Geologico d'Italia, Nuova Tecnica Grafica, Roma.

Dal Piaz, G., Venzo, S., Fabiani, R., Trevisan, L., Pia, J., 1946. Carta Geologica delle Tre Venezie, Foglio 37 – Bassano del Grappa. Ministero dei Lavori Pubblici, Studio Grafico Cartografico L. Boboli, Firenze.

Danciu, L, Tselentis, G-A., 2007. Engineering Ground-Motion Parameters Attenuation Relationships for Greece. Bull Seismo. Soc Am., 97, 162–183, doi: 10.1785/0120050087.

Dawood, HM, Rodriguez-Marek ,A, Bayless, J, Goulet, C, Thompson, E, 2016. A Flatfile for the KiK-net database processed using an automated protocol. Earthq. Spectra; 32:1281–302, doi: 10.1193/071214EQS106.

Del Gaudio, V., Wasowski J., 2011. Advances and problems in understanding the seismic response of potentially unstable slopes. Eng. Geol., 122, 73-83.

Del Gaudio, V, Pierri, P., Chousianitis, K., (under review). Influence of site response and focal mechanism on predictive performance of peak ground motion prediction equations for the Greek region. Soil. Dyn. Earthq. Eng.

EN 1998-1. Eurocode 8: Design of structures for earthquake resistance - Part 1: General rules, seismic actions and rules for buildings. European Committee for Standardization, 2004, pp. 229.

Felicetta, C., Lanzano, G., D'Amico, M., Puglia, R., Luzi, L., Pacor, F., 2018. Ground motion model for reference rock sites in Italy. Soil. Dyn. Earthq. Eng., 110, 276-283, doi: 10.1016/j.soildyn.2018.01.024.

Fukushima, Y., Tanaka., T.,1990. A new attenuation relation for peak horizontal acceleration of strong earthquake ground motion in Japan. Bull. Seismol. Soc. Am., 80, 757-783.

Jacobacci, A., Malatesta, A., Martelli, G., Stampanoni, G., 1967. Note illustrative della Carta Geologica d'Italia. Foglio 163 – Lucera. Servizio Geologico d'Italia, La Litograf, Roma.

Joyner, W.B, Boore, D.M., 1993. Methods for regression analysis of strong-motion data. Bull Seismol Soc Am., 83, 469–487.

Kotha, S.R., Cotton, F., Bindi, D., 2018. A new approach to site classification: Mixed-effects Ground Motion Prediction Equation with spectral clustering of site amplification functions. Soil. Dyn. Earthq. Eng., 110, 318-329, doi: 10.1016/j.soildyn.2018.01.051.

Lanzano, G., Luzi, L., Pacor, F., Felicetta, C., Puglia, R., Sgobba, S., D'Amico, M., 2019a. A revised Ground-Motion Prediction Model for Shallow Crustal Earthquakes in Italy. Bull. Seismol. Soc. Am., 109, 525–540, 10.1785/0120180210.

Lanzano, G., Sgobba, S., Luzi, L., Puglia, R., Pacor, F., Felicetta, C., D'Amico, M., Cotton, F., Bindi, D., 2019b. The pan-European Engineering Strong Motion (ESM) flatfile: compilation criteria and data statistics. Bull. Earthq. Eng., 17, 561-582, doi: 10.1007/s10518-018-0480-z

Luzi, L., Felicetta, C., Puglia, R., Russo, E., Pacor, F., D'Amico, I., Lanzano, G. and Working Group 2014 INGV-DPC Project S2-Task2, 2015. Site characterization in Itaca database. Atti 34° Convegno GNGTS, 17-19 novembre 2015, vol. 2, 122-128, http://www3.ogs.trieste.it/gngts/files/2015/tema_2.pdf.

Luzi, L., Pacor, F., Puglia, R., 2016. Italian Accelerometric Archive v 2.1. Istituto Nazionale di Geofisica e Vulcanologia, Dipartimento della Protezione Civile Nazionale. http://itaca.mi.ingv.it/ItacaNet, doi: 10.13127/ITACA/2.1.

Malatesta, A., Perno, U., Stampanoni, G., 1967. Note illustrative della Carta Geologica d'Italia. Foglio 175 – Cerignola. Servizio Geologico d'Italia, La Litograf, Roma.

Okada, Y, Kasahara, K, Hori, S, Obara, K, Sekiguchi, S, Fujiwara, H, Yamamoto, A., 2004. Recent progress of seismic observation networks in Japan—Hi-net, F-net, K-NET and KiKnet—, Earth Planets Space; 56:BF03353076, doi: 10.1186/BF03353076.

Pacor, F., Paolucci, R., Luzi, L., Sabetta, F., Spinelli, A., Gorini, A., Nicoletti, M., Marcucci, S., Filippi, L., Dolce, M. , 2011. Overview of the Italian strong motion database ITACA 1.0. Bull. Earthq. Eng., 9(6), 1723–1739. doi: 10.1007/s10518-011-9327-6.

Puglia, R., Albarello, D., Luzi,L., Bindi, D., Gallipoli, M.R., Mucciarelli, M., Naso, G., Pacor, F., Peronace, E., 2015. On the Performances of Site Parameters for Soil Classification. In: G. Lollino, D. Giordan, G. B. Crosta, J. Corominas, R. Azzam, J. Wasowski, N. Sciarra (Editors), Engineering

Geology for Society and Territory – Proceedings IAEG XII Congress, Turin, 15-19 September, 2014, Springer, Vol. 5, 1149-1152doi:10.1007/978-3-319-09048-1_219.

Tables

Table 1: Site classification according to the Eurocode 8. Descriptive and geophysical (based on VS,30) criteria of ground type identification are reported together with the amplification factors S applied to seismic actions for each type.

Ground type	Description of stratigraphic profile	Parameters	
		VS,30 (m/s)	Amplification factor S
A	Rock or other rock-like geological formation, including at most 5 m of weaker material at the surface.	> 800	1.0
B	Deposits of very dense sand, gravel, or very stiff clay, at least several tens of metres in thickness, characterised by a gradual increase of mechanical properties with depth.	360 - 800	1.4
C	Deep deposits of dense or medium–dense sand, gravel or stiff clay with thickness from several tens to many hundreds of metres.	180 - 360	1.5
D	Deposits of loose-to-medium cohesionless soil (with or without some soft cohesive layers), or of predominantly soft-to-firm cohesive soil.	< 180	1.8
E	A soil profile consisting of a surface alluvium layer with Vs values of type C or D and thickness varying between about 5 m and 20 m, underlain by stiffer material with Vs > 800 m/s .		1.6

Table 2: List of accelerometric stations of the Italian National Accelerometric Network, selected for this study. Legend: *Code* = station identification code; *Name* = station identification name; *Lat, lon* = station coordinates (latitude and longitude); *Class* = class assigned to station by ITACA; *VS,30* = mean S-wave velocity of the upper 30 m of station subsoil (n.a. when not available); *nrec* = number of recordings available for the station; *M, Dist* = magnitude and distance range of station data.

Code	*Name*	*Lat*	*Lon*	*Class*	*VS,30*	*nrec*	*M*	*Dist(km)*
ACER	ACERENZA	40.787	15.943	B	n.a.	28	3.1-5.2	11-192
ACQ	ACQUI TERME	44.683	8.462	C	n.a.	25	3.1-5.1	40-186
AMN	AMANTEA (CABINA ENEL)	39.137	16.080	B	n.a.	13	3.3-5.2	12-82
AMT	AMATRICE	42.632	13.286	B	670	25	3.1-5.3	6-180
AQP	L'AQUILA - V. ATERNO - M. PETTINO	42.384	13.369	A	836	40	3.1-5.6	1-84
ASO7	ASOLO ROCCA	45.805	11.918	A	n.a.	13	3.1-4.4	45-199
ASOL	ASOLO	45.800	11.902	A	n.a.	33	3.2-6.1	15-196
ASR	ASCOLI SATRIANO	41.199	15.563	A	n.a.	14	3.1-5.0	30-97
AVZ	AVEZZANO	42.027	13.426	C	199	33	3.1-6.3	11-130
BCN	BUCCINO	40.634	15.382	C	n.a.	18	3.1-5.0	10-198
BGN	BAGNONE	44.322	9.992	B	640	27	3.1-6.1	9-171
BOJ	BOJANO (NUOVA)	41.484	14.472	C	306	16	3.1-6.3	23-159
BORM	BORMIO	46.469	10.376	B	n.a.	30	3.2-6.1	70-193
BOTT	BOTTICINO	45.549	10.310	A	n.a.	48	3.2-6.1	50-183
BRZ	BERSEZIO	44.380	6.968	A	1030	11	3.4-4.9	24-77
CADA	CAPODARCO	43.194	13.761	B	n.a.	23	3.4-4.9	32-192
CAR	CARRODANO	44.247	9.620	A	n.a.	24	3.1-5.1	35-147
CASA	CASACALENDA	41.739	14.846	B	n.a.	11	3.1-4.6	3-10
CESM	CESI MONTE	43.005	12.903	A	n.a.	15	3.4-5.6	1.5-20
CHT	CHIETI	42.370	14.148	B	n.a.	33	3.1-6.3	44-134
CIMA	CIVITANOVA MARCHE	43.305	13.670	B	n.a.	25	3.4-4.9	23-176
CMPO	CAMPOTTO PO	44.581	11.805	C (D)	116	16	3.5-4.9	38-194
COR1	CORINALDO	43.632	13.000	B	n.a.	13	3.5-6.1	47-197
CPC	COPPARO (COCCANILE)	44.921	11.876	C	n.a.	22	3.7-6.1	36-116
CPGN	CARPEGNA	43.801	12.321	B	n.a.	74	3.1-6.1	21-199
CRND	CORNUDA	45.836	12.013	C	n.a.	28	3.3-6.1	6-175
CSC	CASCIA	42.719	13.012	B	698	26	3.1-6.0	7-121
CSN	CESENA	44.137	12.241	B	541	35	3.2-6.1	2-189
FAEN	FAENZA	44.290	11.877	C	n.a.	18	3.5-6.1	5-126
FAZ	FAENZA (NUOVA)	44.298	11.891	C	292	22	3.4-6.0	6-179
FLP	FELTRE (PASQUER)	46.027	11.923	C	n.a.	12	3.3-6.1	23-169
FOR	FORLI' (NUOVA)	44.199	12.042	C	296	22	3.2-6.1	6-194
FRC	FORGARIA CORNINO	46.221	12.997	B	454	24	3.4-6.1	1.2-25
FRE8	FREGONA	46.015	12.355	A	n.a.	41	3.1-6.1	21-196
GLT	GUALDO TADINO	43.233	12.789	C	n.a.	10	3.6-4.6	5-55
GMN	GEMONA	46.292	13.123	B	445	13	4.1-6.1	3-24
GSA	GRAN SASSO (LAB. INFN ASSERGI)	42.421	13.519	B	492	71	3.1-6.3	6-142
LARI	LARINO	41.805	14.919	B	n.a.	21	3.1-5.5	9-152
LSS	LEONESSA (NUOVA)	42.558	12.969	A	1091	20	3.2-6.3	10-51
MCEL	MONTICELLO	40.325	15.802	A	n.a.	21	3.2-5.2	37-170
MCR	MACERATA FELTRIA	43.800	12.448	C	n.a.	10	3.4-6.0	34-160

