Geophysical Journal International

*Geophys. J. Int.* (2022) **231,** 1117–1137
Advance Access publication 2022 July 1
GJI Seismology

https://doi.org/10.1093/gji/ggac245

# New reversible relationships between ground motion parameters and macroseismic intensity for Italy and their application in ShakeMap

Ilaria Oliveti [1], Licia Faenza[2] and Alberto Michelini[1]

[1] *Istituto Nazionale di Geofisica e Vulcanologia, sezione ONT, Rome,* 00143*, Italy. E-mail: ilaria.oliveti@ingv.it*
[2] *Istituto Nazionale di Geofisica e Vulcanologia, sezione Bologna, Bologna,* 40128*, Italy*

Accepted 2022 June 28. Received 2022 May 27; in original form 2021 October 27



## SUMMARY

We derived new, reversible relationships between macroseismic intensity (I), expressed in either the European Macroseismic (EMS-98) or the Mercalli–Cancani–Sieberg (MCS) scales and peak ground acceleration (PGA), peak ground velocity (PGV) and the spectral acceleration (SA) at 0.3, 1.0 and 3.0 s [SA(0.3), SA(1.0) and SA(3.0)] for Italy. We adopted the orthogonal distance regression technique to fit a quadratic function. This research aims to improve ground motion and intensity estimates for earthquake hazard applications, and for the calculation of shakemaps in Italy. To this end, the recently published INGe data set was used (https://doi.org/10.13127/inge.2). The new relations are:

$$I = 3.01 \pm 0.12 + 0.86 \pm 0.04 \log^2 \mathrm{PGA},\ \sigma = 0.30,\ \sigma_{\mathrm{PGA}} = 0.25,\ \sigma_I = 0.16$$

$$I = 4.31 \pm 0.15 + 1.99 \pm 0.18 \log \mathrm{PGV} + 0.58 \pm 0.18 \log^2 \mathrm{PGV},\ \sigma = 0.34,\ \sigma_{\mathrm{PGV}} = 0.31,\ \sigma_I = 0.15$$

$$I = 2.77 \pm 0.15 + 0.68 \pm 0.03 \log^2 \mathrm{SA}(0.3),\ \sigma = 0.31,\ \sigma_{\mathrm{SA}(0.3)} = 0.28,\ \sigma_I = 0.14$$

$$I = 3.00 \pm 0.28 + 0.91 \pm 0.55 \log \mathrm{SA}(1.0) + 0.51 \pm 0.20 \log^2 \mathrm{SA}(1.0),\ \sigma = 0.40,\ \sigma_{\mathrm{SA}(1.0)} = 0.38,\ \sigma_I = 0.14$$

$$I = 4.04 \pm 0.20 + 1.63 \pm 0.19 \log \mathrm{SA}(3.0) + 0.66 \pm 0.20 \log^2 \mathrm{SA}(3.0),\ \sigma = 0.38,\ \sigma_{\mathrm{SA}(3.0)} = 0.35,\ \sigma_I = 0.14$$

where PGA and SAs are expressed in cm $s^{-2}$ and PGV is expressed in cm $s^{-1}$. Tests performed to assess the robustness and the accuracy of the results demonstrate that adoption of quadratic relationships for this regression problem is a suitable choice within the range of values of the available data set. Comparison with similar published regressions for Italy evidences that the proposed relations provide statistically significant improved fits to the data. The new relations are also tested by inserting them in the ShakeMap system of the Italian configuration evidencing a significant improvement when compared to those implemented.

**Key words:** Europe; Earthquake ground motions; Seismicity and tectonics.

## 1 INTRODUCTION

Macroseismic intensity represents an important parameter in the engineering, seismological and loss modelling fields. Nowadays, despite abundant strong motion and seismometric data, macroseismic intensities still play an essential role. Their range of applicability covers different fields for several purposes. For example, only through the definition of macroseismic fields obtained from the study of historical earthquakes within their geological setting, and through the development of dedicated algorithms (e.g. Teramo *et al.* 1996; Gasperini *et al.* 2010; Beauval *et al.* 2010; Azzaro *et al.* 2011; Bakun *et al.* 2011; Traversa *et al.* 2018; Sbarra *et al.* 2019; Provost & Scotti 2020; Vannucci *et al.* 2021, amongst others) it is possible to estimate epicentral parameters and magnitude for earthquakes of the pre-instrumental era. This

makes macroseismic fields fundamental for building historical earthquake catalogues that extend back for several decades and even centuries. In addition, current efforts have been focusing on merging and homogenizing historical earthquake catalogues across national boundaries. To this end, two major international initiatives have been promoted—global synthesis of the pre-20th century earthquake history of the world (Albini *et al.* 2014), and the European Archive of Historical Earthquake Data, AHEAD (Locati *et al.* 2014; Rovida & Locati 2015). The latter in particular is the European node that provides key information on historical earthquakes (e.g. location, magnitude, Macroseismic Data Points, MDPs, in different scales) in Europe and surrounding regions complemented by the associated bibliographic sources.

Macroseismic field surveys are still conducted after significant earthquakes to maintain continuity and mapping consistency between past and recent earthquakes. That is, mapping of the intensities of recent earthquakes can be important for understanding those that occurred in the pre-instrumental era. This continuity can provide essential knowledge for impact assessment (Sokolov *et al.* 2010; Pittore *et al.* 2018) and loss estimations models (Cua *et al.* 2010; Tang *et al.* 2019) both crucial for disaster risk managers and civil protection authorities. At the same time, several efforts have been made to re-construct the distribution of the ground shaking for historical events at global (e.g. Allen *et al.* 2008) and local scales (e.g. Faenza *et al.* 2013). Another interesting and valuable application is the use of macroseismic data as a reference for the selection of the most appropriate ground motion model in low-to-moderate seismicity areas (Villani *et al.* 2019; Tang *et al.* 2019).

In general, macroseismic intensities have been found to facilitate the exchange of earthquake impact information from seismologists to the population and vice versa. The increasing popularity of citizen scientists related activities in Earth Science (Lee *et al.* 2020) has led to the development and dissemination of online macroseismic questionnaires, such as the U.S. Geological Survey (USGS) questionnaire (Wald *et al.* 2012; Quitoriano & Wald 2020); the Italian 'Hai Sentito Il Terremoto?' HSIT database (Tosi *et al.* 2007); the LastQuake smartphone app developed by the European Mediterranean Seismological Centre (EMSC) for global earthquake eyewitnesses (Bossu *et al.* 2017). To this regard, the advent of digital seismology has brought to the application of earthquake hazard and impact estimation tools, which use macroseismic data as input, such as the ShakeMap system (Wald *et al.* 1999b; Worden *et al.* 2020) and its derivative product Prompt Assessment of Global Earthquakes for Response (PAGER, Earle *et al.* 2009), Earthquake Loss Estimation Routine (ELER, Corbane *et al.* 2017) and earthQuake Loss Assessment for Response and Mitigation (QLARM, Trendafiloski *et al.* 2011).

Finally, and this is probably the most appealing aspect as it is more related to this research, macroseismic data are also used in earthquake engineering applications. For example, intensities are used for the modelling of vulnerability and for the calibration of intensity prediction equation (IPE) which are valuable in mapping the spatial distribution of damage. In general, intensity prediction relationships have been developed for a specific region and, for example, Bakun (2006) have proposed a model suitable for Western North America; Bakun & Scotti (2006) for France; Baumont *et al.* (2018) for metropolitan France; Bindi *et al.* (2011) for Central Asia; Oros *et al.* (2019) for Romania; Pasolini *et al.* (2008) for Italy and Stromeyer & Grünthal (2009) for Central Europe. The only exception to these relations calibrated at regional or national level is the IPE proposed by Allen *et al.* (2012), which can be adopted for active crustal regions since calibrated using data acquired on a much broader area within a similar tectonic setting. In addition, models for correlating macroseismic intensity data with peak ground motion parameters (PGMs) are playing an increasing role in assessing seismic hazard and in loss modelling. Equations for estimating intensity from the available PGM records (Ground-Motion-to-Intensity Conversion Equations—GMICEs) are, for example, adopted in the USGS-ShakeMap software procedure wherever the maps are converted to instrumental intensity (e.g. Wald *et al.* 2006; Kästli & Fäh 2006; Michelini *et al.* 2008; Schlupp 2016). In the same way, equations in the other direction (Intensity-to-Ground-Motion Conversion Equations—IGMCEs) are usually used for assessing the size of historical earthquakes when few or no recorded ground motions are available, and it is of interest to estimate PGM from the given intensity at a location. These conversions are also applied when "Did You Feel it?" (DYFI) data are employed as inputs into ShakeMap (Quitoriano & Wald 2020). Again, as with IPEs, studies are regionalized, for example, Bilal & Askan (2014) determined a relationship suitable for Turkey; Cramer (2020) for Central and Eastern North America; Du *et al.* (2020) for China; Lesueur *et al.* (2013) for France; Moratalla *et al.* (2021) for New Zealand; Tselentis & Danciu (2008) for Greece; Kaka & Atkinson (2004) for North America; Karim & Yamazaki (2002) for Japan; Worden *et al.* (2012) for California and Wu *et al.* (2003) for Taiwan, just to name a few of the works available in the literature. Caprio *et al.* (2015) derived a global relationship by combining published data sets from different regions. In general, the contributing data sets used to obtain GMICEs differ in their definitions of PGMs, depending on the specific purpose and application of each study: peak ground acceleration (PGA) and velocity (PGV) are often used, but also integral values such as Housner intensity and Arias intensity are reported in the literature. The limitation of the earlier studies is the non-reversibility of the proposed relationships (e.g. Margottini *et al.* 1992; Decanini *et al.* 1995; Panza *et al.* 1997; Wald *et al.* 1999a; Faccioli & Cauzzi 2006). With this in mind, most recent studies have adopted reversible GMICEs, especially for ShakeMap application, using Orthogonal Distance Technique (ODR; Boggs *et al.* 1988) or total least squares (TLS; Golub & Van Loan 1980). Looking at Italy, we can observe that several correlation models based on cross-matching the DataBase of Macroseismic observations of Italy (DBMI), whose intensity data are expressed in either the European Macroseismic Scale 1998 (EMS-98) or the Mercalli–Cancani–Sieberg (MCS) scales, and the ITalian ACceleration Archive (ITACA) have been proposed. Faenza & Michelini (2010, 2011) determined linear relationships for Italy between the MCS (Sieberg 1930) intensity data (DBMI04; Stucchi *et al.* 2007) and the associated PGA, PGV and SA at 0.3, 1.0 and 2.0 s provided by the ITACA1.0 strong motion database (Luzi *et al.* 2008) by adopting the ODR. Zanini *et al.* (2019) introduced reversible macroseismic intensity—PGMs (PGA, PGV, PGD, Arias intensity and Housner intensity) conversion equations based on the EMS-98 (Grünthal 1998). Masi *et al.* (2020) proposed bilinear relationships between macroseismic data (EMS-98 and MCS scales) and PGMs such as PGA, PGV and Housner Intensity by adopting the TLS method. Cataldi *et al.* (2021) derived a new set of linear regression relations between integer MCS intensity classes and eight PGMs (PGA, PGV, PGD, pseudo-SA at 0.3, 1.0 and 2.0 s, Arias intensity and Housner intensity), using the reference data set previously compiled by Faenza & Michelini (2010) and updated

with the recent earthquakes. Among the most recent relationships proposed for Italy, only those by Gomez-Capera *et al.* (2018, see also Gomez-Capera *et al.* 2020, for an update) have been defined in terms of non reversibility. In particular, in this latter study, the non-linear correlation equations between PGA, PGV and SA at 0.2, 0.3, 1.0 and 2.0 s, and macroseismic intensity (and vice versa) were determined by adopting separate regression relationships in the two directions.

All the relationships cited above refer to regressions in which the macroseismic intensity is correlated to the logarithm of a PGM measure since the recorded data are nearly lognormally distributed. The main differences among them depend on the adopted regression approach and on the macroseismic scale to be considered. Concerning the latter, according to some authors (e.g. Codermatz *et al.* 2003) no practical difference exists between the EMS and MCS scales for intensities smaller than 9. By contrast, Molin *et al.* (1995) observed that, for intensities higher than degree 7, EMS and MCS scales may differ by one degree unit or more. It is to be noted that Musson *et al.* (2010) came to the conclusion that assigning EMS intensities to the MCS values does not lead to significant discrepancies since the two scales are mostly comparable. Furthermore, the same authors added that the generally higher MCS intensity assignments compared to those in the EMS scale that are available in the literature, could result from the way in which the scale has been interpreted.

The aim of this work is to determine new reversible correlation relationships between recorded PGM and spectral accelerations (SA), and reported macroseismic intensities in terms of either MCS or EMS-98. These relations are thought to be relevant for both hazard assessment studies and for the generation of shakemaps in Italy. Our study can be viewed as the continuation of the work of Faenza & Michelini (2010, 2011) by using the ODR regression approach and the much larger INGe data set (https://doi.org/10.13127/inge.2; Oliveti *et al.* 2021, 2022). This data set has been recently assembled from the Italian Macroseismic Database DBMI15 (Locati *et al.* 2021) and the Engineering Strong-Motion (ESM) accelerometric data bank flatfile (Lanzano *et al.* 2018). To this regard, an extensive procedure of selection and revision combined with the recent occurrence of destructive earthquakes makes this data set much larger and more homogeneous when compared to that used previously by Faenza & Michelini (2010) and also to those assembled and adopted by the above cited authors. In addition, the availability of a data set like INGe makes it possible to perform benchmark studies investigating the correlations between ground motion parameters and macroseismic intensities.

The novel and main element of our work is the more robust and statistically sound approach that was followed to develop the GMICEs. In summary, the main steps of our analysis are:

(i) adoption of the strategy introduced by Kuehn & Scherbaum (2010) to assign more realistic estimates of the standard deviation associated to the PGM data. This method compensates for the lack of data at the higher intensity classes;

(ii) application of shuffle-split cross-validation analysis (e.g. Chollet 2018) to evaluate the proposed relationships while avoiding data overfitting. This technique, lent from machine learning good practices, subdivides the data set into training and test sets—the former is used to determine the coefficients of the relationships whereas the latter is used to validate them;

(iii) exploration of several modelling approaches by adopting the Akaike Information Criterion (AIC) (Anderson *et al.* 1998) as best fit criterion. We found that adoption of a quadratic relation for the regressions provided the best-fitting correlations between PGMs and macroseismic intensities;

(iv) performance comparison between our resolved GMICEs and those presented in other studies (tested on INGe) through the use of the mean squared error (MSE) and the standard deviation ($\sigma_r$) of the residuals as metrics;

(v) use of the conversion equations obtained in our work for the generation of shakemaps in order to evaluate their applicability in Italy.