MCS	MERCATO SARACENO (NUOVA)	43.994	12.107	B	568	39	3.1-5.1	15-164
MDN	MODENA	44.647	10.890	C	213	26	3.6-6.1	24-98
MELA	MELANICO - S. CROCE DI MAGLIANO	41.706	15.127	A	n.a.	17	3.1-5.0	14-181
MLC	MALCESINE	45.808	10.849	B	430	30	3.1-6.4	5.3-196
MLD	MELDOLA	44.118	12.071	C	214	20	3.1-6.1	15-120
MMUR	MONTE MURANO	43.442	12.997	A	n.a.	31	3.1-5.2	38-197
MNS	MONSELICE	45.252	11.722	C	227	14	3.5-6.4	54-166
MNTP	MONTAPPONE	43.137	13.469	B	n.a.	21	3.4-4.9	29-176
MNTV	MANTOVA	45.150	10.790	C	237	48	3.2-6.1	31-175
MOCO	BICCARI MONTE CORNACCHIA	41.370	15.158	B	n.a.	22	3.6-5.2	29-185
MPAG	MONTE PAGANUCCIO	43.629	12.760	B	n.a.	63	3.1-6.1	32-199
MSAG	MONTE S. ANGELO	41.712	15.910	A	n.a.	19	3.1-5.0	29-199
NTE	NOTO (AREA ENEL)	36.910	15.069	B	659	10	3.1-4.4	8-129
PCRO	PIETRA CROCE	43.608	13.532	B	n.a.	42	3.1-4.9	18-193
PLAC	PLACANICA	38.449	16.438	A	n.a.	14	3.4-4.4	45-189
PNN	PENNABILLI	43.818	12.263	C	335	24	3.1-6.1	10-146
PP3	MAROLINO (POTENZA PICENA)	43.378	13.610	C	n.a.	19	3.4-4.6	29-193
PTT	PATTI - CABINA PRIMARIA	38.134	14.975	C	251	33	3.1-5.8	6-115
RDG	RODI GARGANICO	41.926	15.879	A	n.a.	11	3.1-4.6	31-194
RNC	RINCINE (LONDA)	43.870	11.607	A	859	11	3.1-6.1	21-124
SANR	SANDRIGO	45.640	11.610	C	321	64	3.1-6.1	41-199
SCF	SCAFA	42.265	13.998	B	n.a.	16	3.2-5.9	12-74
SCM	S. CROCE DI MAGLIANO	41.711	14.984	B	n.a.	16	3.1-6.3	7-150
SENI	SENIGALLIA	43.705	13.233	C	n.a.	66	3.3-6.3	41-197
SGMA	S. GIULIANO - PALAZZO MARCHESALE	41.685	14.964	B	n.a.	10	3.6-5.0	2-150
SGSC	S. GIULIANO - SCUOLA NUOVA	41.689	14.958	B	n.a.	11	3.6-5.0	2-178
SGTA	SANT'AGATA DI PUGLIA	41.135	15.365	B	n.a.	24	3.5-5.2	36-177
SIRI	MONTE SIRINO - MOLITERNO	40.182	15.868	B	n.a.	19	3.5-5.2	16-186
SLD	SALUDECIO	43.874	12.674	B	n.a.	29	3.1-4.9	50-168
SMAP	S. MARTINO IN PENSILIS	41.870	15.011	B	n.a.	20	3.1-5.5	18-143
SMU	SOMPLAGO CENTRALE - USCITA GALLERIA	46.340	13.061	B	n.a.	10	3.6-5.3	1-25
SNN	SANNICANDRO GARGANICO	41.832	15.571	A	865	15	3.1-5.7	11-117
SNS	SANSEPOLCRO	43.567	12.143	C	310	12	3.3-6.3	5-178
SPS	SPEZZANO DELLA SILA (CAMIGL.)	39.340	16.449	C	318	16	3.2-5.2	10-120
SRT	SORTINO	37.167	15.054	A	871	17	3.1-5.6	20-135
SSV	S. SEVERO	41.681	15.386	B	386	11	3.6-6.9	10-103
STAL	STALIGIAL	46.260	12.710	B	n.a.	25	3.3-6.1	12-197
STR	STURNO	41.021	15.116	B	382	11	3.2-6.9	12-58
T0815	T0815	44.873	11.720	C	n.a.	18	3.7-6.0	38-157
TERO	TERAMO	42.623	13.604	B	n.a.	21	3.5-5.0	20-156
TLM1	TOLMEZZO CENTRALE - DIGA AMBIESTA 1	46.381	12.984	B	458	12	3.1-6.4	10-32
TOR	TORTORICI	38.044	14.815	B	455	25	3.1-5.8	2-193
VAGA	VALLE AGRICOLA	41.415	14.234	A	n.a.	26	3.1-5.0	15-172
VSD	VIESTE (DANTE)	41.881	16.170	A	800	11	3.1-5.0	24-154
VULT	VULT	40.955	15.616	B	n.a.	13	3.6-5.0	69-196
ZOVE	ZOVENCEDO	45.454	11.488	B	n.a.	73	3.1-6.1	40-199

Table 3: Comparison between *PHA*-based and ITACA classifications of accelerometric stations. Cell reporting classes with different ranking in terms of expected amplification level are marked with different grey shades (white for A/A′, light grey for B/B′, dark grey for C/C′). Legend: *Code* = station code; *reclass* = class assigned from the analysis of *PHA* residuals; *ITACA* = class reported by ITACA; *means* = average of *PHA* residuals; *sds* = standard deviation of *PHA* residuals; *difn* = minimum difference between *means* and values within the class having the same ranking as the class assigned by ITACA to the station, normalized by the *sds* value.

Code	*reclass*	*ITACA*	*means*	*sds*	*difn*	*Code*	*reclass*	*ITACA*	*means*	*sds*	*difn*
MOCO	**A'**	**B**	-0.828	0.239	-2.633	**PTT**	**B'**	**C**	0.040	0.223	-0.798
BCN	**A'**	**C**	-0.605	0.365	-2.256	**CSN**	**B'**	**B**	0.055	0.316	0.000
SGMA	**A'**	**B**	-0.550	0.217	-1.616	**FRE8**	**B'**	**A**	0.056	0.268	0.953
SGTA	**A'**	**B**	-0.532	0.301	-1.107	**PCRO**	**B'**	**B**	0.057	0.269	0.000
CASA	**A'**	**B**	-0.517	0.212	-1.502	**CHT**	**B'**	**B**	0.061	0.213	0.000
CPGN	**A'**	**B**	-0.484	0.209	-1.361	**T0815**	**B'**	**C**	0.064	0.256	-0.600
MPAG	**A'**	**B**	-0.458	0.243	-1.063	**MDN**	**B'**	**C**	0.071	0.395	-0.371
PLAC	**A'**	**A**	-0.454	0.256	0.000	**PNN**	**B'**	**C**	0.073	0.212	-0.682
SGSC	**A'**	**B**	-0.385	0.258	-0.720	**BORM**	**B'**	**B**	0.074	0.197	0.000
SCM	**A'**	**B**	-0.380	0.216	-0.836	**MNTP**	**B'**	**B**	0.079	0.272	0.000
LARI	**A'**	**B**	-0.374	0.207	-0.843	**CADA**	**B'**	**B**	0.085	0.234	0.000
LSS	**A'**	**A**	-0.356	0.262	0.000	**FAZ**	**B'**	**C**	0.097	0.214	-0.563
MCEL	**A'**	**A**	-0.343	0.224	0.000	**TLM1**	**B'**	**B**	0.101	0.257	0.000
RNC	**A'**	**A**	-0.320	0.211	0.000	**MSAG**	**B'**	**A**	0.113	0.225	1.387
ACER	**A'**	**B**	-0.314	0.252	-0.452	**STR**	**B'**	**B**	0.121	0.378	0.000
VULT	**A'**	**B**	-0.309	0.369	-0.298	**GSA**	**B'**	**B**	0.127	0.260	0.000
VAGA	**A'**	**A**	-0.308	0.287	0.000	**SNS**	**B'**	**C**	0.133	0.315	-0.267
GLT	**A'**	**C**	-0.245	0.296	-1.562	**ZOVE**	**B'**	**B**	0.157	0.296	0.000
CESM	**A'**	**A**	-0.221	0.140	0.000	**GMN**	**B'**	**B**	0.173	0.289	0.000
TERO	**A'**	**B**	-0.219	0.378	-0.051	**AMN**	**B'**	**B**	0.213	0.158	0.000
AVZ	**A'**	**C**	-0.208	0.238	-1.788	**MCR**	**C'**	**C**	0.222	0.334	0.000
TOR	**B'**	**B**	-0.191	0.198	0.000	**FRC**	**C'**	**B**	0.255	0.200	0.186
SIRI	**B'**	**B**	-0.165	0.248	0.000	**CRND**	**C'**	**C**	0.257	0.216	0.000
CAR	**B'**	**A**	-0.147	0.316	0.166	**BOJ**	**C'**	**C**	0.272	0.275	0.000
CMPO	**B'**	**C**	-0.105	0.195	-1.657	**MNS**	**C'**	**C**	0.283	0.437	0.000
COR1	**B'**	**B**	-0.105	0.296	0.000	**SRT**	**C'**	**A**	0.303	0.242	2.074
PP3	**B'**	**C**	-0.084	0.212	-1.422	**CPC**	**C'**	**C**	0.311	0.412	0.000
CIMA	**B'**	**B**	-0.076	0.249	0.000	**VSD**	**C'**	**A**	0.322	0.432	1.208
FAEN	**B'**	**C**	-0.075	0.231	-1.265	**ACQ**	**C'**	**C**	0.327	0.392	0.000
CSC	**B'**	**B**	-0.074	0.248	0.000	**MLD**	**C'**	**C**	0.358	0.185	0.000
MNTV	**B'**	**C**	-0.071	0.304	-0.950	**ASO7**	**C'**	**A**	0.381	0.226	2.572
SPS	**B'**	**C**	-0.069	0.287	-1.000	**SCF**	**C'**	**B**	0.384	0.201	0.829
SMAP	**B'**	**B**	-0.064	0.252	0.000	**SENI**	**C'**	**C**	0.398	0.294	0.000
BOTT	**B'**	**A**	-0.007	0.286	0.674	**ASOL**	**C'**	**A**	0.413	0.214	2.867
AMT	**B'**	**B**	-0.004	0.219	0.000	**BRZ**	**C'**	**A**	0.418	0.064	9.705
NTE	**B'**	**B**	-0.001	0.058	0.000	**STAL**	**C'**	**B**	0.433	0.306	0.703
MCS	**B'**	**B**	-0.001	0.451	0.000	**SSV**	**C'**	**B**	0.484	0.280	0.952
BGN	**B'**	**B**	0.000	0.296	0.000	**SANR**	**C'**	**C**	0.485	0.317	0.000
FOR	**B'**	**C**	0.017	0.246	-0.815	**MLC**	**C'**	**B**	0.548	0.254	1.303
AQP	**B'**	**A**	0.025	0.328	0.685	**SNN**	**C'**	**A**	0.561	0.199	3.815
SMU	**B'**	**B**	0.028	0.138	0.000	**ASR**	**C'**	**A**	0.621	0.380	2.161
SLD	**B'**	**B**	0.032	0.195	0.000	**RDG**	**C'**	**A**	0.689	0.502	1.768
MELA	**B'**	**A**	0.036	0.237	0.993	**FLP**	**C'**	**C**	0.709	0.284	0.000
MMUR	**B'**	**A**	0.038	0.272	0.872						

Table 4: Comparison between *PHV*-based and ITACA classifications of accelerometric stations. Cell reporting classes with different ranking in terms of expected amplification level are marked with different grey shades (white for A/A″, light grey for B/B″, dark grey for C/C″). Legend: *Code* = station code; *reclass* = class assigned from the analysis of *PHV* residuals; *ITACA* = class reported by ITACA; *means* = average of *PHV* residuals; *sds* = standard deviation of *PHV* residuals; *difn* = minimum difference between *means* and values within the class having the same ranking as the class assigned by ITACA to the station, normalized by the *sds* value.

Code	*reclass*	*ITACA*	*means*	*sds*	*difn*	*Code*	*reclass*	*ITACA*	*means*	*sds*	*difn*
MOCO	**A"**	**B**	-0.472	0.198	-1.707	**MELA**	**B"**	**A**	0.080	0.263	0.816
LARI	**A"**	**B**	-0.434	0.280	-1.071	**CHT**	**B"**	**B**	0.090	0.200	0.000
MPAG	**A"**	**B**	-0.413	0.230	-1.214	**AMT**	**B"**	**B**	0.091	0.193	0.000
LSS	**A"**	**A**	-0.413	0.265	0.000	**CIMA**	**B"**	**B**	0.093	0.175	0.000
BCN	**A"**	**C**	-0.409	0.286	-2.044	**FRE8**	**B"**	**A**	0.097	0.266	0.869
SGMA	**A"**	**B**	-0.389	0.205	-1.240	**MSAG**	**B"**	**A**	0.098	0.206	1.133
GLT	**A"**	**C**	-0.358	0.289	-1.843	**CRND**	**B"**	**C**	0.100	0.241	-0.312
SGSC	**A"**	**B**	-0.333	0.211	-0.942	**MCR**	**B"**	**C**	0.120	0.324	-0.170
SGTA	**A"**	**B**	-0.316	0.190	-0.956	**TLM1**	**B"**	**B**	0.123	0.315	0.000
CASA	**A"**	**B**	-0.312	0.314	-0.566	**FOR**	**B"**	**C**	0.127	0.267	-0.179
PLAC	**A"**	**A**	-0.300	0.226	0.000	**ACQ**	**B"**	**C**	0.129	0.369	-0.123
MCEL	**A"**	**A**	-0.257	0.137	0.000	**FAEN**	**B"**	**C**	0.138	0.213	-0.173
VAGA	**A"**	**A**	-0.247	0.228	0.000	**CADA**	**B"**	**B**	0.150	0.224	0.000
SCM	**A"**	**B**	-0.234	0.243	-0.410	**BRZ**	**B"**	**A**	0.165	0.217	1.382
TERO	**A"**	**B**	-0.230	0.258	-0.368	**VSD**	**C"**	**A**	0.184	0.417	0.764
CPGN	**A"**	**B**	-0.223	0.218	-0.406	**STR**	**C"**	**B**	0.192	0.306	0.057
CESM	**A"**	**A**	-0.212	0.042	0.000	**PNN**	**C"**	**C**	0.193	0.206	0.000
CAR	**A"**	**A**	-0.208	0.258	0.000	**GMN**	**C"**	**B**	0.200	0.201	0.128
TOR	**A"**	**B**	-0.194	0.210	-0.286	**AMN**	**C"**	**B**	0.201	0.146	0.180
VULT	**A"**	**B**	-0.186	0.251	-0.207	**PTT**	**C"**	**C**	0.201	0.178	0.000
NTE	**A"**	**B**	-0.152	0.145	-0.123	**STAL**	**C"**	**B**	0.204	0.314	0.093
ACER	**A"**	**B**	-0.142	0.202	-0.037	**FAZ**	**C"**	**C**	0.214	0.220	0.000
BOTT	**B"**	**A**	-0.127	0.259	0.029	**FLP**	**C"**	**C**	0.231	0.347	0.000
RNC	**B"**	**A**	-0.123	0.186	0.059	**SRT**	**C"**	**A**	0.236	0.179	2.071
SIRI	**B"**	**B**	-0.121	0.172	0.000	**T0815**	**C"**	**C**	0.238	0.217	0.000
BGN	**B"**	**B**	-0.071	0.215	0.000	**SCF**	**C"**	**B**	0.242	0.201	0.333
BORM	**B"**	**B**	-0.068	0.224	0.000	**SNS**	**C"**	**C**	0.249	0.310	0.000
SMAP	**B"**	**B**	-0.045	0.305	0.000	**MNTV**	**C"**	**C**	0.251	0.268	0.000
SPS	**B"**	**C**	-0.043	0.270	-0.806	**SSV**	**C"**	**B**	0.255	0.236	0.339
MMUR	**B"**	**A**	-0.040	0.241	0.390	**FRC**	**C"**	**B**	0.265	0.255	0.355
CSC	**B"**	**B**	-0.039	0.187	0.000	**MLC**	**C"**	**B**	0.278	0.247	0.420
MCS	**B"**	**B**	-0.037	0.417	0.000	**CMPO**	**C"**	**C**	0.283	0.145	0.000
SMU	**B"**	**B**	-0.030	0.222	0.000	**RDG**	**C"**	**A**	0.299	0.435	0.998
GSA	**B"**	**B**	-0.003	0.254	0.000	**SNN**	**C"**	**A**	0.332	0.178	2.626
COR1	**B"**	**B**	0.017	0.319	0.000	**SENI**	**C"**	**C**	0.367	0.287	0.000
CSN	**B"**	**B**	0.019	0.283	0.000	**MDN**	**C"**	**C**	0.367	0.389	0.000
ZOVE	**B"**	**B**	0.022	0.296	0.000	**MLD**	**C"**	**C**	0.369	0.157	0.000
MNS	**B"**	**C**	0.041	0.434	-0.309	**SANR**	**C"**	**C**	0.371	0.295	0.000
SLD	**B"**	**B**	0.045	0.213	0.000	**CPC**	**C"**	**C**	0.372	0.373	0.000
MNTP	**B"**	**B**	0.048	0.215	0.000	**ASO7**	**C"**	**A**	0.377	0.232	2.204
AQP	**B"**	**A**	0.049	0.329	0.556	**ASOL**	**C"**	**A**	0.379	0.219	2.349
PCRO	**B"**	**B**	0.050	0.227	0.000	**ASR**	**C"**	**A**	0.400	0.371	1.440
AVZ	**B"**	**C**	0.072	0.197	-0.523	**BOJ**	**C"**	**C**	0.495	0.315	0.000
PP3	**B"**	**C**	0.073	0.169	-0.599						

Table 5: Comparison between the results of reclassifications of accelerometric stations. Cell reporting classes with different ranking in terms of expected amplification level are marked with different grey shades (white for A′/A″, light grey for B′/B″, dark grey for C′/C″). Legend: *Code* = station code; *cl(PHV), cl(PHA)* = class assigned from the analysis of *PHV* and *PHA* residuals, respectively; *means* = average of station *PHV* residuals: *sds* = standard deviation of station on *PHV* residuals; *difn* = minimum difference between *means* and values within the class having the same ranking as the class assigned to the station according to *PHA* residuals, normalized by the *sds value.*