## 2 METHODOLOGY

### 2.1 The PGM-macroseismic intensity data set

The recently published INGe data set (available at https://doi.org/10.13127/inge.2) that includes observations for Italy in the time span 1972–2016 has been used for the regression analysis described below; full detail on its compilation is provided by Oliveti *et al.* (2021, 2022). In essence, INGe includes PGM-intensity pairs of all the localities reporting intensity data which are located within 3 km from the strong motion stations that recorded the data. INGe contains pairs for $I \geq 3$ since the parent Italian Macroseismic Database, DBMI15, (Locati *et al.* 2021) does not include any MDP with lower values within a 3-km radius from the stations for our selected earthquakes. To this regard and in the future, a broader reevaluation of the HSIT database (Sbarra *et al.* 2019), while accounting for the different definitions therein provided for the 'locality' of the MDPs, may reveal valuable to constrain better the low intensity values. The data set includes 519 data pairs with intensities ranging between 3 and 11 and PGA between 850.8 and 0.4 cm s$^{-2}$. PGM parameters were taken from the maximum between the two horizontal components. In our analysis, we have selected only the MDP locality closest to the recording station. The station-to-MDP distances range between 0.01 km and nearly 3 km, showing an average value of about 1 km and a median value of 0.73 km. The resulting data set used in the analysis consists of a total of 323 associated PGM-intensity pairs derived from 65 earthquakes and 227 stations (Fig. 1). Considering only PGM-intensity pairs with the shortest distance ensures that the possible difference in site conditions between the station and the macroseismic point is the minimum acceptable for the analysis. This choice also seems to be the most effective in light of the different spatial definitions of the two sets of data: for the stations, the values are pointwise, whereas, for the macroseismic data, the values are always represented in a comprehensive perspective for the entire location. See Oliveti *et al.* (2022) for details. The events, located at distances within 300 km from the stations, feature magnitudes in the range 4.1–6.8 and depths in the range 0–55 km. The resulting macroseismic intensity

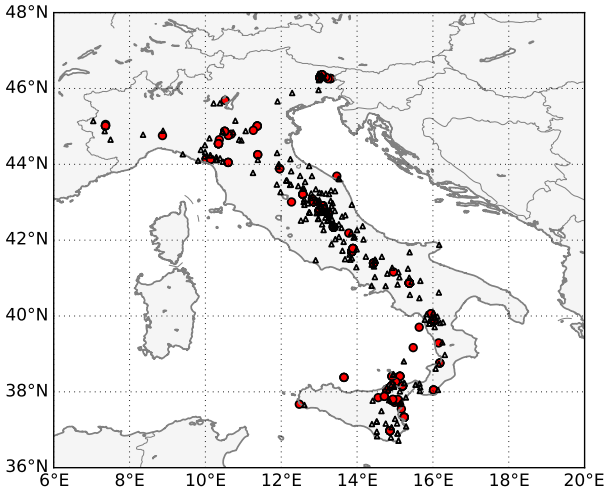


**Figure 1.** Spatial distribution of the selected seismic events (red dots) and of the stations (open solid triangles) used to assemble the intensity-PGM pair data set.

values range between 3 and 10. The intensity is usually provided in the MCS scale but, especially for recent earthquakes, the EMS-98 has been used, as reported in DBMI15 (Locati *et al.* 2021).

### 2.2 Data processing

Data binning is a customary pre-processing technique in several scientific fields whenever it is necessary to reduce the effects of data outliers or observational errors. It is also widely used in combination with the orthogonal distance regression (ODR, e.g. Fu *et al.* 2017; Gatzsche *et al.* 2018; Pylak *et al.* 2021).

In our analysis, considering the dispersion of the intensity data, we followed the same procedure used previously by various authors (Faenza & Michelini 2010, 2011; Bilal & Askan 2014; Zanini *et al.* 2019; Gomez-Capera *et al.* 2020) to minimize the uncertainties in the values by arranging the intensity observations in classes and, for each level, by calculating the mean value of the recorded ground motion parameter. Moreover, the within-intensity-bin normalization is entirely consistent with other studies of the ground motion to intensity relationships in which an orthogonal regression is applied to ensure their reversibility (Faenza & Michelini 2010, 2011; Worden *et al.* 2012; Caprio *et al.* 2015; Zanini *et al.* 2019; Cataldi *et al.* 2021). As a result, the data were binned into 0.5 intensity intervals and for each class the mean was calculated as:

$$\mu_k = \frac{1}{n_k}\sum_{i=1}^{n_k} \log \mathrm{PGM}_i, \tag{1}$$

where $n_k$ is the number of data points for each intensity class $k$.

The use of the logarithm is motivated by the distribution of log PGM values: as shown in Fig. 2, for each intensity bin of the whole data set, the standard normal variable (i.e. $z = \frac{x-\mu}{\sigma}$) for the logarithmic values is in agreement with the theoretical standard normal distribution. In contrast, the great majority of the distributions about the arithmetic means are concentrated on the lower side of the mean value.

Since intensities have been arranged in classes spaced by 0.5 (ranging from 3 to 10), the standard deviations of each intensity bin have been set equal to 0.5 as a conservative but reasonable value. In addition, rather than adopting the individual standard deviation of the average ground motion values for each intensity class, we calculated a common standard deviation of the distribution of each ground motion parameter over the different intensity classes. This solution was adopted because only a few MDPs are available for some intensity levels (cf. classes 9 and 10 versus class 5 in Fig. 3) inhibiting the robust estimation of the standard deviation at each intensity class value. Following the approach introduced by Kuehn & Scherbaum (2010), the common standard deviation was determined by eq. (2):

$$\sigma_{\mathrm{com}} = \sqrt{\frac{1}{N-1}\sum_{k}\sum_{i=1}^{n_k}(\log \mathrm{PGM}_{ik} - \mu_k)^2}, \tag{2}$$

where $N$ is the total number of data points and $k$ is the intensity level. The calculated values are listed in Table 1.

### 2.3 The orthogonal distance regression technique

The ODR technique used in this study warrants fully reversible ground motion-to-intensity conversion equations. The approach is the same followed by Faenza & Michelini (2010, 2011) in which observational errors on both dependent and independent variables are taken into account. This methodology consists of solving the weighted ODR problem by minimizing the residual sum of the squares of the weighted

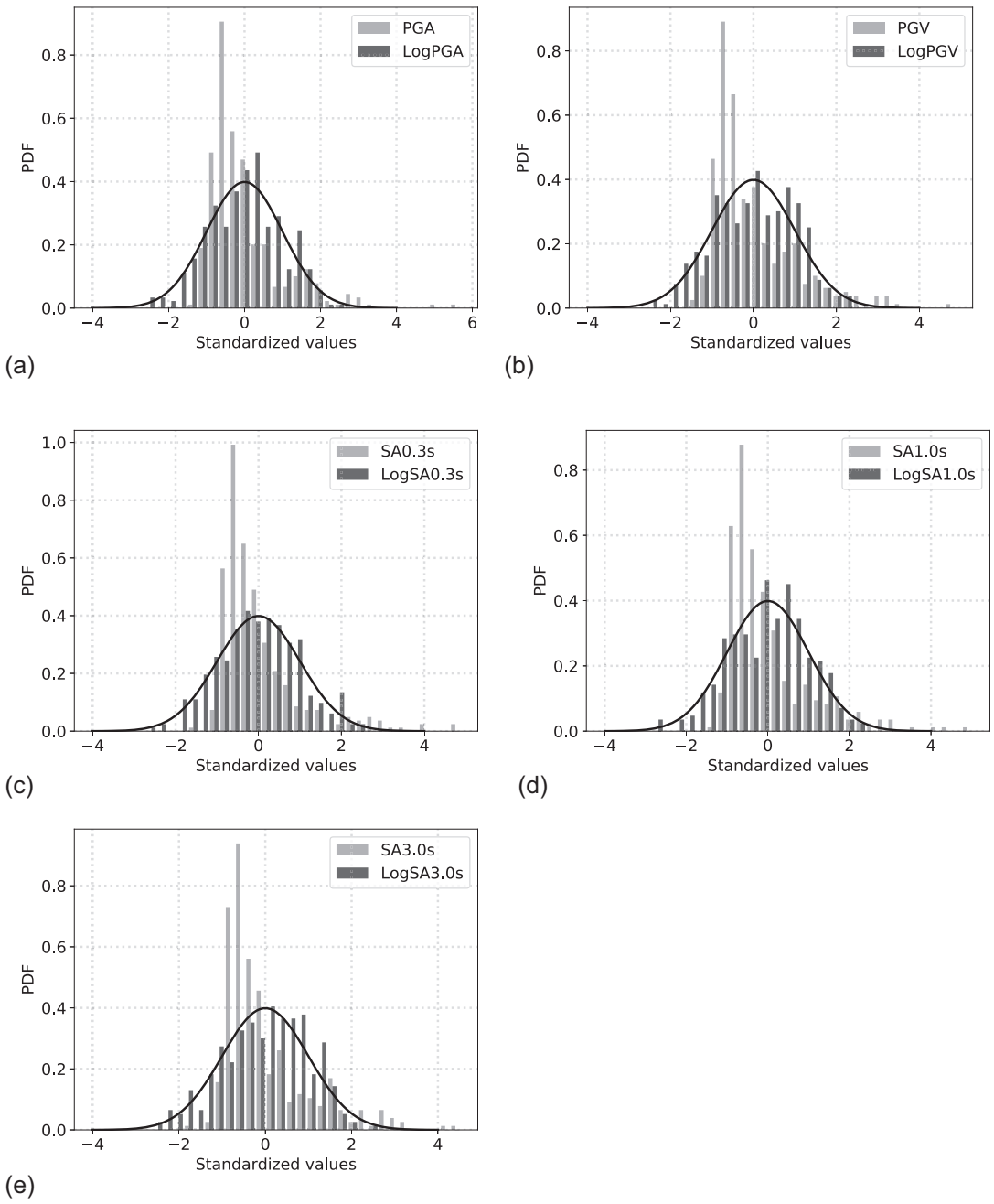


**Figure 2.** (a) PGA, (b) PGV, SA at periods (c) 0.3 s, (d) 1.0 s and (e) 3.0 s value distribution. For each intensity bin, the data set is normalized to obtain standardized values, having zero mean and unit standard deviation. In each panel, both the original data (light grey) and the logarithm in base 10 of those (dark grey) are represented. As reference, the expected normal distribution curves are also shown as solid black line.

orthogonal distances between each data point and the curve described by the model equation (Boggs *et al.* 1988). Given a simple data set consisting of $n$ points $(x_i, y_i)$, $i = 1,..., n$, where $x_i$ and $y_i$ are the independent and dependent variables, respectively, the ODR algorithm for an explicit function can be expressed as:

$$\min\left(\sum_{i=1}^{n}\left(w_{yi}\cdot\epsilon_i^2 + w_{xi}\cdot\delta_i^2\right)\right) \tag{3}$$

subject to the constraints:

$$y_i = f(x_i + \delta_i; \beta) - \epsilon_i \qquad i = 1, ..., n \tag{4}$$

where $w_{xi}$ and $w_{yi}$ are the user input weights of $x_i$ and $y_i$, $\delta_i$ and $\epsilon_i$ are the residual of the corresponding $x_i$ and $y_i$, and $\beta$ is the fitting parameters vector. In this work, we used the algorithm developed by Boggs *et al.* (1987)—a FORTRAN code wrapped within the SciPy Python module

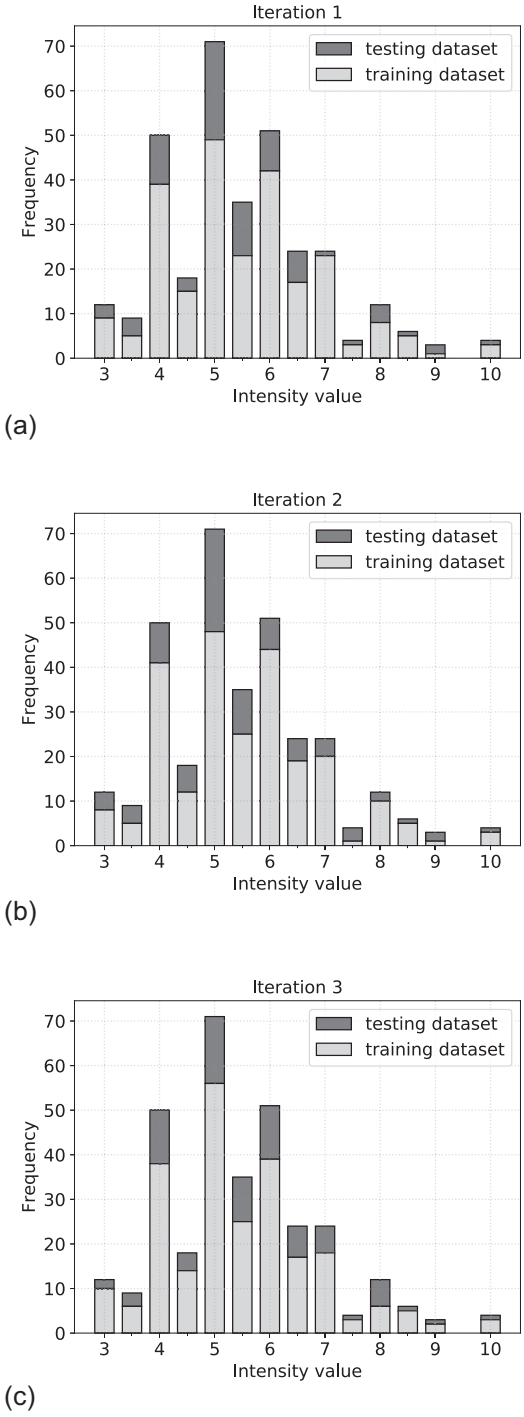


**Figure 3.** Distribution of strong-motion data in terms of macroseismic intensities for each cross-validation iteration.