Code	***cl(PHV)***	***cl(PHA)***	***means***	***sds***	***difn***
MOCO	**A"**	**A'**	-0.472	0.198	0.000
LARI	**A"**	**A'**	-0.434	0.280	0.000
MPAG	**A"**	**A'**	-0.413	0.230	0.000
LSS	**A"**	**A'**	-0.413	0.265	0.000
BCN	**A"**	**A'**	-0.409	0.286	0.000
SGMA	**A"**	**A'**	-0.389	0.205	0.000
GLT	**A"**	**A'**	-0.358	0.289	0.000
SGSC	**A"**	**A'**	-0.333	0.211	0.000
SGTA	**A"**	**A'**	-0.316	0.190	0.000
CASA	**A"**	**A'**	-0.312	0.314	0.000
PLAC	**A"**	**A'**	-0.300	0.226	0.000
MCEL	**A"**	**A'**	-0.257	0.137	0.000
VAGA	**A"**	**A'**	-0.247	0.228	0.000
SCM	**A"**	**A'**	-0.234	0.243	0.000
TERO	**A"**	**A'**	-0.230	0.258	0.000
CPGN	**A"**	**A'**	-0.223	0.218	0.000
CESM	**A"**	**A'**	-0.212	0.042	0.000
CAR	**A"**	**B'**	-0.208	0.258	-0.285
TOR	**A"**	**B'**	-0.194	0.210	-0.286
VULT	**A"**	**A'**	-0.186	0.251	0.000
NTE	**A"**	**B'**	-0.152	0.145	-0.123
ACER	**A"**	**A'**	-0.142	0.202	0.000
BOTT	**B"**	**B'**	-0.127	0.259	0.000
RNC	**B"**	**A'**	-0.123	0.186	0.059
SIRI	**B"**	**B'**	-0.121	0.172	0.000
BGN	**B"**	**B'**	-0.071	0.215	0.000
BORM	**B"**	**B'**	-0.068	0.224	0.000
SMAP	**B"**	**B'**	-0.045	0.305	0.000
SPS	**B"**	**B'**	-0.043	0.270	0.000
MMUR	**B"**	**B'**	-0.040	0.241	0.000
CSC	**B"**	**B'**	-0.039	0.187	0.000
MCS	**B"**	**B'**	-0.037	0.417	0.000
SMU	**B"**	**B'**	-0.030	0.222	0.000
GSA	**B"**	**B'**	-0.003	0.254	0.000
COR1	**B"**	**B'**	0.017	0.319	0.000
CSN	**B"**	**B'**	0.019	0.283	0.000
ZOVE	**B"**	**B'**	0.022	0.296	0.000
MNS	**B"**	**C'**	0.041	0.434	-0.309
SLD	**B"**	**B'**	0.045	0.213	0.000
MNTP	**B"**	**B'**	0.048	0.215	0.000
AQP	**B"**	**B'**	0.049	0.329	0.000
PCRO	**B"**	**B'**	0.050	0.227	0.000
AVZ	**B"**	**A'**	0.072	0.197	1.048
PP3	**B"**	**B'**	0.073	0.169	0.000
MELA	**B"**	**B'**	0.080	0.263	0.000
CHT	**B"**	**B'**	0.090	0.200	0.000
AMT	**B"**	**B'**	0.091	0.193	0.000
CIMA	**B"**	**B'**	0.093	0.175	0.000
FRE8	**B"**	**B'**	0.097	0.266	0.000
MSAG	**B"**	**B'**	0.098	0.206	0.000
CRND	**B"**	**C'**	0.100	0.241	-0.312
MCR	**B"**	**C'**	0.120	0.324	-0.170
TLM1	**B"**	**B'**	0.123	0.315	0.000
FOR	**B"**	**B'**	0.127	0.267	0.000
ACQ	**B"**	**C'**	0.129	0.369	-0.123
FAEN	**B"**	**B'**	0.138	0.213	0.000
CADA	**B"**	**B'**	0.150	0.224	0.000
BRZ	**B"**	**C'**	0.165	0.217	-0.045
VSD	**C"**	**C'**	0.184	0.417	0.000
STR	**C"**	**B'**	0.192	0.306	0.057
PNN	**C"**	**B'**	0.193	0.206	0.087
GMN	**C"**	**B'**	0.200	0.201	0.128
AMN	**C"**	**B'**	0.201	0.146	0.180
PTT	**C"**	**B'**	0.201	0.178	0.149
STAL	**C"**	**C'**	0.204	0.314	0.000
FAZ	**C"**	**B'**	0.214	0.220	0.179
FLP	**C"**	**C'**	0.231	0.347	0.000
SRT	**C"**	**C'**	0.236	0.179	0.000
T0815	**C"**	**B'**	0.238	0.217	0.293
SCF	**C"**	**C'**	0.242	0.201	0.000
SNS	**C"**	**B'**	0.249	0.310	0.238
MNTV	**C"**	**B'**	0.251	0.268	0.285
SSV	**C"**	**C'**	0.255	0.236	0.000
FRC	**C"**	**C'**	0.265	0.255	0.000
MLC	**C"**	**C'**	0.278	0.247	0.000
CMPO	**C"**	**B'**	0.283	0.145	0.750
RDG	**C"**	**C'**	0.299	0.435	0.000
SNN	**C"**	**C'**	0.332	0.178	0.000
SENI	**C"**	**C'**	0.367	0.287	0.000
MDN	**C"**	**B'**	0.367	0.389	0.495
MLD	**C"**	**C'**	0.369	0.157	0.000
SANR	**C"**	**C'**	0.371	0.295	0.000
CPC	**C"**	**C'**	0.372	0.373	0.000
ASO7	**C"**	**C'**	0.377	0.232	0.000
ASOL	**C"**	**C'**	0.379	0.219	0.000
ASR	**C"**	**C'**	0.400	0.371	0.000
BOJ	**C"**	**C'**	0.495	0.315	0.000

Table 6: Root mean square of errors in the prediction of *PHA* and *PHV* values of the validation dataset using the basic equations, with no site terms, or including such terms but following different classification criteria (ITACA criteria or analysis of *PHA* and *PHV* regression residuals).

Site classification	**rms_{val} (*PHA*)**	**rms_{val} (*PHV*)**
No	0.443	0.390
ITACA	0.392	0.384
Re-classification	0.359	0.262

Figures

Figure 1: Sketch illustrating the procedure for site reclassification. Open circles and black dot at the same position on the x-axis represent the residuals of GMPE predictions and the residual average, respectively, for recordings at the station whose code is reported at the same position below the axis. In diagram a), stations are alphabetically ordered along the x-axis, in diagram b) are ordered by increasing values of residual averages. On the latter diagram, vertical dashed lines separate the stations classified into different site classes and horizontal solid lines mark the average of residuals for the stations of each class.

Figure 2: Geographical distribution of (a) stations of the ITACA database, (b) recording stations selected for the implementation of the classification method, (c) seismic events selected for the training dataset and (d) for the validation dataset. In (b) the site classification reported by ITACA for the selected stations is represented by different symbols (A – white circles, B – grey squares, C – black triangles).

Figure 3: Magnitude of recorded events as function of distance of the recording station for the "training" dataset (a), used in equation regressions, and for the "validation" dataset (b), used in the equation effectiveness evaluation. White circles, grey squares and black triangles are used for data acquired at sites classified by ITACA as of type A, B and C, respectively, according to the EC8 soil categories.

Figure 4: Residuals of regressions for site-independent equations [4] (PHA) and [5] (PHV) as function of epicentral distance of recordings used in regressions. Colors refer to magnitude of the recorded events, according to the scale below the diagrams. Black and red lines represent the running mean of residuals and of magnitude of the recorded events, respectively, both averaged over 21 data of increasing epicentral distance, as function of mean distance.

Figure 5: Diagram of the scattering of observations within the classes (expressed through the quadratic sums *q* of deviations of observation residuals from class average) as function of the number *l* of classes used in clustering of PHA (top) and PHV (bottom) residuals.

Figure 6: Results of cluster analyses carried out on the regression residuals relative to *PHA* (a) and *PHV* (b) predictions. Open circles and black dots represent the residuals of single GMPE predictions and their average, respectively, for each station whose code is reported on the x-axis. Vertical dashed lines separate the stations classified into different site classes and horizontal solid lines mark the average of residuals for the stations of each site class.

Figure 7: Residuals of log(*PHA*) predictions obtained applying a site-independent GMPE to the validation dataset. Residuals are averaged over ranges of recording distances (top) and event magnitudes (bottom), from an increasing lower bound (indicated on the x-axis), to the maximum of distance and magnitude, respectively. Results obtained for site of different classes are plotted with different symbols (according to the legend), following the ITACA classification (to the right) and that proposed in the present study (to the left).

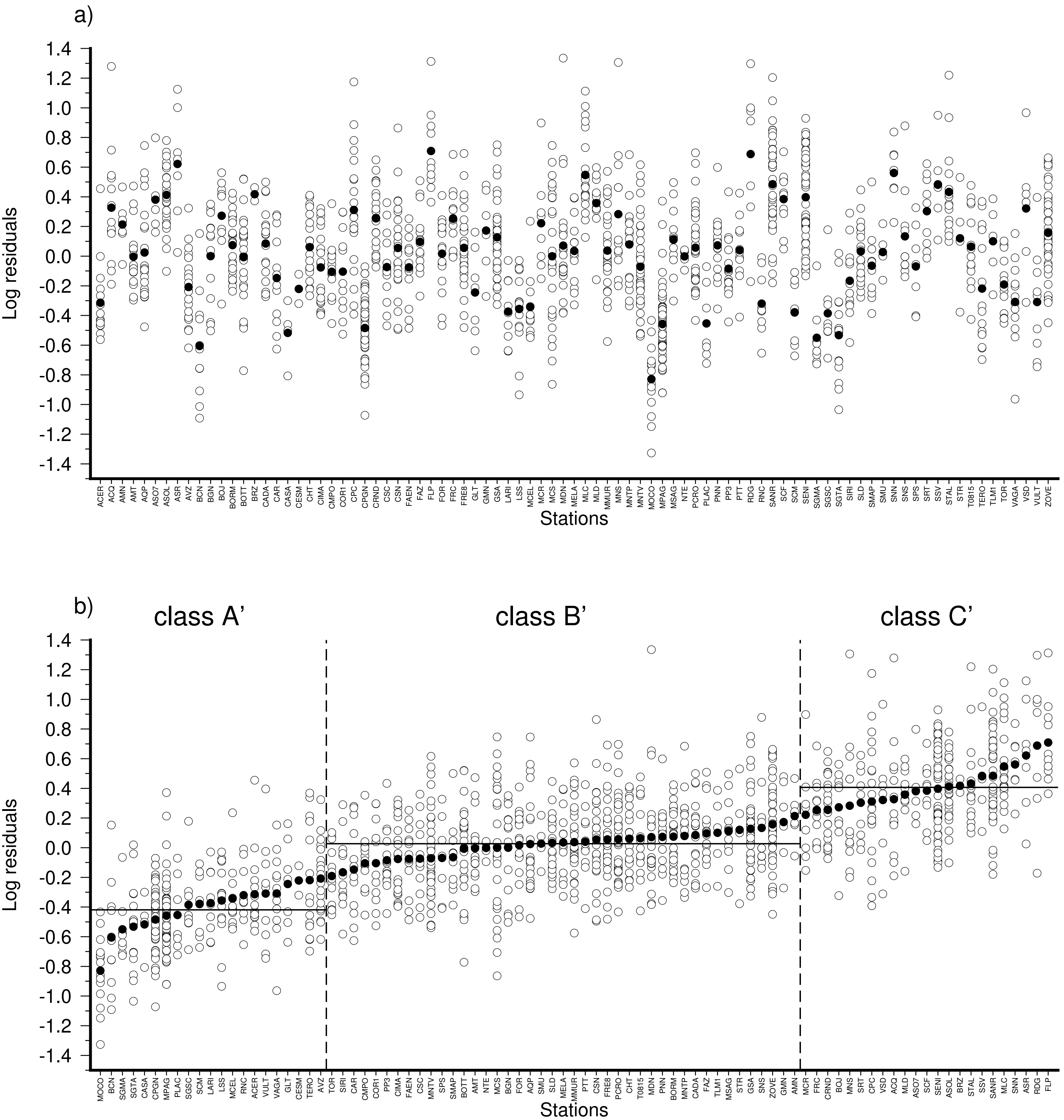
Figure 8: Residuals of log(*PHV*) predictions obtained applying a site-independent GMPE to the validation dataset. Residuals are averaged over ranges of recording distances (top) and event magnitudes (bottom), from an increasing lower bound (indicated on the x-axis), to the maximum of distance and magnitude, respectively. Results obtained for site of different classes are plotted with different symbols (according to the legend), following the ITACA classification (to the right) and that proposed in the present study (to the left).

Figure 9: Curves of variation with distance of predictions from site-dependent GMPEs derived adopting the new classification (solid lines) based on the clustering analysis of residuals of PHA (to the left) and PHV (to the right), or the ITACA classification (dashed lines), for different site classes (A-A′-A″, top, B-B′-B″, centre, C-C′-C″, bottom). Comparatively, the observations acquired at

stations classified from PHA and PHV residuals are reported as circles coloured according to the magnitude of the recorded events.

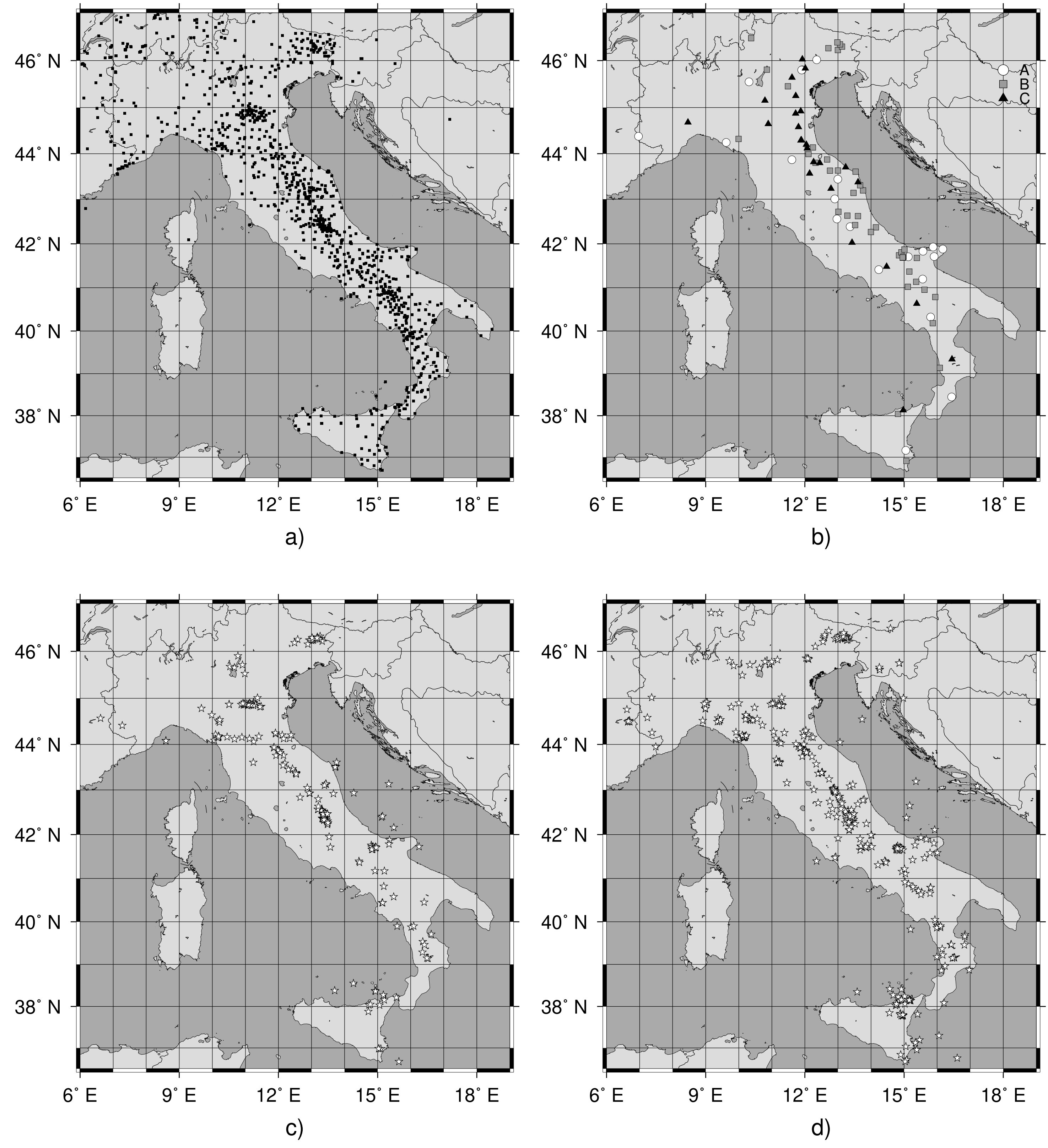
Figure 10: Diagrams of mean logarithmic residuals at single stations for site-independent *PHA* (a) and *PHV* (b) regressions, as function of VS,30 reported by the ITACA database for the same stations (see Table 2).

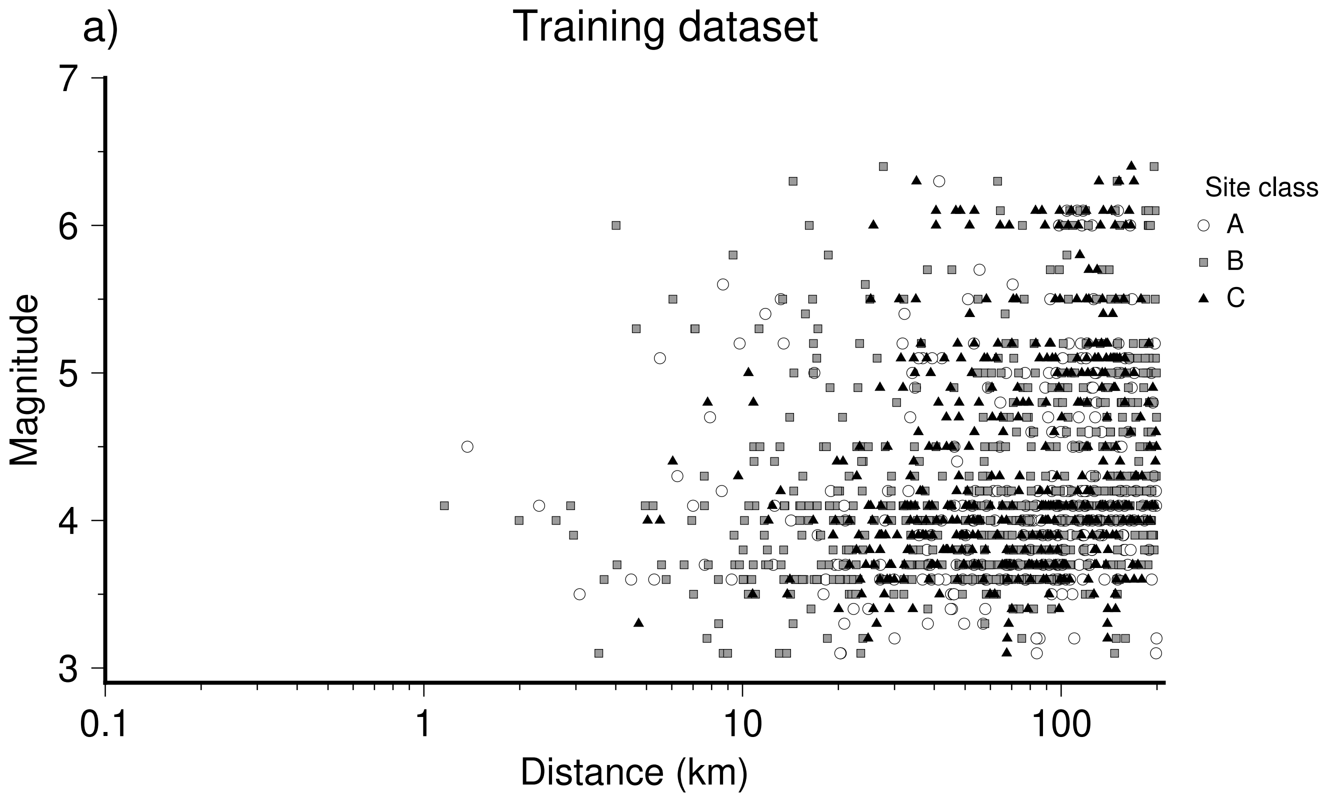
Fig. 1



a)
Log residuals
Stations
b)
class A'
class B'
class C'
Log residuals
Stations

Fig. 2




a)
Training dataset
Magnitude
Distance (km)
Site class
A
B
C
7
6
5
4
3
0.1
1
10
100