**Table 1.** Regression coefficients ($a$, $b$, $c$), their standard deviation ($\sigma_a$, $\sigma_b$, $\sigma_c$) and common standard deviation of the distribution of the PGM parameters for all intensity classes ($\sigma_{\text{com}}$: see eq. 2)

| Functional form: $I = a \pm \sigma_a + b\log \text{PGM} \pm \sigma_b + c\log^2\text{PGM} \pm \sigma_c$ | | | | | | | |
|---|---|---|---|---|---|---|---|
| PGM | $a$ | $b$ | $c$ | $\sigma_a$ | $\sigma_b$ | $\sigma_c$ | $\sigma_{\text{com}}$ |
| PGA | 3.01 | – | 0.86 | 0.12 | – | 0.04 | 0.41 |
| PGV | 4.31 | 1.99 | 0.58 | 0.15 | 0.18 | 0.18 | 0.42 |
| SA(0.3) | 2.77 | – | 0.68 | 0.15 | – | 0.03 | 0.44 |
| SA(1.0) | 3.00 | 0.91 | 0.51 | 0.28 | 0.55 | 0.20 | 0.50 |
| SA(3.0) | 4.04 | 1.63 | 0.66 | 0.20 | 0.19 | 0.20 | 0.60 |

**Table 2.** The averaged error estimation over all three trials to get total effectiveness of the proposed ODR model.

| | | $R^2$ score | | |
|---|---|---|---|---|
| PGA | PGV | SA(0.3) | SA(1.0) | SA(3.0) |
| 0.94 | 0.92 | 0.93 | 0.87 | 0.83 |

(http://www.scipy.org), which adjusts both fitting parameters and values of the independent variable in the iterative process. In particular, we chose the standard deviations presented in Section 2.2 as weights to account for different variances of the observations.

### 2.4 Shuffle-split cross-validation

To tackle the problem of analysing a limited size data set and with the goal of obtaining robust model coefficient estimates without the occurrence of overfitting, we applied a shuffle-split cross-validation procedure—a methodology routinely adopted in machine learning analysis. In general, cross-validation is advocated to assess the performance of a model. To this end, the data set is split in $K$ parts, and in each of the $K$ iterations, a different part is used as a *testing data set* (called *validation data set*), while the remaining portions are treated as a *training data set*. In each iteration, predictions are made, and the final prediction is their average (e.g. Chollet 2018).

With shuffle-split cross-validation, we randomly sampled the entire data set during each iteration to generate a *training data set* and a *testing data set*. This approach allowed us to control how large the two sets should be for each iteration in order to ensure a good representation of the whole data set. In this application, samples were first shuffled and then split into three pairs of train and test sets. In each of these partitions, 1/4 represents the chosen proportion of the data set included in the test split, while the remaining 3/4 constitutes the *training data set* size. Thus, we had three sets of data to train and test our model. Fig. 3 shows the distribution of the samples for the three subsets of the testing data in terms of macroseismic intensity. The data appear to be distributed among all the intensity classes demonstrating that each subset can be representative of the whole data set. The error metric computed to determine the overall accuracy of the model is the average of the $R^2$ scores obtained at each run of shuffle-split cross-validation. $R^2$ score or coefficient of determination is, by definition, the ratio of the difference between the actual and predicted value and the difference between the actual and mean of the target variable. Specifically, at each iteration, we performed the orthogonal regressions on the training data, and used the modelled parameters to estimate the $R^2$ score by taking the differences between the predicted and observed intensity values of the test set. That is, the model parameters are estimated using the three training sets and each time validated against the corresponding test set. The prediction error is calculated by taking the mean average of the $R^2$ scores of the three test sets (see Table 2).

## 3 APPLICATION

In our implementation of the ODR regression (eqs 3–4), we used a quadratic functional form. Once applied to the pre-processed data we obtained quadratic relationships between macroseismic intensity and base-10 logarithm of PGMs. The following fully reversible ground motion-to-intensity conversion equations, derived using binned values on the training data sets, were obtained for PGA, PGV, SA(0.3), SA(1.0) and SA(3.0), such that

$$I = 3.01 \pm 0.12 + 0.86 \pm 0.04 \log^2 \mathrm{PGA},\ \sigma = 0.30,\ \ \sigma_{\mathrm{PGA}} = 0.25,\ \ \sigma_I = 0.16 \tag{5}$$

$$I = 4.31 \pm 0.15 + 1.99 \pm 0.18 \log \mathrm{PGV} + 0.58 \pm 0.18 \log^2 \mathrm{PGV},\ \sigma = 0.34,\ \ \sigma_{\mathrm{PGV}} = 0.31,\ \ \sigma_I = 0.15 \tag{6}$$

$$I = 2.77 \pm 0.15 + 0.68 \pm 0.03 \log^2 \mathrm{SA}(0.3),\ \sigma = 0.31,\ \ \sigma_{\mathrm{SA}(0.3)} = 0.28,\ \ \sigma_I = 0.14 \tag{7}$$

$$I = 3.00 \pm 0.28 + 0.91 \pm 0.55 \log \mathrm{SA}(1.0) + 0.51 \pm 0.20 \log^2 \mathrm{SA}(1.0),\ \sigma = 0.40,\ \ \sigma_{\mathrm{SA}(1.0)} = 0.38,\ \ \sigma_I = 0.14 \tag{8}$$

$$I = 4.04 \pm 0.20 + 1.63 \pm 0.19 \log \mathrm{SA}(3.0) + 0.66 \pm 0.20 \log^2 \mathrm{SA}(3.0),\ \sigma = 0.38,\ \ \sigma_{\mathrm{SA}(3.0)} = 0.35,\ \ \sigma_I = 0.14 \tag{9}$$

where $\sigma$ is the standard deviation of each regression model, $\sigma_{\mathrm{PGM}}$ and $\sigma_I$ are the standard deviations of the independent and dependent variables, respectively. The unit of PGA and SAs is cm s$^{-2}$, whereas for PGV is cm s$^{-1}$.

Whereas the functional form of the regression equations is $I = a + b\log \mathrm{PGM} + c\log^2 \mathrm{PGM}$, the inverse empirical conversion relations between macroseismic intensity and PGMs corresponds to $\log \mathrm{PGM} = (-b + \sqrt{b^2 - 4c(a - I)})/2c$ for $\mathrm{I} \geq (4ac - b^2)/4$ (see Fig. S1, Supporting Information).

The coefficients of the equations (see Table 1) were obtained by taking the mean of the outcome of the regression coefficients obtained using the individual training data sets as explained in Section 2.4. Moreover, it is worth pointing out that eqs (5) and (7) are binomial quadratic expressions because the standard deviation of the *log* term is greater than the value of the regression coefficient.

In Table 1, we report the main parameters derived for each of the eqs (5)–(9) (i.e. $a$, $b$, $c$ regression coefficients, coefficient's standard deviation $\sigma_a$, $\sigma_b$, $\sigma_c$) and the common standard deviation of the distribution of the PGM ($\sigma_{\mathrm{com}}$: see eq. 2).

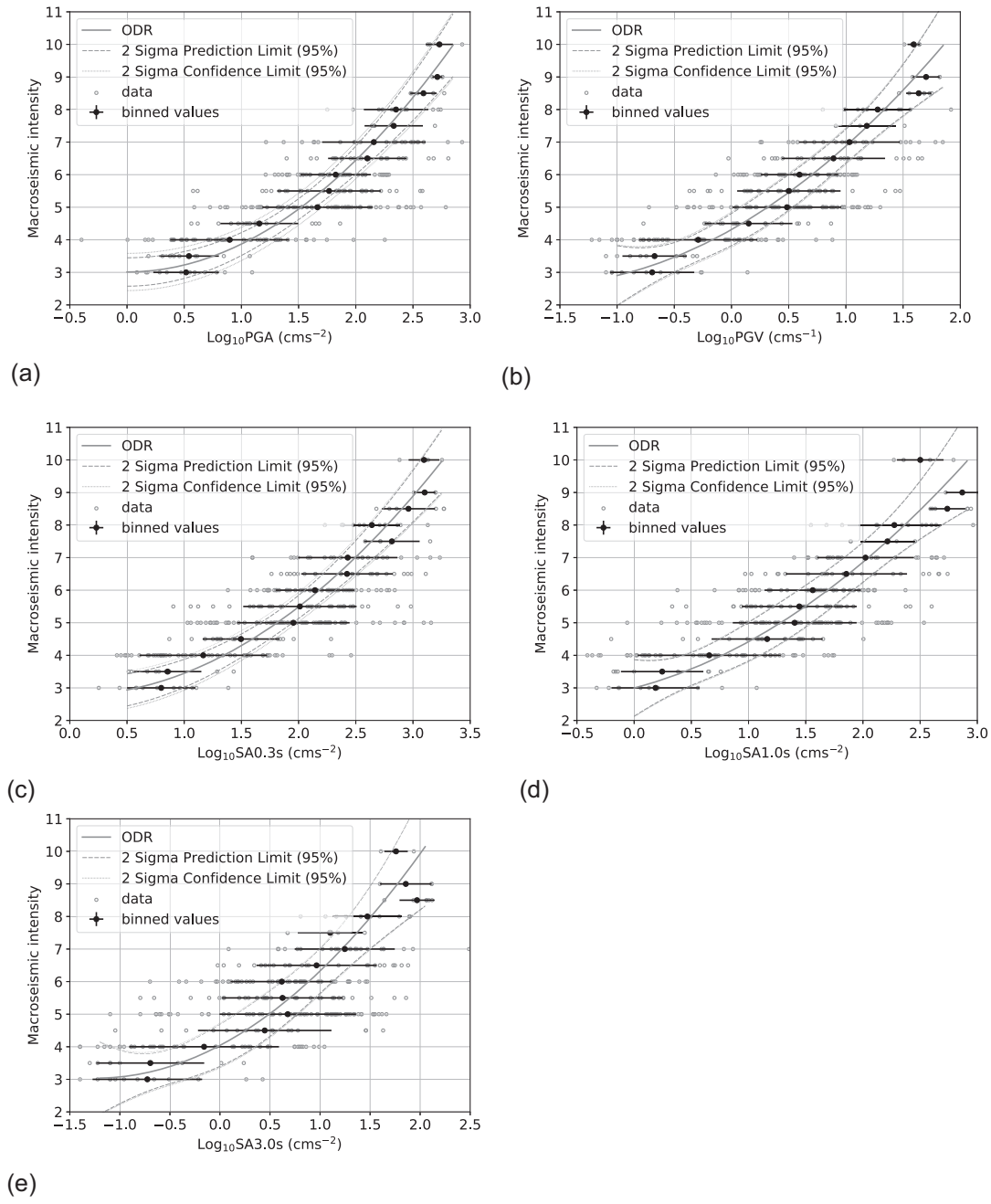


**Figure 4.** Macroseismic intensities versus the logarithm in base 10 of (a) PGA, (b) PGV, SA at periods (c) 0.3 s, (d) 1.0 s and (e) 3.0 s. Data (grey open circle), binned values (black solid circle), standard deviations (black error bars), quadratic function (solid grey line) and the associated 95 per cent confidence (dotted lines) and prediction (dashed lines) intervals are also shown.

A graphical representation of eqs (5)–(9) is displayed in Fig. 4: the grey solid line represents the averaged ODR model, while the data (grey open circles) and their binned values (black solid circles) and standard deviations (black horizontal error bars), shown as a general representation only, refer to the whole original data set.

In order to explore further our choice of adopting a quadratic function, we have tested other functional forms (linear and exponential) derived using binned values on the training data sets. For each functional form, we have calculated the AIC value by considering the perpendicular distance between the 323 data set points and the fitted curve. Fig. 5 shows, as an example for the PGA-intensity pairs, the comparison between the quadratic model with the linear and exponential regressions within the validity range of the models. It is possible to observe that the linear model does not provide an optimal correlation between the PGM parameters and the macroseismic intensity. By contrast, the exponential and the quadratic functions better capture the trend of the data over the entire range of validity of the input data set. AIC values of all empirical relationships are available in Table 3.

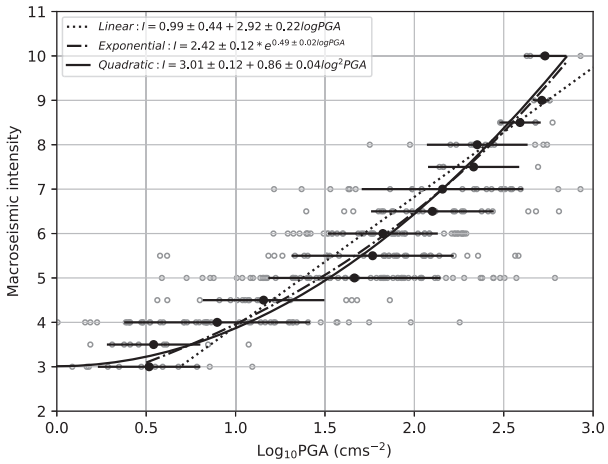


**Figure 5.** Comparison between the proposed macroseismic intensity versus PGA relationship and the linear and exponential regressions performed on data.

**Table 3.** Comparison of the linear, exponential and quadratic relationships in terms of AIC value for each PGM parameter.

| PGM | AIC | | | | | |
|---|---|---|---|---|---|---|
| | Linear | | Exponential | | Quadratic | |
| | $I$[3,10] | $I \geq 7$ | $I$[3,10] | $I \geq 7$ | $I$[3,10] | $I \geq 7$ |
| PGA | −532 | −631 | −588 | −673 | −599 | −675 |
| PGV | −530 | −619 | −577 | −664 | −577 | −670 |
| SA(0.3) | −514 | −638 | −563 | −670 | −571 | −671 |
| SA(1.0) | −451 | −545 | −492 | −593 | −492 | −594 |
| SA(3.0) | −318 | −513 | −358 | −556 | −368 | −557 |

Additional analysis was carried out to explore the behaviour of the calculated relationships at the high end of the intensity values. Improved accuracy in this range of values is crucial not only for generating accurate shakemaps in the near source region but also, for example, for seismic hazard studies that exploit the rich set of intensity measurements of the historical earthquakes of the pre-instrumental era. To this end, we selected PGM-intensity pairs with intensity values $\geq 7$ (50 on a total of 323) and we calculated the AIC value of all empirical relationships. The best fit of the quadratic model at the largest values of intensity is confirmed by the lowest AIC values listed in Table 3.