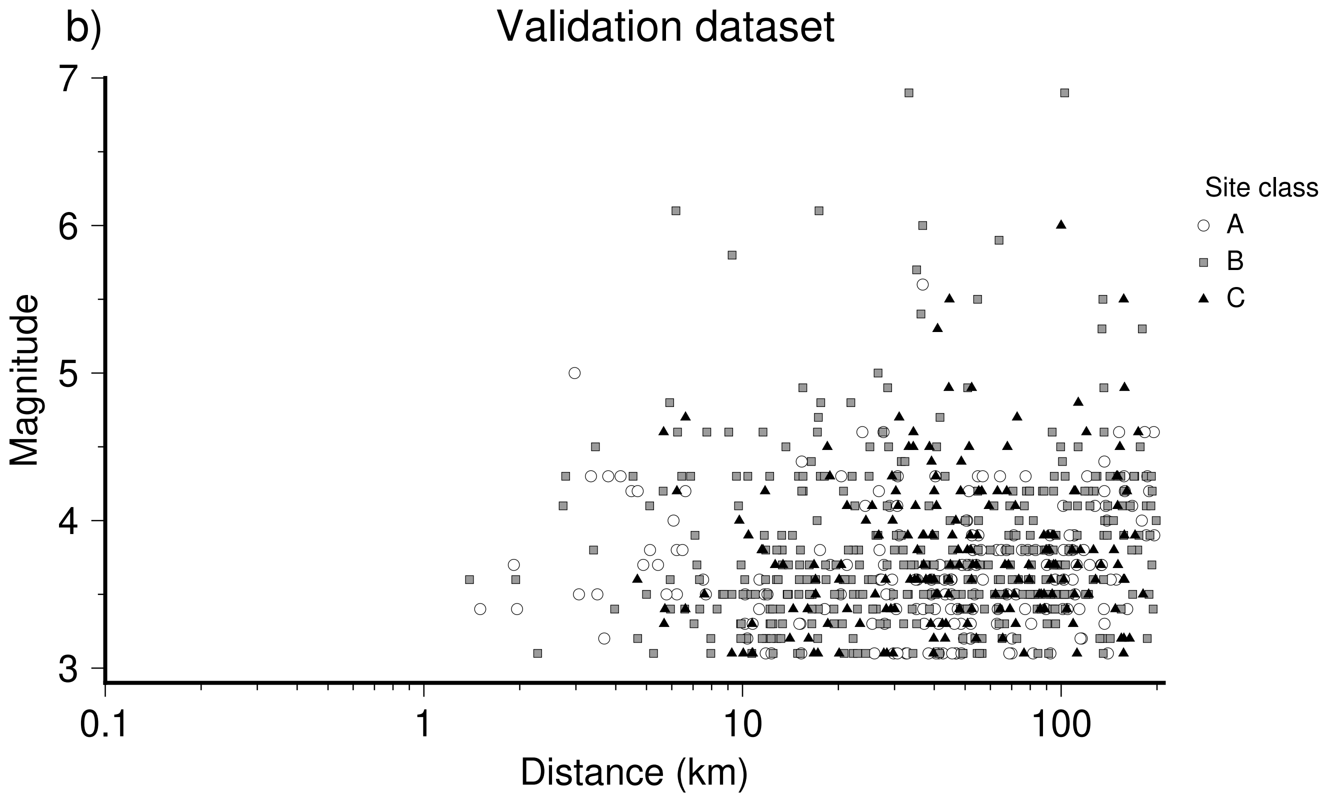
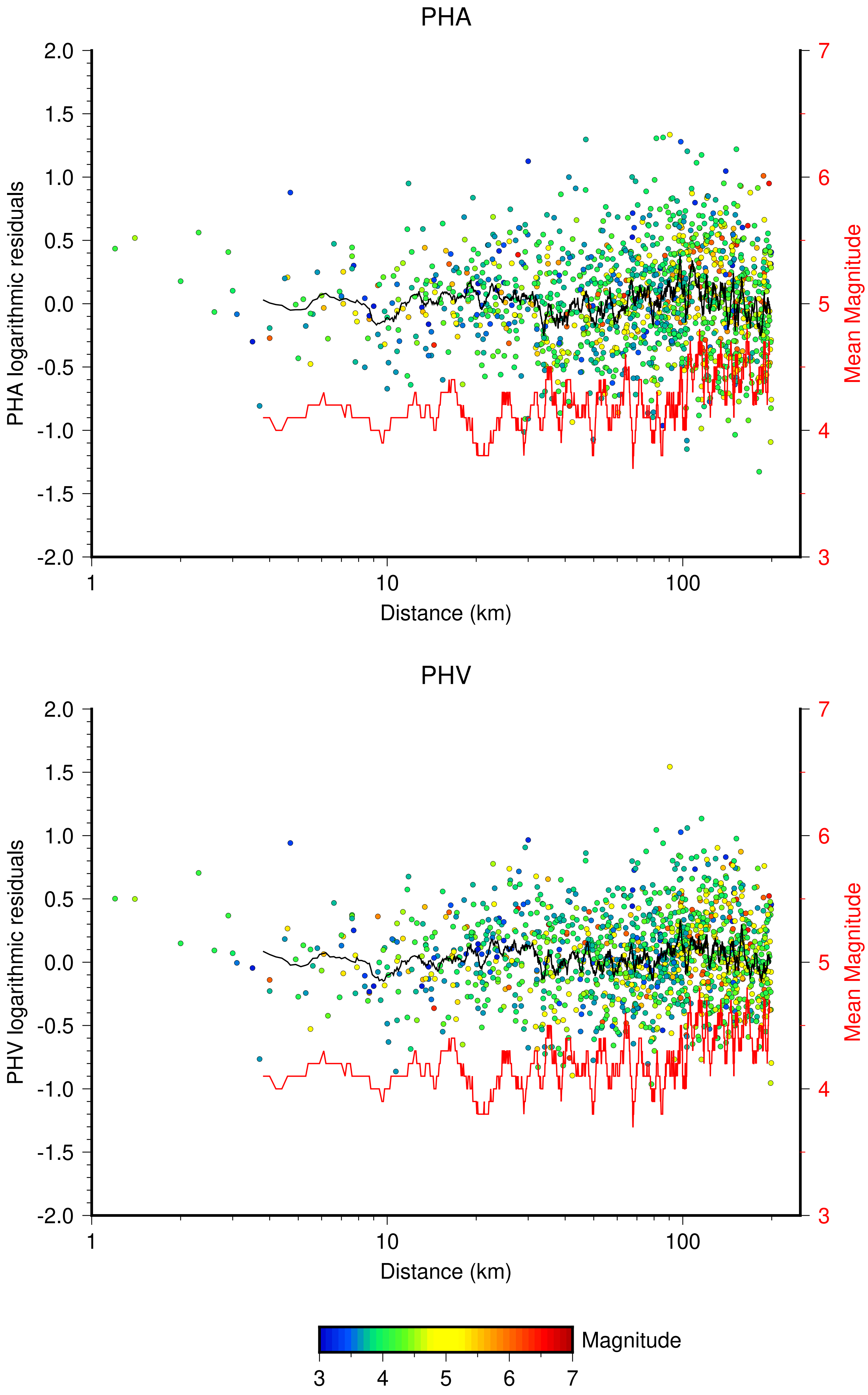
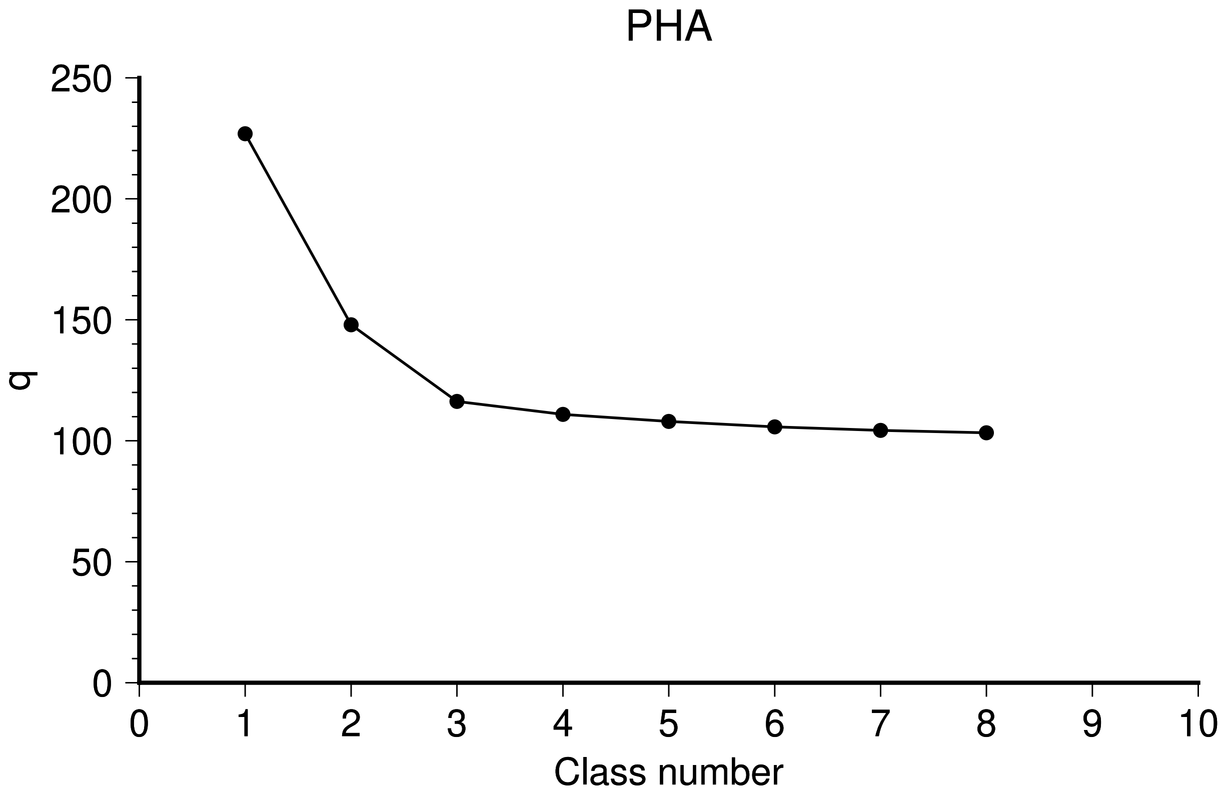

b)
Validation dataset
Magnitude
Distance (km)
Site class
A
B
C
7
6
5
4
3
0.1
1
10
100

Fig. 4



PHA
PHA logarithmic residuals
Mean Magnitude
Distance (km)
2.0
1.5
1.0
0.5
0.0
-0.5
-1.0
-1.5
-2.0
1
10
100
7
6
5
4
3
PHV
PHV logarithmic residuals
Mean Magnitude
Distance (km)
Magnitude
3
4
5
6
7