Finally, we would like to mention that we have also tried to fit a cubic model to the purpose of comparison with the results of the models presented above. We have found that the uncertainties of the coefficients of both the quadratic and the cubic terms were very large to indicate the high indeterminacy of the resolved values (cf. the $\sigma_c$ and $\sigma_d$ values in Table S1, Supporting Information, with the uncertainties of the coefficients in eqs 5–9).

## 4 COMPARISON WITH PREVIOUS STUDIES

### 4.1 Regression relations

The relationships between macroseismic intensity and PGM parameters of eqs (5)–(9) (Fig. 4) were compared to a subset of regression functions that were derived by Faenza & Michelini (2010, 2011), Gomez-Capera *et al.* (2018), Masi *et al.* (2020), Gomez-Capera *et al.* (2020) and Cataldi *et al.* (2021). Noteworthy, the above studies were chosen because they also adopted data sets compiled by joining the same macroseismic and strong motion data sets although by using different criteria. It is thus expected a good quantitative consistency between the results of these studies and ours. With respect to the specific macroseismic scale to be considered, we selected works based on intensity values that are fully or mainly expressed in MCS scale, with the exception of a few recent earthquakes for which only EMS-98 data points are available in DBMI.

Figs 6 and 7 show the comparison of the regressions for PGA and PGV. These two figures can be considered representative also for the relations of SA at the 0.3 and 1.0 s periods, respectively. In general, we find that the rate of change of our quadratic relationship between PGA and PGV, and macroseismic intensity increases more gradually compared to the other models. Specifically, in Fig. 6 we observe that, the relationships of Faenza & Michelini (2010) and Cataldi *et al.* (2021) predict larger intensity values than ours in the range 5 cm s$^{-2}$ < PGA < 200 cm s$^{-2}$ and 6 cm s$^{-2}$ < PGA < 300 cm s$^{-2}$, respectively. The maximum difference is about 0.6 intensity units at about 30 cm s$^{-2}$. By contrast, the regression model of Masi *et al.* (2020) provides lower macroseismic intensities starting from about 25 cm s$^{-2}$, but approaches

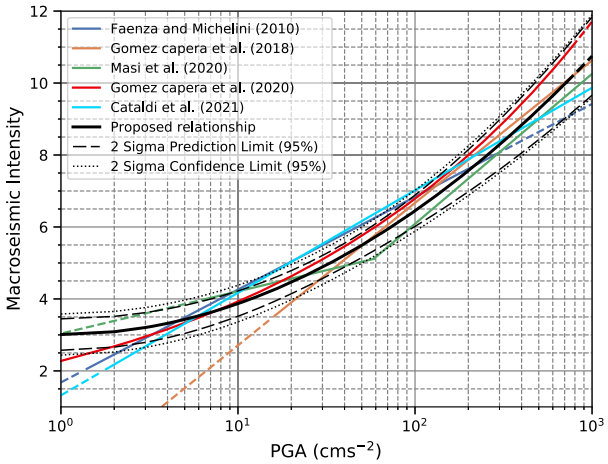


**Figure 6.** Comparison between existing and proposed relations for PGA. The solid lines show the functions for PGM and intensity ranges constrained by their respective data sets; the dashed lines show the GMICEs when extrapolated beyond their data sets. For our conversion equation, the associated 95 per cent confidence (dotted lines) and prediction (dashed lines) intervals are also shown. The plotted relationships are listed in the legend of each panel.

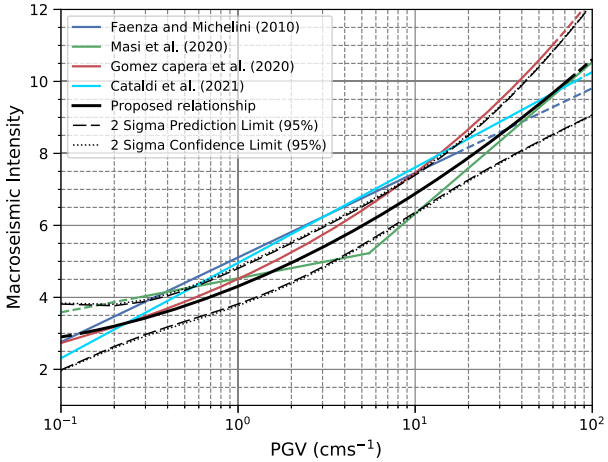


**Figure 7.** Comparison between existing and proposed relations for PGV. The solid lines show the functions for PGM and intensity ranges constrained by their respective data sets; the dashed lines show the GMICEs when extrapolated beyond their data sets. For our conversion equation, the associated 95 per cent confidence (dotted lines) and prediction (dashed lines) intervals are also shown. The plotted relationships are listed in the legend of each panel.

the same values as ours at the very large PGAs. At higher values of PGA ($> 200$ cm s$^{-2}$), we observe that the relationship by Gomez-Capera *et al.* (2020) predicts intensity values progressively larger than those obtained in this study.

The same basic behaviour described for PGA is also observed for PGV (Fig. 7). The regressions proposed by Faenza & Michelini (2010) and Cataldi *et al.* (2021) display higher values of intensities in the range comprised between 0.14 and 30 cm s$^{-1}$, and between 0.25 and 50 cm s$^{-1}$, respectively. By contrast, the relation developed by Masi *et al.* (2020) features intensity values smaller than ours between 1.5 cm s$^{-1}$ and the very large values of PGV where the two nearly coincide. We also observe that the relation implemented by Gomez-Capera *et al.* (2020) predicts much higher intensity values than that proposed in this work at large values of PGV (i.e. intensity 11 at 60 cm s$^{-1}$).

In summary, the regressions described by eqs (5)–(9) provide predictions of the intensity values that lie in between those determined using similar sets of intensity–ground motion pairs. The observed differences among all relationships can be explained by the criteria adopted to pair the intensity values with the recorded ground motion parameters, by the chosen regression technique and by the selected functional form. The general behaviour observed for PGA appears replicated in great part in the regressions performed using PGV. The relevant difference seems to be in the trend of the exponential curve proposed by Gomez-Capera *et al.* (2020) which, while similar to our relation at intensities about 3.5, it overestimates our data set at high intensity values.

### 4.2 Differences between predicted values

In this subsection, we appraise the results obtained through the direct comparison between observed and predicted values for both intensity and PGM parameters in terms of PGA, PGV, SA at $T = 0.3$, 1.0 and 3.0 s for all the relationships presented in the previous subsection. The same set of 323 data pairs has been used for all the relations tested. To this end, and to the purpose of analysing the variance of the error of the regressions, we have proposed residual plots of the difference between (i) observed and predicted macroseismic intensities

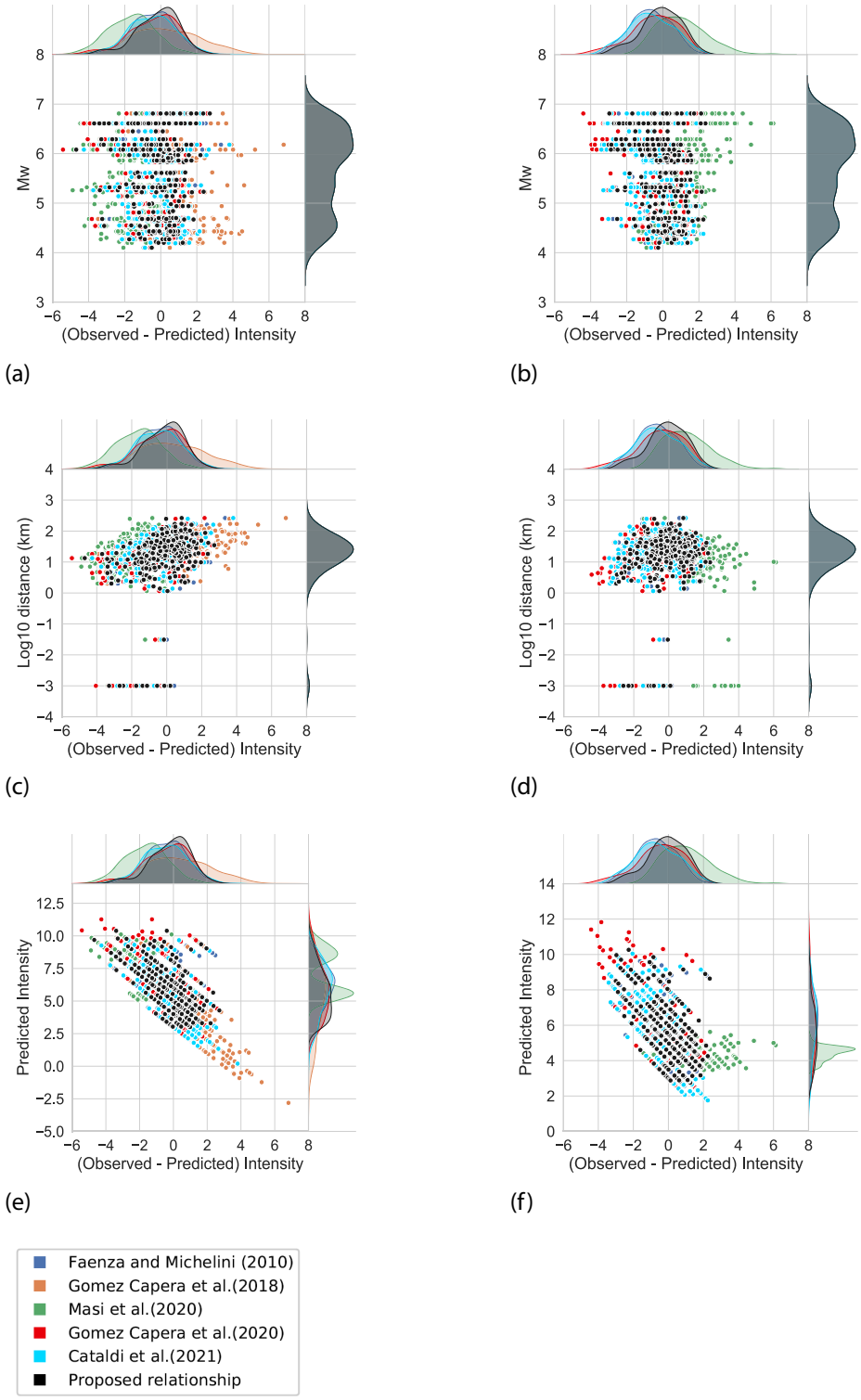


**Figure 8.** Plots of the intensity residuals obtained from (a), (c) and (e) PGA and (b), (d) and (f) PGV represented against moment magnitude, Joyner–Boore distance and the predicted macroseismic intensity, respectively. The selected relationships are listed in the legend.

with respect to moment magnitude, Joyner–Boore distance and the predicted intensities themselves (Figs 8, 10 and 12), and (ii) observed and predicted PGMs again as function of moment magnitude, Joyner–Boore and regression predicted PGM values (Figs 9, 11 and 12). To predict the PGM values from intensity, we exploited the bi-uniqueness property of orthogonal relations, whereas for the non-reversible relationships, we adopted the formulations proposed by Gomez-Capera *et al.* (2018) and Gomez-Capera *et al.* (2020). To this regard, we note that it is not possible to apply the strategy adopted in these last two cases in codes like ShakeMap (Wald *et al.* 2006; Kästli & Fäh 2006; Michelini *et al.* 2008; Worden *et al.* 2020; Michelini *et al.* 2020), which exploit the reversibility between intensity and PGM characteristic

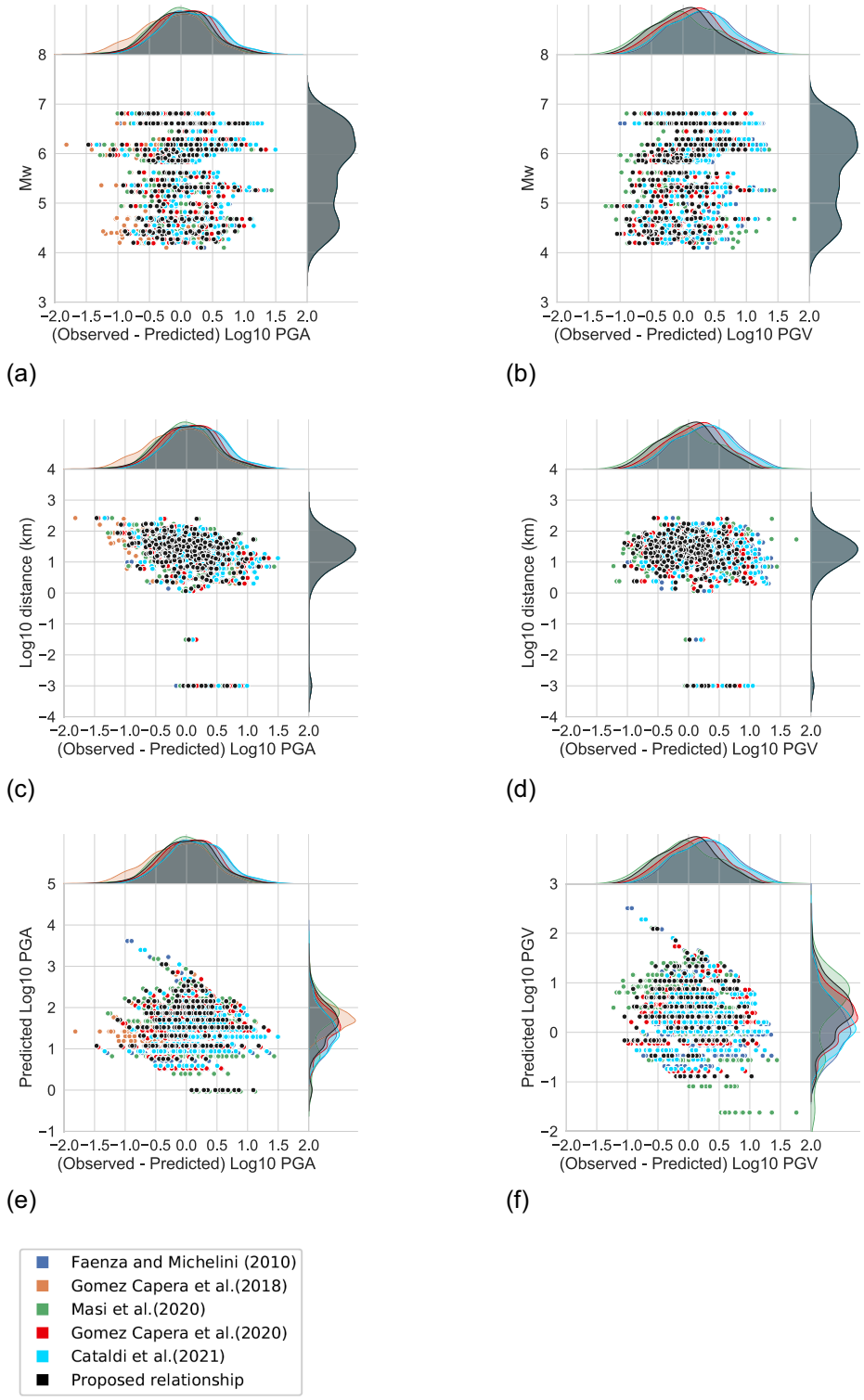


**Figure 9.** Plots of (a), (c) and (e) PGA and (b), (d) and (f) PGV residuals obtained from the intensities and represented against moment magnitude, Joyner–Boore distance and the corresponding predicted PGM value, respectively. The selected relationships are listed in the legend.

of the GMICE regressions to avoid inconsistent results. To summarise the results for each regression relationship, we determined the MSE and the standard deviation of the residuals ($\sigma_r$) (Tables 4 and 5). MSE and $\sigma_r$ values have been evaluated through the following expressions:

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i) \tag{10}$$

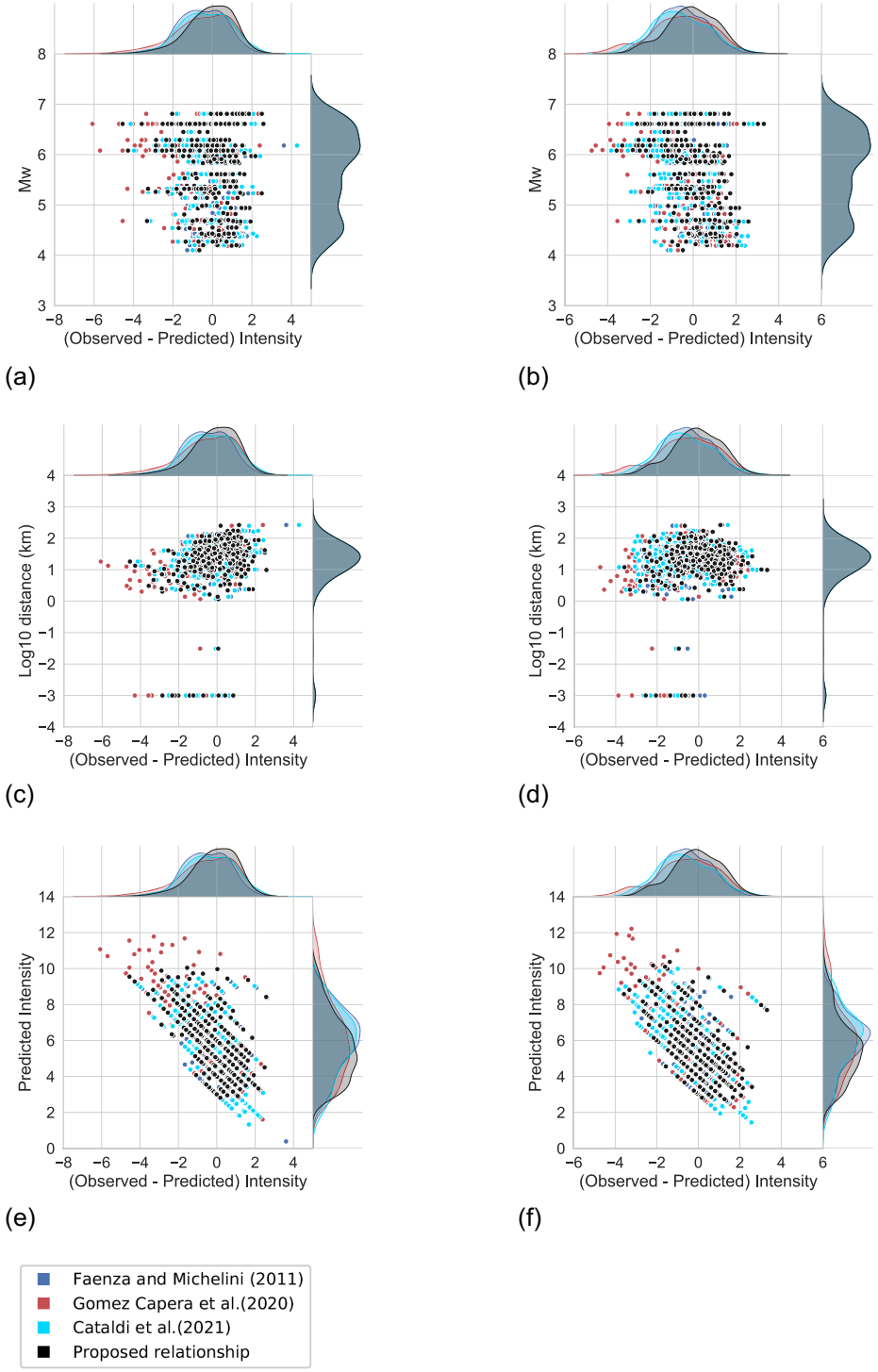


**Figure 10.** Plots of the intensity residuals obtained from (a), (c) and (e) SA(0.3) and (b), (d) and (f) SA(1.0) represented against moment magnitude, Joyner–Boore distance and the predicted macroseismic intensity, respectively. The selected relationships are listed in the legend.

$$\sigma_r = \sqrt{\frac{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}{n-1}} \tag{11}$$

where $Y_i$ and $\hat{Y}_i$ ($i = 1, 2, \ldots \cdot\cdot, n$) are the observed data and the predicted values, respectively, and $\bar{Y}$ is the mean of $Y_i$. Overall, this analysis estimates the error in the predicted values (eq. 10) and its dispersion (eq. 11) of all the relationships examined in this study.

The results shown in Fig. 8 for intensities predicted from PGA and PGV indicate that the relationships determined in this work are performing better than others as confirmed by both MSE and standard deviation values listed in Table 4. In fact, our models show both the

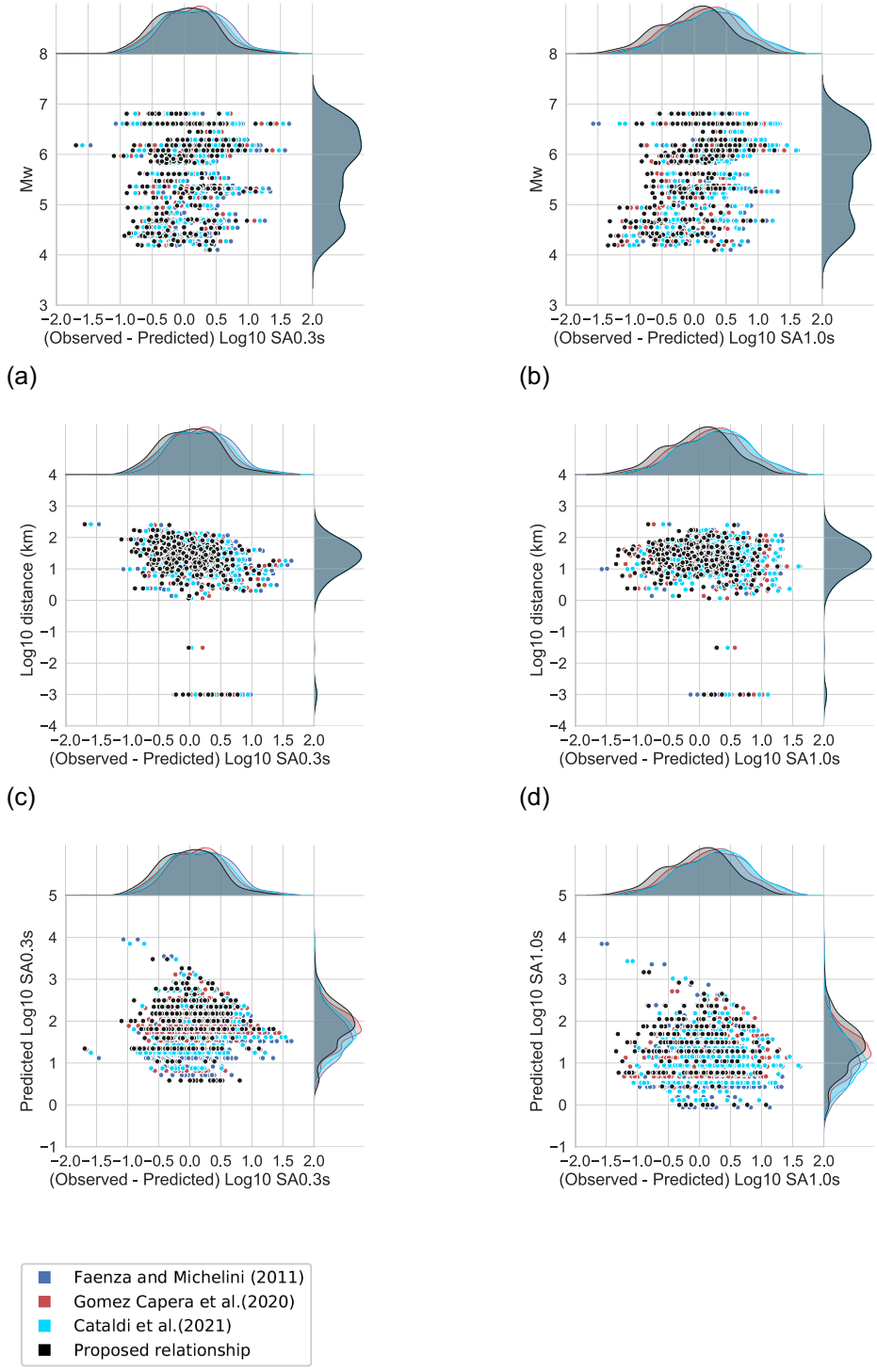


**Figure 11.** Plots of (a), (c) and (e) SA(0.3) and (b), (d) and (f) SA(1.0) residuals obtained from the intensities and represented against moment magnitude, Joyner–Boore distance and the corresponding predicted PGM value, respectively. The selected relationships are listed in the legend.

lowermost error in the predicted values ($MSE_{PGA} = 1.45$ and $MSE_{PGV} = 1.22$) and a small dispersion ($\sigma_{rPGA} = 1.19$, $\sigma_{rPGV} = 1.11$) as illustrated in Fig. 8 where our error appears to be almost normally distributed around zero. In contrast, the corresponding values of MSE obtained using the other authors' relations range between 1.51 and 4.22 for PGA and between 1.65 and 2.90 for PGV. It is worth noting that, overall, the regressions of Faenza & Michelini (2010) perform slightly worse than those determined here, as shown by the distribution curve of the residuals and by the values of MSE and $\sigma_r$. Furthermore, for PGA and PGV our results do not show any significant trend of the residuals for both magnitude and distance. With regard to the intensity predictions (bottom row of Fig. 8), it is evident that an overprediction

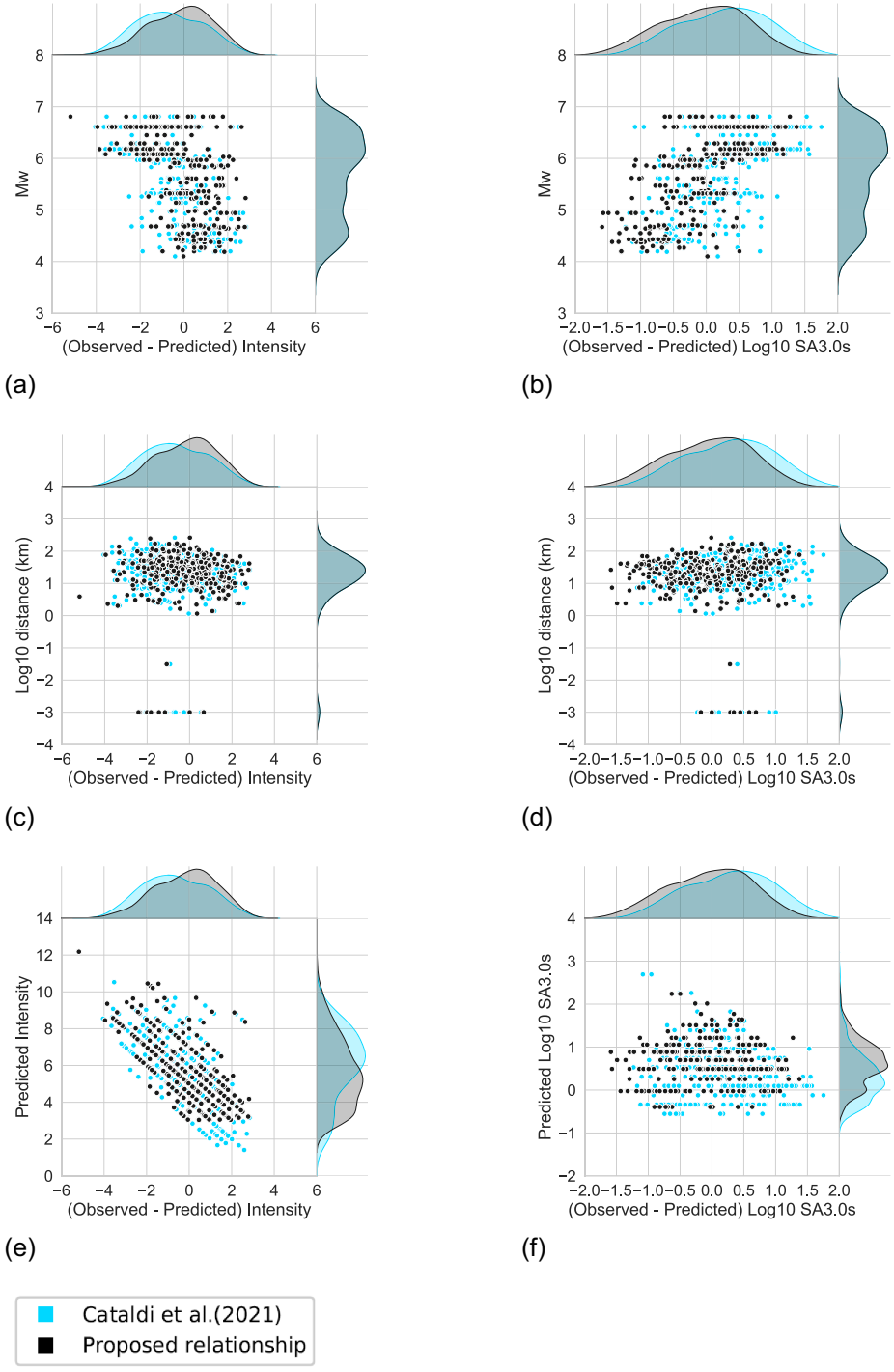


**Figure 12.** Plots of the intensity residuals for (a), (c) and (e) SA(3.0) and (b), (d) and (f) of SA(3.0) residuals represented against moment magnitude, Joyner–Boore distance and, respectively, the predicted macroseismic intensity and the corresponding predicted PGM value. The selected relationships are listed in the legend of the last panel.