PHA
250
200
150
100
50
0
q
0
1
2
3
4
5
6
7
8
9
10
Class number

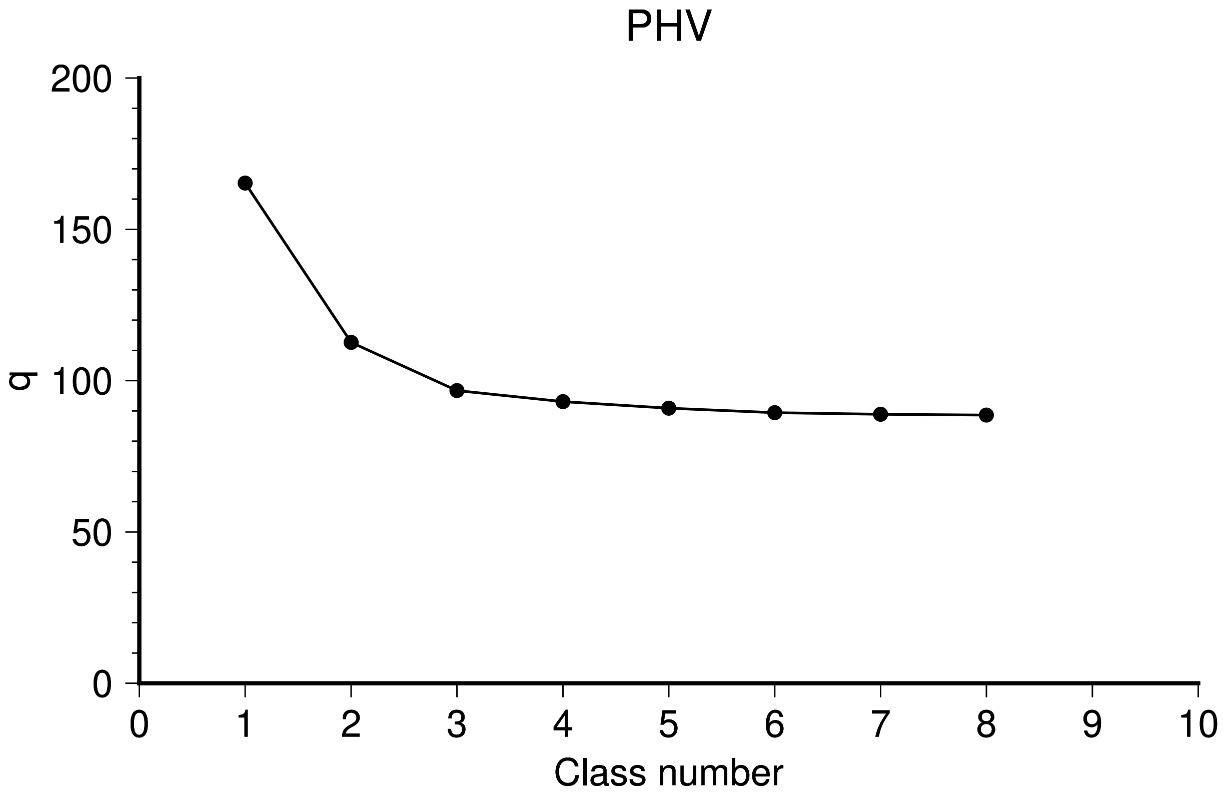

PHV
200
150
100
50
0
q
0
1
2
3
4
5
6
7
8
9
10
Class number

Fig. 6

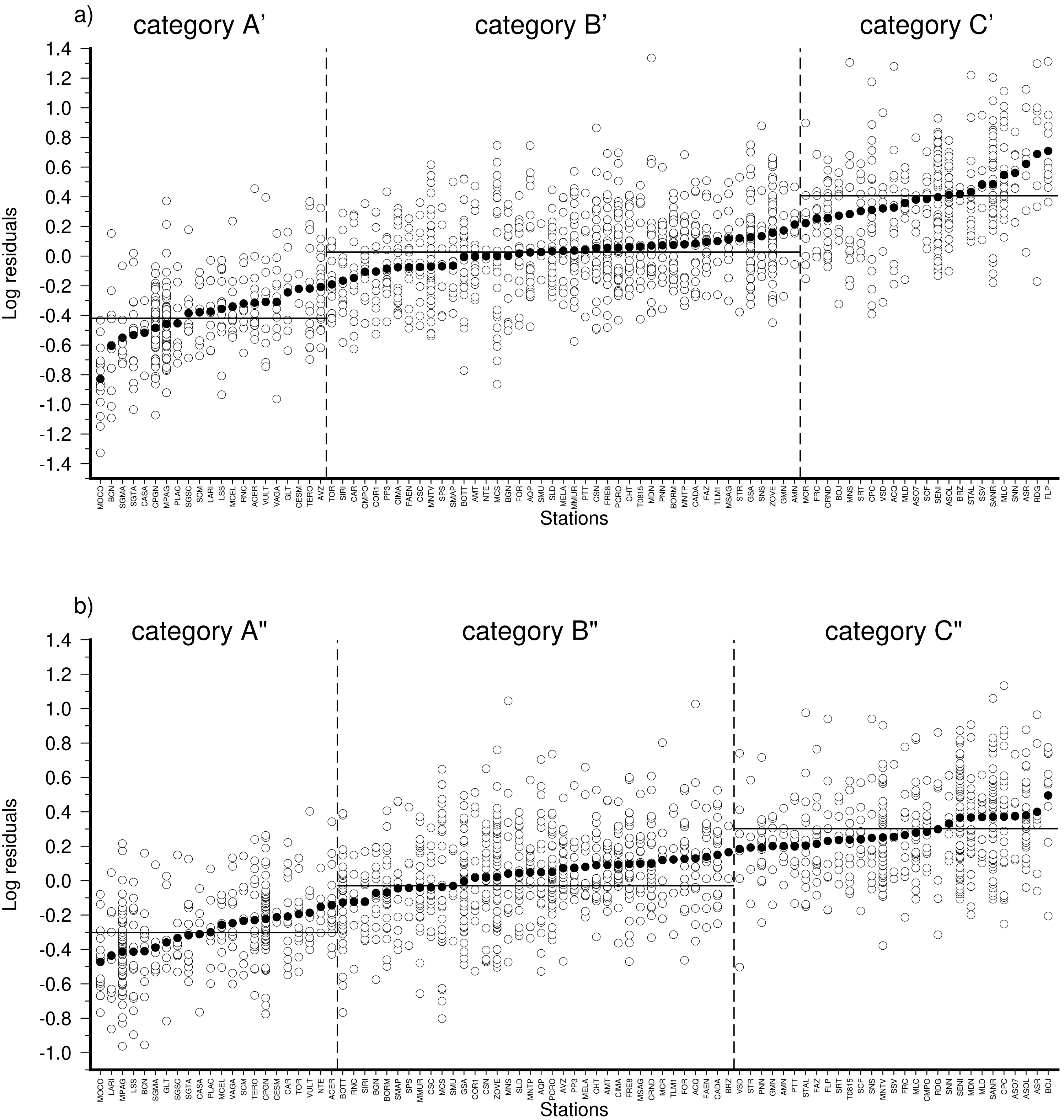

a)
category A'
category B'
category C'
Log residuals
Stations
MOCO BCN SGMA SGTA CASA CPGN MPAG PLAC SGSC SCM LARI LSS MCEL RNC ACER VULT VAGA GLT CESM TERO AVZ TOR SIRI CAR CMPO COR1 PP3 CIMA FAEN CSC MNTV SPS SMAP BOTT AMT NTE MCS BGN FOR AQP SMU SLD MELA MMUR PTT CSN FRE8 PCRO CHT T0815 MDN PNN BORM MNTP CADA FAZ TLM1 MSAG STR GSA SNS ZOVE GMN AMN MCR FRC CRND BOJ MNS SRT CPC VSD ACQ MLD ASO7 SCF SENI ASOL BRZ STAL SSV SANR MLC SNN ASR RDG FLP
b)
category A"
category B"
category C"
Log residuals
Stations
MOCO LARI MPAG LSS BCN SGMA GLT SGSC SGTA CASA PLAC MCEL VAGA SCM TERO CPGN CESM CAR TOR VULT NTE ACER BOTT RNC SIRI BGN BORM SMAP SPS MMUR CSC MCS SMU GSA COR1 CSN ZOVE MNS SLD MNTP AQP PCRO AVZ PP3 MELA CHT AMT CIMA FRE8 MSAG CRND MCR TLM1 FOR ACQ FAEN CADA BRZ VSD STR PNN GMN AMN PTT STAL FAZ FLP SRT T0815 SCF SNS MNTV SSV FRC MLC CMPO RDG SNN SENI MDN MLD SANR CPC ASO7 ASOL ASR BOJ