**Table 4.** Statistical results for the direct relationships (i.e. from PGMs to intensity) considered in the comparison in terms of MSE and standard deviation ($\sigma_r$) of the residuals

| Relationship | PGA | | PGV | | SA(0.3) | | SA(1.0) | | SA(3.0) | |
|---|---|---|---|---|---|---|---|---|---|---|
| | MSE | $\sigma_r$ | MSE | $\sigma_r$ | MSE | $\sigma_r$ | MSE | $\sigma_r$ | MSE | $\sigma_r$ |
| Faenza and Michelini (2010) | 1.51 | 1.17 | 1.65 | 1.10 | 1.61 | 1.17 | 1.50 | 1.12 | – | – |
| Gomez Capera et al. (2018) | 3.68 | 1.89 | – | – | – | – | – | – | – | – |
| Gomez Capera et al. (2020) | 1.99 | 1.35 | 1.95 | 1.31 | 2.46 | 1.49 | 2.35 | 1.42 | – | – |
| Masi et al. (2020) | 4.22 | 1.27 | 2.90 | 1.36 | – | – | – | – | – | – |
| Cataldi et al. (2021) | 1.87 | 1.29 | 1.92 | 1.22 | 1.71 | 1.27 | 2.00 | 1.28 | 2.66 | 1.51 |
| Proposed | 1.45 | 1.19 | 1.22 | 1.11 | 1.40 | 1.18 | 1.39 | 1.18 | 2.11 | 1.44 |

**Table 5.** Statistical results for the inverse relationships (i.e. from intensity to PGMs) considered in the comparison in terms of MSE and standard deviation ($\sigma_r$) of the residuals.

| Relationship | PGA | | PGV | | SA(0.3) | | SA(1.0) | | SA(3.0) | |
|---|---|---|---|---|---|---|---|---|---|---|
| | MSE | $\sigma_r$ | MSE | $\sigma_r$ | MSE | $\sigma_r$ | MSE | $\sigma_r$ | MSE | $\sigma_r$ |
| Faenza and Michelini (2010) | 0.23 | 0.45 | 0.30 | 0.47 | 0.27 | 0.47 | 0.36 | 0.55 | – | – |
| Gomez Capera et al. (2018) | 0.26 | 0.49 | – | – | – | – | – | – | – | – |
| Gomez Capera et al. (2020) | 0.20 | 0.44 | 0.21 | 0.45 | 0.22 | 0.46 | 0.29 | 0.52 | – | – |
| Masi et al. (2020) | 0.19 | 0.44 | 0.27 | 0.52 | – | – | – | – | – | – |
| Cataldi et al. (2021) | 0.23 | 0.45 | 0.27 | 0.46 | 0.24 | 0.47 | 0.35 | 0.53 | 0.51 | 0.65 |
| Proposed | 0.20 | 0.44 | 0.20 | 0.45 | 0.21 | 0.46 | 0.27 | 0.52 | 0.40 | 0.64 |

occurs using our relationships for intensities larger than ≈ 6, and, vice versa, a smaller underprediction is observable for intensities less than ≈ 6.

Fig. 9 illustrates the results of the residual analysis when predicting PGA and PGV from the intensity values. The examined residuals do not show a strong dependence upon magnitude, Joyner–Boore distance and the predicted values of PGA and PGV. In summary, the MSE and $\sigma_r$ values calculated using our relations for PGA are close to those obtained using the models of Gomez-Capera *et al.* (2020) and Masi *et al.* (2020) (see Table 5). The estimates of PGA and PGV obtained using the other authors' relations also do not differ significantly. Although our predicted values of PGA (bottom row in Fig. 9), appear to match quite well the observed values, our relations tend to underestimate them by roughly 1 units (logarithm in base 10 of acceleration expressed in cm $s^{-2}$) at log PGA = 0. The same behaviour is found using the model proposed by Masi *et al.* (2020). For what concerns the MSE and $\sigma_r$ determined using our relations for PGV, similar values have been found for the models proposed by Gomez-Capera *et al.* (2020). This is in line with the previous qualitative comparison that highlighted that the regressions introduced by Gomez-Capera *et al.* (2020) and by this study are comparable (partly) in terms of trend.

Moving to the description of the macroseismic intensity residuals plots for SA(0.3) and SA(1.0) (Fig. 10), it is possible to observe that no correlation of the residuals with distance and magnitude occurs in our analysis. By contrast, for what concerns the intensity predictions (bottom row of Fig. 10), our model is overpredicting for intensities larger than ≈ 6 and is underpredicting for intensities less than this value. In general, we see a fairly random, uniform distribution of our residuals against the target, while, for the relationship proposed by Gomez-Capera *et al.* (2020), the residuals are not normally distributed because there are too many extreme negative residuals coinciding with very large and somewhat unrealistic values of predicted intensities ($I > 11$). These results are confirmed by the statistical values listed in Table 4: in fact, the relationships of Gomez-Capera *et al.* (2020) show both a greater error in the predicted values ($MSE_{SA(0.3)} = 2.46$, $MSE_{SA(1.0)} = 2.35$) and a larger dispersion ($\sigma_{rSA(0.3)} = 1.49$, $\sigma_{rSA(1.0)} = 1.42$) than our regressions ($MSE_{SA(0.3)} = 1.40$, $MSE_{SA(1.0)} = 1.39$ and $\sigma_{rSA(0.3)} = 1.18$, $\sigma_{rSA(1.0)} = 1.18$). By contrast, the standard deviations of the residuals obtained using our regressions and those calculated adopting the relations proposed by Faenza & Michelini (2011) appear to be comparable in value. With regard to the MSE and $\sigma_r$ calculated using the relations of Cataldi *et al.* (2021), we can note that their values lie in between those determined for our regressions and the relationships of Gomez-Capera *et al.* (2020).

The residuals plots shown in Fig. 11 indicate that the relationships proposed in this study for predicting SA(0.3) and SA(1.0) from the intensity represent the best fit model compared to the other models. Indeed, we can see a fairly random, uniform dispersion of the residuals for both SA(0.3) and SA(1.0) to demonstrate that our errors tend to be quite normally distributed around zero. These results are also confirmed by both MSE and standard deviation values listed in Table 5. In fact, our relationships in terms of SA(0.3) and SA(1.0) show both a smaller error in the predicted values ($MSE_{SA(0.3)} = 0.21$, $MSE_{SA(1.0)} = 0.27$) and a lower or equal dispersion ($\sigma_{r(SA0.3)} = 0.46$, $\sigma_{rSA(1.0)} = 0.52$) than the other models. Furthermore, for both SA(0.3) and SA(1.0), our results show some small trend for magnitude as illustrated in Figs 11(a) and (b).

For what concerns the statistical analyses of SA(3.0), we show both the graphical representation of the macroseismic intensity residuals (Figs 12a, c and e) and the residuals plots when predicting SA(3.0) from the intensity values (Figs 12b, d and f). As confirmed by both MSE and standard deviation values listed in Tables 4 and 5, the relationships determined in this work result in smaller values of the error both for intensities from SA(3.0) and for SA(3.0) from intensity values. With regard to the standard deviations, it appears that our relationships (1.44, 0.64) perform slightly better than those proposed by Cataldi *et al.* (2021) (1.51, 0.65). There does not appear to be any significant trend of the residuals for both magnitude and distance, except some small trend for magnitude when predicting SA(3.0) from the intensity values (Fig. 12b). In addition, there appears an overall overprediction when using both our new relationships and those of Cataldi *et al.* (2021) for intensities larger than ≈6 and, vice versa, underprediction for intensities less than ≈6.

## 5 RESIDUAL ANALYSIS USING SHAKEMAP

Because one of the main goals that motivated this study was the calibration of new reliable regression equations which can be adopted in the USGS-ShakeMap procedure (Wald *et al.* 1999b; Worden *et al.* 2020) for the Italian territory (Michelini *et al.* 2020), in this subsection we performed a residual analysis to compare our relationships against those of Faenza & Michelini (2010, 2011) currently implemented in the INGV ShakeMap site for Italy (http://shakemap.ingv.it).

To show the improvements resulting from the use of the relations developed in this work, we have applied eqs (5)–(9) to the data of all the 65 earthquakes listed in our data set (Oliveti *et al.* 2021, 2022). It is important to note that the vertices of all our quadratic expressions

are located near intensity 3 because the regression models were obtained using a data set with macroseismic intensities in the range [3, 10]. Overall, this issue does not appear particularly relevant considering that low levels of the ground shaking ($I < 3$) are only slightly felt by the population, they are often unreported and, as a consequence, they appear very marginally in a database like the DBMI15 (Locati *et al.* 2021) which has been compiled based on tangible and persistent effects of the ground shaking. In other words, there is no data for $I < 3$ in the DBMI15 within a 3-km radius from the stations for our selected earthquakes (and consequently in our data set as explained in Oliveti *et al.* (2022)) which makes highly discretionary the relation to be adopted. In previous studies on GMICEs that used very similar data sets with possibly only very few MDPs at $I < 3$, the regression relations valid at higher values have been just extended towards these low values (e.g. Gomez-Capera *et al.* 2020; Cataldi *et al.* 2021). Since this indeterminacy can represent a problem when generating shakemaps at very low intensities, a proper strategy is to insert reasonable constraints of the macroseismic field at low intensity values (e.g. Worden *et al.* 2012). In our case, we have interpolated linearly between the values at intensity 1 predicted by the regressions of Faenza & Michelini (2010, 2011) and the values at the vertices from our new relations (see Fig. S2, Supporting Information).

For each earthquake, we computed the shakemaps using the strong ground motion parameters and the macroseismic intensity data separately as input. To quantify the accuracy of each set of GMICEs, we took the difference between the shakemaps of the corresponding parameter (e.g. PGA, PGV, intensities, etc.) obtained from the two types of input data. Thus, this procedure was applied to the GMICEs developed by Faenza & Michelini (2010, 2011) and those in this work. For each earthquake, we calculated four different shakemaps (i.e. using the two sets of relations, from instrumental data to predict the intensities and, vice versa, from macroseismic intensities to predict the PGM maps). To generate the shakemaps we used the same values for the source parameters (hypocenter, magnitude and the fault when available), and the same configuration for ground motion models and site effects. Therefore, we evaluated the goodness of GMICEs in converting data from intensity to strong motion and vice versa, within the ShakeMap algorithm. This test is crucial in order to cross-verify the consistency of the resolved GMICEs. If the shakemaps calculated starting from PGM data and those obtained from macroseismic intensity data are very similar, the GMICEs adopted is likely very appropriate. By contrast, significant differences between the corresponding shakemaps are an indication of biased or incorrect regression relations.

In our analysis, we evaluated the difference (i.e. residuals) for all the shakemap gridpoints. The values of ground motion calculated at these gridpoints are determined from the application of the conditional multivariate normal distribution (MVN, Worden *et al.* 2018) to the input data. Each shakemap features more than 200 000 geographical gridpoints. Fig. 13 shows the boxplots of the differences for each ground motion parameter derived from all the shakemap grids calculated for the 65 earthquakes. Based on the corresponding intensity maps obtained from macroseismic surveys, we classified each ground motion value assigned to a gridpoint into one of six intensity categories (i.e. 3–4 ($3 \leq I \leq 3.9$), 4–5 ($4 \leq I \leq 4.9$), 5–6 ($5 \leq I \leq 5.9$), 6–7 ($6 \leq I \leq 6.9$), 7–8 ($7 \leq I \leq 7.9$) and >8 ($8 \leq I \leq 11$)). We excluded from the analysis the intensities less than 3 because, for these values, we used the same conversion equations (by Faenza & Michelini 2010, 2011) for generating all the shakemaps.

The figure panels demonstrate that the median value of the residuals obtained with our new GMICEs are nearly zero, especially for intensity and PGV (Figs 13a and c), for all the six classes. This result indicates that the proposed GMICEs do not produce systematic biases. If we compare the results presented above with those obtained with the relations proposed by Faenza & Michelini (2010, 2011), we observe that the latter conversion equations featured median values very close to zero for PGA when taking into account all the epicentral distances (Fig. 13b) as the intensity decreases. In contrast, larger discrepancies are observed for intensity, PGV and SA(1.0) (Figs 13a, c and e) at the lower intensity classes and for SA(0.3) and SA(3.0) (Figs 13d and f) at higher intensity values. Thus, the only ground motion variable for which the GMICEs of Faenza & Michelini (2010, 2011) perform slightly better than our new relationships is PGA at low intensities (e.g. < 7). Since correct estimations of the ground shaking in the areas closest to the epicentre are crucial for disaster risk management, we also investigated the behaviour of the PGA residuals with source distance by selecting only those gridpoints within 50 km (Fig. 13g). We observed that the residuals of the proposed regression equations are close to zero and perform better than those of Faenza & Michelini (2010, 2011) in all cases except for 3–4 and 4–5 categories. This leads us to state that shakemaps for small-medium earthquakes (i.e. those producing shaking in the intensity range between 3 and 5 at short distances) computed with only macroseismic data result in slightly overestimated PGA values as compared to those derived from instrumental data. We have obtained, however, a very accurate estimate of shaking for strong earthquakes even for PGA at epicentral distances less than 50 km. Thus, the results in Fig. 13 demonstrate that the GMICEs developed in this study can greatly improve the estimation of the ground shaking especially for the historical earthquakes of the pre-instrumental era.

## 6 CONCLUSIONS

In this study, we developed a set of fully reversible (bi-unique) relations that correlate the maximum horizontal component of recorded PGM data of PGA, PGV and SA at $T = 0.3$, 1.0 and 3.0 s to macroseismic intensity values for Italy. One of the main objectives was the determination of regression models which can be adopted in the USGS-ShakeMap procedure implemented in Italy (Michelini *et al.* 2020). These new relationships can also be important for estimating the ground motion parameters of historical events and for producing seismic hazard maps in Italy, both in terms of peak values of ground motion and in terms of spectral components.

In our analysis we have used the recently published data set (Oliveti *et al.* 2022, 2021) compiled by joining two thoroughly verified data sources, the Italian Macroseismic Database DBMI15 (Locati *et al.* 2021) and the ESM accelerometric flatfile (Lanzano *et al.* 2018). For our

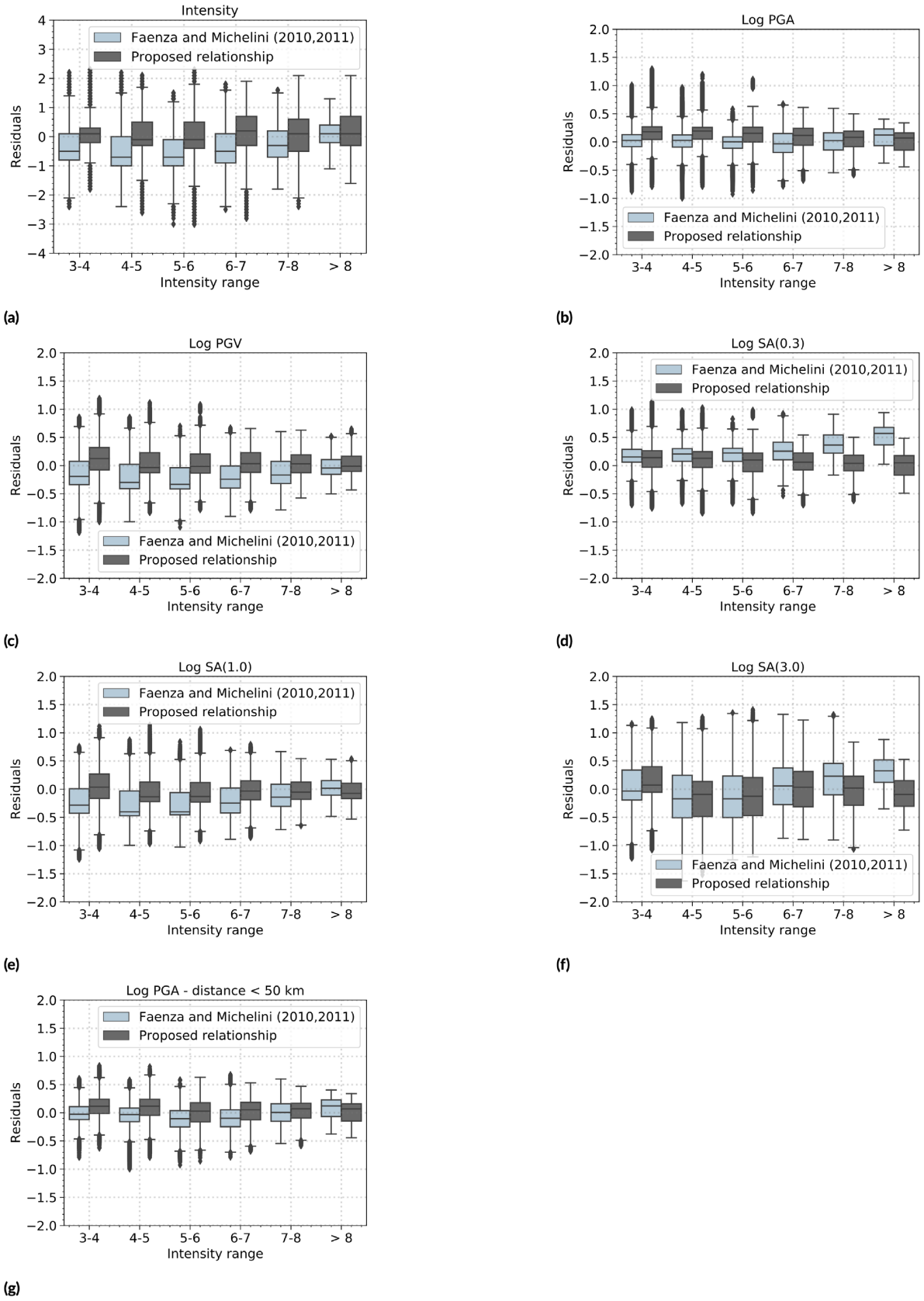


**Figure 13.** Boxplots of the gridpoint residuals for the shakemaps of the 65 earthquakes, using the regressions of Faenza & Michelini (2010, 2011) and those developed in this study (see the text for detail). The residuals are calculated as: $ShakeMap\_from\_Macroseismic\_Data - ShakeMap\_from\_Instrumental\_Data$ for (a) intensity and each ground motion parameter [PGA (b), PGV (c), SA(0.3) (d), SA(1.0) (e) and SA(3.0) (f)] at each gridpoint. Panel (g) displays PGA gridpoint differences at distances less than 50 km. The residuals in PGMs are expressed in logarithmic scale.

purposes and in line with previous studies, we have selected only PGM-macroseismic intensity pairs associated to the MDP closest locality to the recording station.

Because both the intensity and the PGM data are affected by inherent uncertainties, we adopted the ODR approach (e.g. Faenza & Michelini 2010, 2011), in which observational errors on both dependent and independent variables are taken into account and the reversibility of the proposed formulations is ensured. In order to apply the technique, we have grouped the data in classes at 0.5 intensity bins calculating the mean of logarithm of PGMs since the data set values are in agreement with a lognormal distribution.

We have used the shuffle-split cross-validation, a statistical resampling procedure where data are divided in a training set, implemented to build up the model, while a test set is utilized to validate the performance of the model. Different functional forms were assessed and the best fit was found by adopting a quadratic functional form. The AIC was used to verify the goodness of the fit.

The analysis of the residuals between observed and predicted values confirms an overall consistency of our regression equations, compared to other relationships, both for intensity versus PGMs and, PGMs versus intensity. These results are also validated by both MSE and standard deviation values of the residuals, and suggest that the residuals tend to be normally distributed. The residuals plots have also shown that the estimated regressions do not depend significantly on either magnitude or distance. We have illustrated that the regressions derived in this study provide more accurate predictions of the largest PGM values and of the macroseismic intensity.

Finally, in order to further assess the quality of the relations found here and to investigate their applicability for the generation of shakemaps in Italy, we calculated the differences between the shakemaps determined using recorded data and those derived from the macroseismic surveys. The same analysis applied to the shakemaps obtained from the GMICEs developed by Faenza & Michelini (2010, 2011) show that the new relations are more accurate evidencing minimal bias. It follows that the regression models developed here can be inserted in the configuration of the USGS-ShakeMap currently in use at INGV (Michelini *et al.* 2020).

## ACKNOWLEDGMENTS

This research was supported by the “Istituto Nazionale di Geofisica e Vulcanologia” and the “Dipartimento Protezione Civile” under 2019-2021 B2-WP1 Task 5 ShakeMap adjournment project. We would like to thank the referees for reviewing the manuscript and providing constructive criticism.

## DATA AVAILABILITY

The data set used for the regression analysis in this study is available in Zenodo at https://doi.org/10.13127/inge.2 (Oliveti *et al.* 2021). This data set was derived from two primary databases: the ESM accelerometric data bank available at https://esm.mi.ingv.it/flatfile-2018/ and the DBMI15 intensity database available at https://emidius.mi.ingv.it/CPTI15-DBMI15.

## REFERENCES

Albini, P., Musson, R.M., Rovida, A., Locati, M., Gomez Capera, A.A. & Viganò, D., 2014. The global earthquake history, *Earthq. Spectra,* **30**(2), 607–624.

Allen, T.I., Wald, D.J., Hotovec, A.J., Lin, K., Earle, P.S. & Marano, K.D., 2008. *An Atlas of ShakeMaps for Selected Global Earthquakes,* US Department of the Interior, US Geological Survey.

Allen, T.I., Wald, D.J. & Worden, C.B., 2012. Intensity attenuation for active crustal regions, *J. Seismol.,* **16**(3), 409–433.

Anderson, D., Burnham, K. & White, G., 1998. Comparison of Akaike Information Criterion and consistent Akaike Information Criterion for model selection and statistical inference from capture-recapture studies, *J. appl. Stat.,* **25**(2), 263–282.

Azzaro, R., D’Amico, S. & Tuvè, T., 2011. Estimating the magnitude of historical earthquakes from macroseismic intensity data: new relationships for the volcanic region of mount etna (italy), *Seismol. Res. Lett.,* **82**(4), 533–544.

Bakun, W., 2006. Estimating locations and magnitudes of earthquakes in southern california from modified mercalli intensities, *Bull. seism. Soc. Am.,* **96**(4A), 1278–1295.

Bakun, W., Gómez Capera, A. & Stucchi, M., 2011. Epistemic uncertainty in the location and magnitude of earthquakes in italy from macroseismic data, *Bull. seism. Soc. Am.,* **101**(6), 2712–2725.

Bakun, W.H. & Scotti, O., 2006. Regional intensity attenuation models for france and the estimation of magnitude and location of historical earthquakes, *Geophys. J. Int.,* **164**(3), 596–610.

Baumont, D., Manchuel, K., Traversa, P., Durouchoux, C., Nayman, E. & Ameri, G., 2018. Empirical intensity attenuation models calibrated in mw for metropolitan france, *Bull Earthq Eng*.

Beauval, C., Yepes, H., Bakun, W.H., Egred, J., Alvarado, A. & Singaucho, J.-C., 2010. Locations and magnitudes of historical earthquakes in the sierra of ecuador (1587–1996), *Geophys. J. Int.,* **181**(3), 1613–1633.

Bilal, M. & Askan, A., 2014. Relationships between felt intensity and recorded ground-motion parameters for Turkey, *Bull. seism. Soc. Am.,* **104**(1), 484–496.

Bindi, D., Parolai, S., Oth, A., Abdrakhmatov, K., Muraliev, A. & Zschau, J., 2011. Intensity prediction equations for central asia, *Geophys. J. Int.,* **187**(1), 327–337.

Boggs, P.T., Byrd, R.H., Donaldson, J.R. & Schnabel, R.B., 1987. *ODRPACK Software for Weighted Orthogonal Distance Regression,* Tech. rep., The Department of Computer Science at the University of Colorado Boulder.

Boggs, P.T., Spiegelman, C.H., Donaldson, J.R. & Schnabel, R.B., 1988. A computational examination of orthogonal distance regression, *J. Econom.,* **38**(1-2), 169–201.

Bossu, R., Landès, M., Roussel, F., Steed, R., Mazet-Roux, G., Martin, S.S. & Hough, S., 2017. Thumbnail-based questionnaires for the rapid and efficient collection of macroseismic data from global earthquakes, *Seismol. Res. Lett.,* **88**(1), 72–81.

Caprio, M., Tarigan, B., Worden, C.B., Wiemer, S. & Wald, D.J., 2015. Ground motion to intensity conversion equations (GMICEs): a global relationship and evaluation of regional dependency, *Bull. seism. Soc. Am.,* **105**(3), 1476–1490.

Cataldi, L., Tiberi, L. & Costa, G., 2021. Estimation of mcs intensity for italy from high quality accelerometric data, using gmices and gaussian naïve bayes classifiers, *Bull. Earthq. Eng.,* **19**(6), 2325–2342.

Chollet, F., 2018. *Deep Learning mit Python und Keras: Das Praxis-Handbuch vom Entwickler der Keras-Bibliothek,* MITP-Verlags GmbH & Co. KG.

Codermatz, R., Nicolich, R. & Slejko, D., 2003. Seismic risk assessments and gis technology: applications to infrastructures in the friuli–venezia giulia region (ne italy), *Earthq. Eng. Struct. Dynam.,* **32**(11), 1677–1690.

Corbane, C., Hancilar, U., Ehrlich, D. & De Groeve, T., 2017. Pan-European seismic risk assessment: a proof of concept using the Earthquake Loss Estimation Routine (ELER), *Bull. Earthq. Eng.,* **15**(3), 1057–1083.

Cramer, C.H., 2020. Updated gmice for central and eastern North America extending to higher intensities, *Seismol. Soc. Am.,* **91**(6), 3518–3527.

Cua, G., Wald, D., Allen, T., Garcia, D., Worden, C., Gerstenberger, M., Lin, K. & Marano, K., 2010. *Best Practices" for using Macroseismic Intensity and Ground Motion-intensity Conversion Equations for Hazard and Loss Models in Gem1,* Tech. rep., GEM Technical Report 2010-4, GEM Foundation, Pavia, Italy .

Decanini, L., Gavarini, C. & Mollaioli, F., 1995. Proposta di definizione delle relazioni tra intensita macrosismica e parametri del moto del suolo, in *Atti del 7° Convegno Nazionale l'Ingegneria Sismica in Italia,* **1,** 63–72.

Du, K., Ding, B., Bai, W., Sun, J. & Bai, J., 2020. Quantifying uncertainties in ground motion-macroseismic intensity conversion equations. A probabilistic relationship for western china, *J. Earthq. Eng.,* 1–25.

Earle, P.S., Wald, D.J., Jaiswal, K.S., Allen, T.I., Hearne, M.G., Marano, K.D., Hotovec, A.J. & Fee, J.M., 2009. Prompt Assessment of Global Earthquakes for Response (PAGER): a system for rapidly determining the impact of earthquakes worldwide, *US Geol. Surv. Open-File Rep.,* **1131,** 1–15.

Faccioli, E. & Cauzzi, C., 2006. Macroseismic intensities for seismic scenarios estimated from instrumentally based correlations, in *Proc. First European Conference on Earthquake Engineering and Seismology,* Swiss Society for Earthquake Engineering and Structural Dynamics (SGEB), paper no. 569, 3064–3073.