Fig. 7

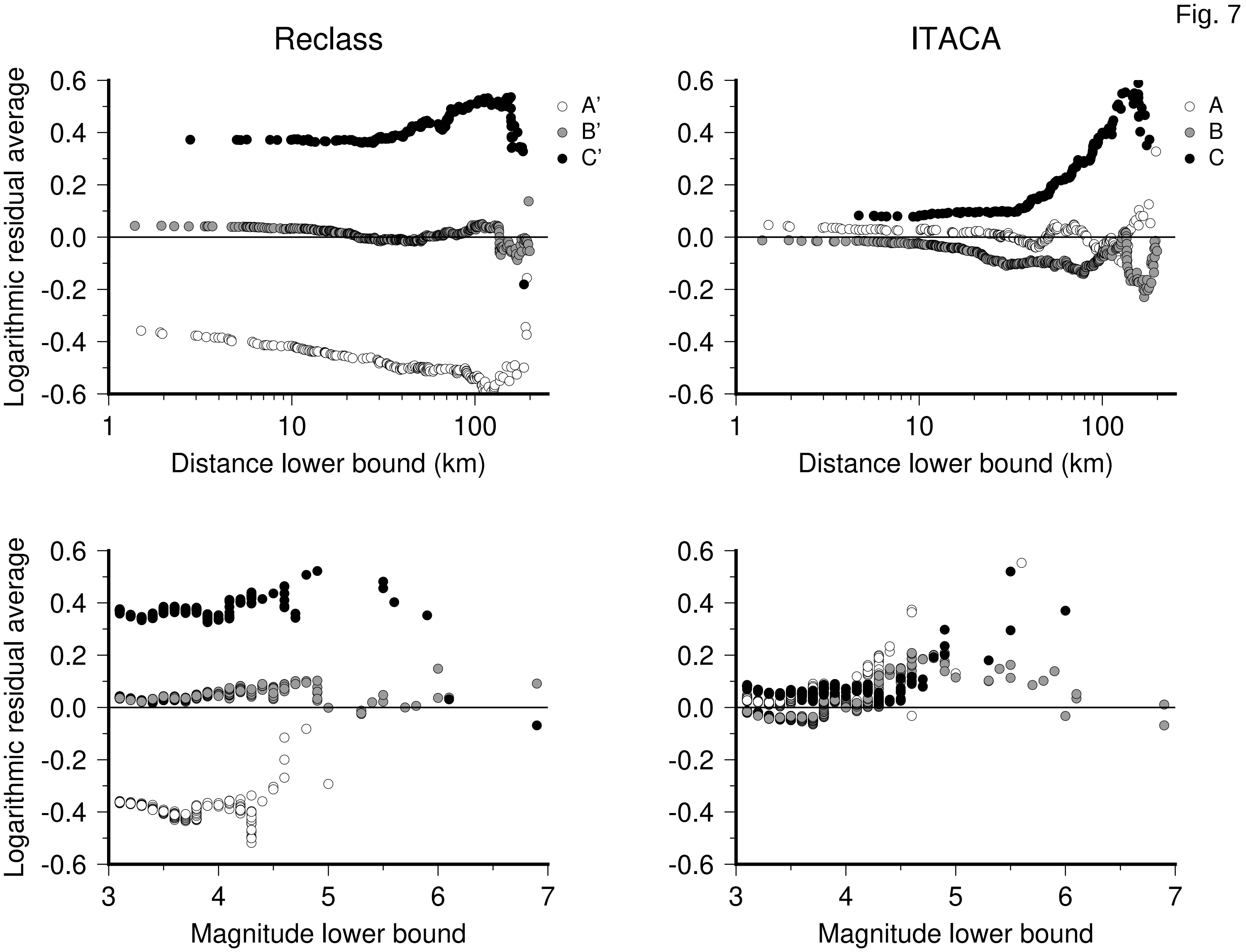
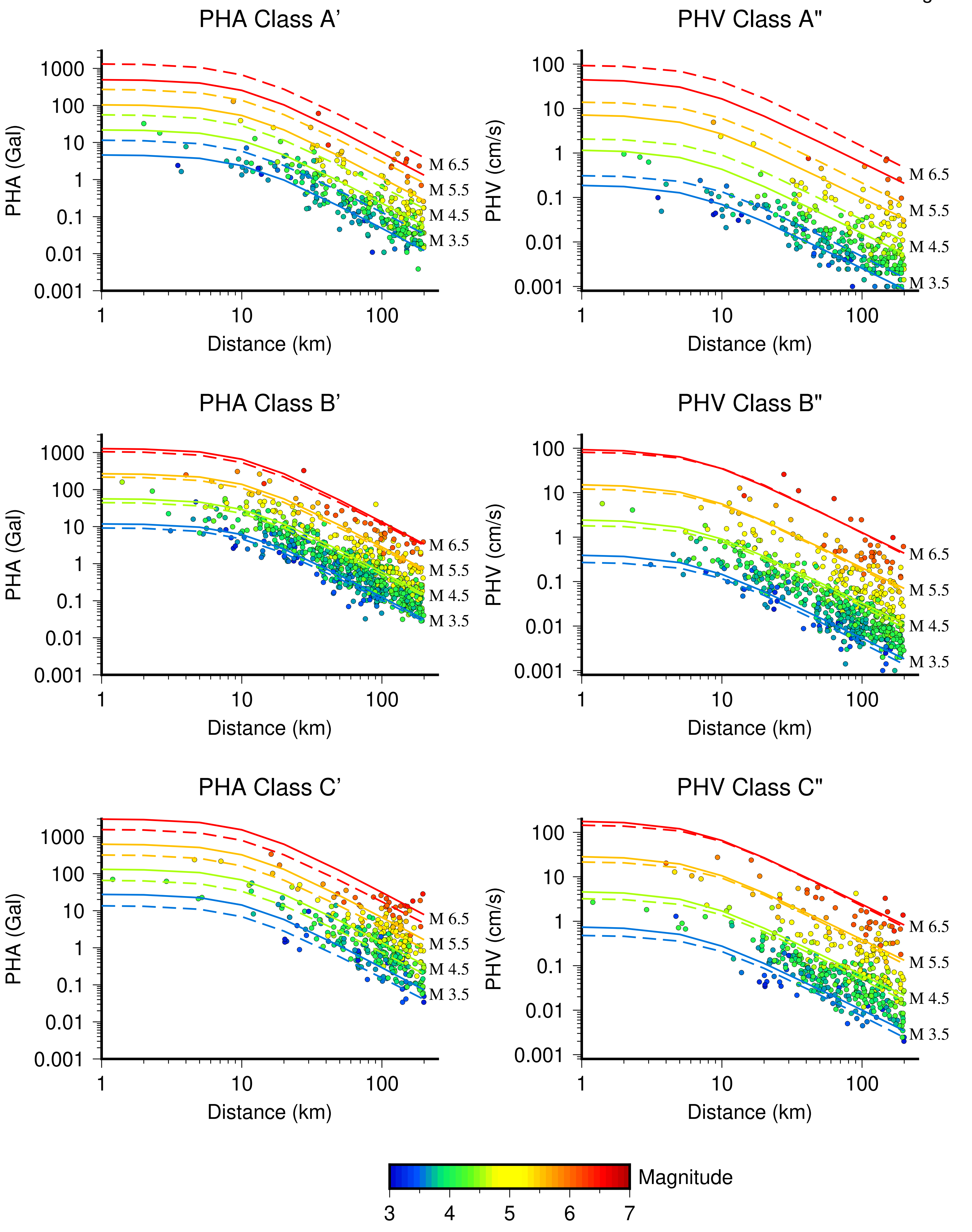
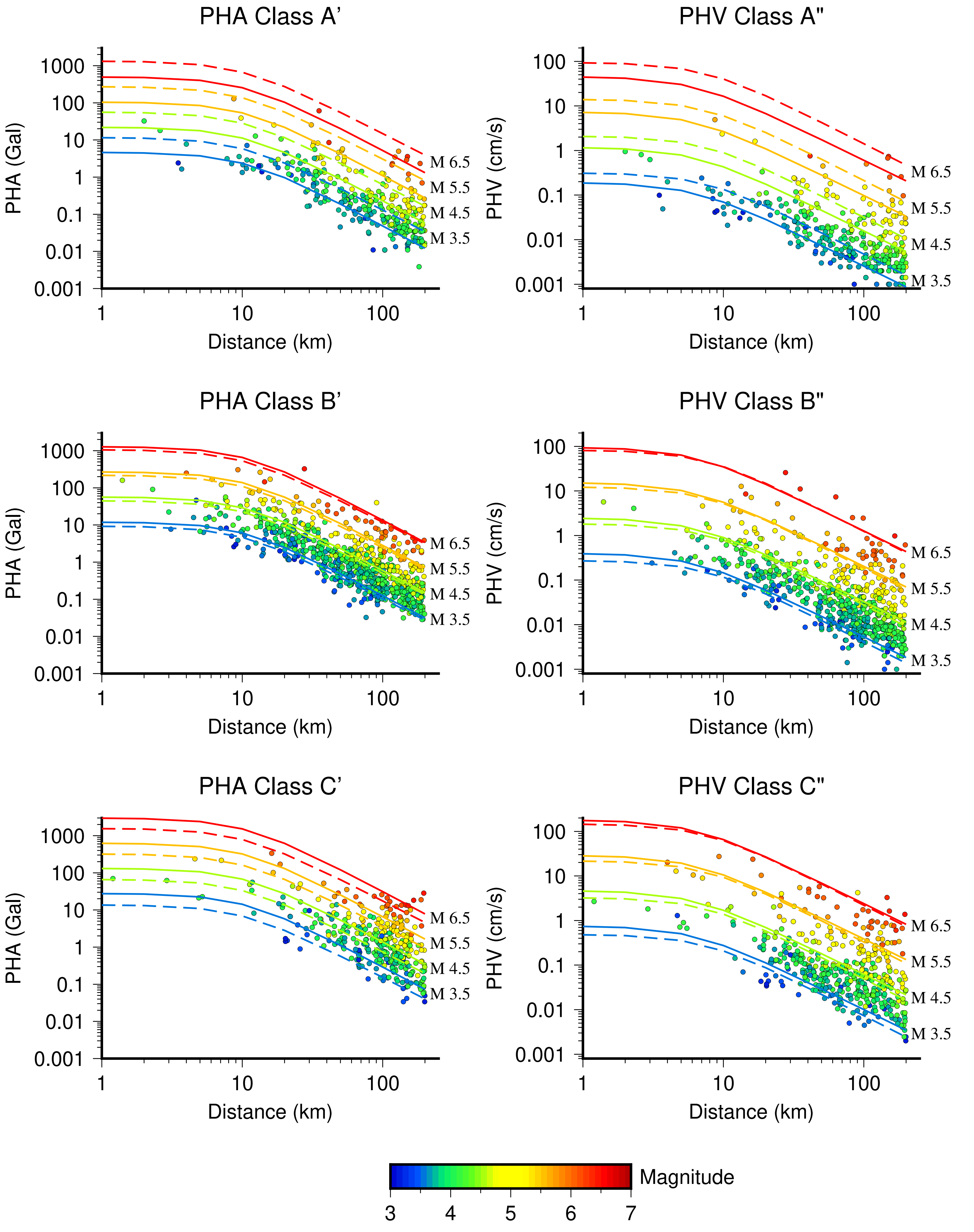
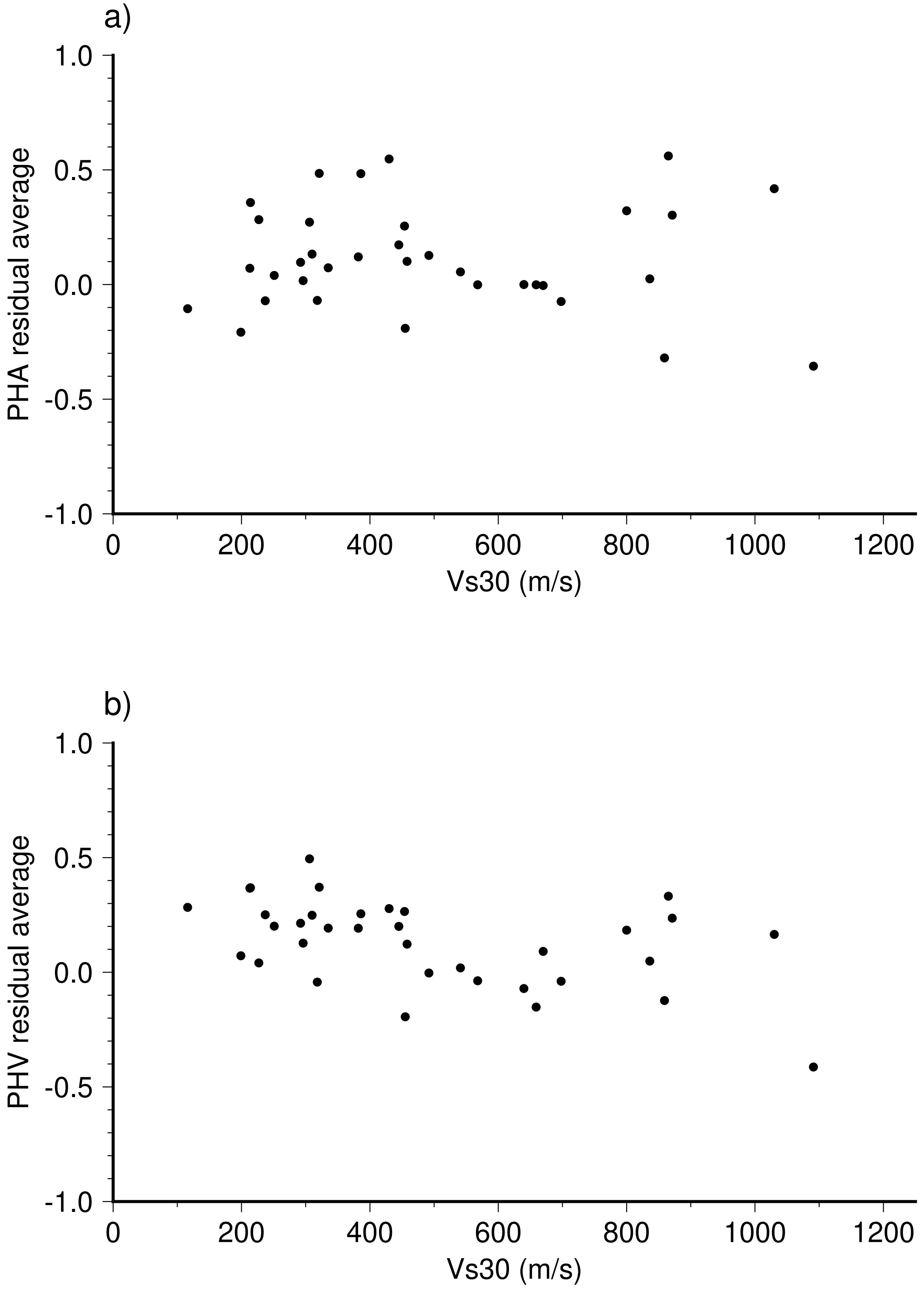

Reclass
ITACA
A'
B'
C'
A
B
C
Logarithmic residual average
Distance lower bound (km)
Magnitude lower bound
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
1
10
100
3
4
5
6
7

Fig. 8


PHA Class A'
PHV Class A"
PHA Class B'
PHV Class B"
PHA Class C'
PHV Class C"
PHA (Gal)
PHV (cm/s)
Distance (km)
M 6.5
M 5.5
M 4.5
M 3.5
1000
100
10
1
0.1
0.01
0.001
Magnitude
3
4
5
6
7


PHA Class A'
PHV Class A"
PHA Class B'
PHV Class B"
PHA Class C'
PHV Class C"
PHA (Gal)
PHV (cm/s)
Distance (km)
M 6.5
M 5.5
M 4.5
M 3.5
1000
100
10
1
0.1
0.01
0.001
Magnitude
3
4
5
6
7


a)
1.0
0.5
0.0
-0.5
-1.0
PHA residual average
0
200
400
600
800
1000
1200
Vs30 (m/s)
b)
1.0
0.5
0.0
-0.5
-1.0
PHV residual average
0
200
400
600
800
1000
1200
Vs30 (m/s)