Faenza, L. & Michelini, A., 2010. Regression analysis of MCS intensity and ground motion parameters in Italy and its application in ShakeMap, *Geophys. J. Int.,* **180**(3), 1138–1152.

Faenza, L. & Michelini, A., 2011. Regression analysis of MCS intensity and ground motion spectral accelerations (SAs) in Italy, *Geophys. J. Int.,* **186**(3), 1415–1430.

Faenza, L., Pierdominici, S., Camassi, R., Michelini, A., Ercolani, E. & Lauciani, V., 2013. The shakemap atlas for the city of naples, Italy, *Seismol. Res. Lett.,* **84**(6), 963–972.

Fu, G., Deming, D., Knutson, H., Madhusudhan, N., Mandell, A. & Fraine, J., 2017. Statistical analysis of Hubble/WFC3 transit spectroscopy of extrasolar planets, *Astrophys. J.,* **847**(2), L22 (6 pp).

Gasperini, P., Vannucci, G., Tripone, D. & Boschi, E., 2010. The location and sizing of historical earthquakes using the attenuation of macroseismic intensity with distance, *Bull. seism. Soc. Am.,* **100**(5A), 2035–2066.

Gatzsche, K., Babel, W., Falge, E., Pyles, R.D., Paw U, K.T., Raabe, A. & Foken, T., 2018. Footprint-weighted tile approach for a spruce forest and a nearby patchy clearing using the ACASA model, *Biogeosciences,* **15**(9), 2945–2960.

Golub, G.H. & Van Loan, C.F., 1980. An analysis of the total least squares problem, *SIAM J. Numer. Anal.,* **17**(6), 883–893.

Gomez-Capera, A., Santulin, M., D'Amico, M., D'Amico, V., Locati, M., Luzi, L., Massa, M. & Puglia, R., 2018. Macroseismic intensity to ground motion empirical relationships for Italy, *Proceedings, 37° Convegno Nazionale GNGTS, Bologna (Italy),* ISBN, 289–291.

Gomez-Capera, A.A., D'Amico, M., Lanzano, G., Locati, M. & Santulin, M., 2020. Relationships between ground motion parameters and macroseismic intensity for Italy, *Bull. Earthq. Eng.,* **18,** 5143–5164.

Grünthal, G., 1998. *European Macroseismic Scale 1998,* Tech. Rep., European Seismological Commission (ESC).

Kaka, S.I. & Atkinson, G.M., 2004. Relationships between instrumental ground-motion parameters and modified mercalli intensity in eastern north america, *Bull. seism. Soc. Am.,* **94**(5), 1728–1736.

Karim, K.R. & Yamazaki, F., 2002. Correlation of jma instrumental seismic intensity with strong motion parameters, *Earthq. Eng. Struct. Dynam.,* **31**(5), 1191–1212.

Kästli, P. & Fäh, D., 2006. Rapid estimation of macroseismic effects and shake maps combining macroseismic and instrumental data, in *Proceedings of the First European Conference on Earthquake Engineering and Seismology (ECEES),* Geneva, Switzerland, Swiss Society for Earthquake Engineering and Structural Dynamics (SGEB), 7341–7350.

Kuehn, N. & Scherbaum, F., 2010. A Naive Bayes classifier for intensities using peak ground velocity and acceleration, *Bull. seism. Soc. Am.,* **100**(6), 3278–3283.

Lanzano, G. *et al.*, 2018. ESM strong-motion flat-file 2018, *Istituto Nazionale di Geofisica e Vulcanologia (INGV), Helmholtz-Zentrum Potsdam Deutsches GeoForschungsZentrum (GFZ), Observatories and Research Facilities for European Seismology (ORFEUS).*

Lee, K.A., Lee, J.R. & Bell, P., 2020. A review of citizen science within the earth sciences: potential benefits and obstacles, in *Proceedings of the Geologists' Association,* Elsevier, **131,** 605–617..

Lesueur, C., Cara, M., Scotti, O., Schlupp, A. & Sira, C., 2013. Linking ground motion measurements and macroseismic observations in France: a case study based on accelerometric and macroseismic databases, *J. Seismol.,* **17**(2), 313–333.

Locati, M., Rovida, A., Albini, P. & Stucchi, M., 2014. The ahead portal: a gateway to european historical earthquake data, *Seismol. Res. Lett.,* **85**(3), 727–734.

Locati, M. *et al.*, 2021. Database Macrosismico Italiano DBMI15, versione 3.0, *Istituto Nazionale di Geofisica e Vulcanologia*.

Luzi, L., Hailemikael, S., Bindi, D., Pacor, F., Mele, F. & Sabetta, F., 2008. ITACA (ITalian ACcelerometric Archive): a web portal for the dissemination of Italian strong-motion data, *Seismol. Res. Lett.,* **79**(5), 716–722.

Margottini, C., Molin, D. & Serva, L., 1992. Intensity versus ground motion: a new approach using italian data, *Eng. Geol.,* **33**(1), 45–58.

Masi, A., Chiauzzi, L., Nicodemo, G. & Manfredi, V., 2020. Correlations between macroseismic intensity estimations and ground motion measures of seismic events, *Bull. Earthq. Eng.,* **18**(5), 1899–1932.

Michelini, A., Faenza, L., Lauciani, V. & Malagnini, L., 2008. ShakeMap implementation in Italy, *Seismol. Res. Lett.,* **79**(5), 688–697.

Michelini, A., Faenza, L., Lanzano, G., Lauciani, V., Jozinović, D., Puglia, R. & Luzi, L., 2020. The new ShakeMap in Italy: progress and advances in the last 10 yr, *Seismol. Res. Lett.,* **91**(1), 317–333.

Molin, D. *et al.*, 1995. *Considerations on the Assessment of Macroseismic Intensity, Annals of Geophysics,* **38**(5-6),805–810.

Moratalla, J.M., Goded, T., Rhoades, D.A., Canessa, S. & Gerstenberger, M.C., 2021. New ground motion to intensity conversion equations (gmices) for new zealand, *Seismol. Soc. Am.,* **92**(1), 448–459.

Musson, R.M., Grünthal, G. & Stucchi, M., 2010. The comparison of macroseismic intensity scales, *J. Seismol.,* **14**(2), 413–428.

Oliveti, I., Faenza Licia, A. & Michelini, A., 2021. *INGe: Intensity-ground Motion Dataset for Italy [dataset] version 2,* Zenodo.

Oliveti, I., Faenza, L. & Michelini, A., 2022. Inge: intensity-ground motion dataset for Italy, *Ann. Geophys.,* **65**(1), DM102.

Oros, E., Placinta, A.O., Popa, M., Rogozea, M. & Paulescu, D., 2019. Attenuation of macroseismic intensity for crustal romanian earthquakes: Calibrating the bakun-wentworth's method, in *IOP Conference Series: Earth and Environmental Science,* Vol. **362,** pp. 012026, IOP Publishing.

Panza, G.F., Cazzaro, R., Vaccari, F., *et al.*, 1997. Correlation between macroseismic intensities and seismic ground motion parameters, *Ann. Geophys.,* **40**(5), 1371–1382.

Pasolini, C., Gasperini, P., Albarello, D., Lolli, B. & D'Amico, V., 2008. The attenuation of seismic intensity in Italy, part i: theoretical and empirical backgrounds, *Bull. seism. Soc. Am.,* **98**(2), 682–691.

Pittore, M., Graziani, L., Maramai, A., Haas, M., Parolai, S. & Tertulliani, A., 2018. Bayesian estimation of macroseismic intensity from post-earthquake rapid damage mapping, *Earthq. Spectra,* **34**(4), 1809–1828.

Provost, L. & Scotti, O., 2020. Quake-md: open-source code to quantify uncertainties in magnitude–depth estimates of earthquakes from macroseismic intensities, *Seismol. Res. Lett.,* **91**(5), 2520–2530.

Pylak, M., Fornalski, K.W., Reszczyńska, J., Kukulski, P., Waligórski, M. P.R. & Dobrzyński, L., 2021. Analysis of indoor radon data using Bayesian, Random Binning, and Maximum Entropy Methods, *Dose-Response: Publ. Int. Hormesis Soc.,* **19**(2).

Quitoriano, V. & Wald, D.J., 2020. Usgs "Did You Feel It?" science and lessons from 20 years of citizen science-based macroseismology, *The Power of Citizen Seismology: Science and Social Impacts*.

Rovida, A. & Locati, M., 2015. Archive of historical earthquake data for the european-mediterranean area, in *Perspectives on European Earthquake Engineering and Seismology,* 359–369, Springer, Cham.

Sbarra, P., Burrato, P., Tosi, P., Vannoli, P., De Rubeis, V. & Valensise, G., 2019. Inferring the depth of pre-instrumental earthquakes from macroseismic intensity data: a case-history from northern italy, *Sci. Rep.,* **9**(1), 1–13.

Schlupp, A., 2016. Shakemap fed by macroseismic data in france: feedbacks and contribution for improving SHA., in *Fall Meeting Abstracts,* American Geophysical Union, S11B–2445.

Sieberg, A., 1930. Scala MCS (Mercalli-Cancani-Sieberg), *Geologie der Erdbeben, Handbuch der Geophysik,* **2**(4), 552–555.

Sokolov, V., Furumura, T. & Wenzel, F., 2010. On the use of JMA intensity in earthquake early warning systems, *Bull. Earthq. Eng.,* **8**(4), 767–786.

Stromeyer, D. & Grünthal, G., 2009. Attenuation relationship of macroseismic intensities in central europe, *Bull. seism. Soc. Am.,* **99**(2A), 554–565.

Stucchi, M. *et al.*, 2007. DBMI04, il database delle osservazioni macrosismiche dei terremoti italiani utilizzate per la compilazione del catalogo parametrico cpti04, *Quaderni di Geofisica,* **49,** 1–38.

Tang, Y., Lam, N., Tsang, H.-H. & Lumantarna, E., 2019. Use of macroseismic intensity data to validate a regionally adjustable ground motion prediction model, *Geosciences,* **9**(10), 422.

Teramo, A., Termini, D., Stillitani, E. & Bottari, A., 1996. The determination of the epicentre by a vectorial modelling of macroseismic intensity distribution, *Nat. Hazards,* **13**(2), 101–117.

Tosi, P., De Rubeis, V., Sbarra, P. & Sorrentino, D., 2007. *Hai Sentito Il Terremoto (HSIT). Istituto Nazionale di Geofisica e Vulcanologia (INGV),* doi:10.13127/hsit.

Traversa, P., Baumont, D., Manchuel, K., Nayman, E. & Durouchoux, C., 2018. Exploration tree approach to estimate historical earthquakes mw and depth, test cases from the french past seismicity, *Bull. Earthq. Eng.,* **16**(6), 2169–2193.

Trendafiloski, G., Wyss, M. & Rosset, P., 2011. Loss estimation module in the second generation software QLARM, in *Human Casualties in Earthquakes,* pp. 95–106, Springer.

Tselentis, G.-A. & Danciu, L., 2008. Empirical relationships between modified mercalli intensity and engineering ground-motion parameters in greece, *Bull. seism. Soc. Am.,* **98**(4), 1863–1875.

Vannucci, G., Lolli, B. & Gasperini, P., 2021. Inhomogeneity of macroseismic intensities in Italy and consequences for macroseismic magnitude estimation, *Seismol. Res. Lett,* **92,** 2234–2244.

Villani, M. *et al.*, 2019. A selection of gmpes for the united kingdom based on instrumental and macroseismic datasets, *Bull. seism. Soc. Am.,* **109**(4), 1378–1400.

Wald, D.J., Quitoriano, V., Heaton, T.H. & Kanamori, H., 1999a. Relationships between peak ground acceleration, peak ground velocity, and modified Mercalli intensity in California, *Earthq. Spectra,* **15**(3), 557–564.

Wald, D.J., Quitoriano, V., Heaton, T.H., Kanamori, H., Scrivner, C.W. & Worden, C.B., 1999b. TriNet "ShakeMaps" rapid generation of peak ground motion and intensity maps for earthquakes in southern California, *Earthq. Spectra,* **15**(3), 537–555.

Wald, D.J., Worden, C., Quitoriano, V. & Pankow, K., 2006. *ShakeMap Manual: Technical Mmanual, User's Guide, and Software Guide,* Tech. rep., U. S. Geological Survey.

Wald, D.J., Quitoriano, V., Worden, C.B., Hopper, M. & Dewey, J.W., 2012. USGS "Did You Feel It?" internet-based macroseismic intensity maps, *Ann. Geophys.,* **54**(6), 1–13.

Worden, C., Gerstenberger, M., Rhoades, D. & Wald, D., 2012. Probabilistic relationships between ground-motion parameters and modified Mercalli intensity in California, *Bull. seism. Soc. Am.,* **102**(1), 204–221.

Worden, C.B., Thompson, E.M., Baker, J.W., Bradley, B.A., Luco, N. & Wald, D.J., 2018. Spatial and spectral interpolation of ground-motion intensity measure observations, *Bull. seism. Soc. Am.,* **108**(2), 866–875.

Worden, C.B., Thompson, E.M., Hearne, M. & Wald, D.J., 2020. ShakeMap Manual Online: Technical Manual, User's Guide, and Software Guide, *U. S. Geological Survey*.

Wu, Y.-M., Teng, T.-l., Shin, T.-C. & Hsiao, N.-C., 2003. Relationship between peak ground acceleration, peak ground velocity, and intensity in taiwan, *Bull. seism. Soc. Am.,* **93**(1), 386–396.

Zanini, M.A., Hofer, L. & Faleschini, F., 2019. Reversible ground motion-to-intensity conversion equations based on the EMS-98 scale, *Eng. Struct.,* **180,** 310–320.

## SUPPORTING INFORMATION

Supplementary data are available at *GJI* online.

Figure S1: Intensity to ground motion conversion equations (IGMCEs) proposed in this study for (a) PGA, (b) PGV, SA at periods (c) 0.3 s, (d) 1.0 s and (e) 3.0 s. Quadratic function (solid grey line) and the associated 95 per cent confidence (dotted lines) and prediction (dashed lines) intervals are shown.

Figure S2: Ground motion to intensity conversion equations (GMICEs) applied in Section 4.3 for (a) PGA, (b) PGV, SA at periods (c) 0.3 s, (d) 1.0 s and (e) 3.0 s. Quadratic function (solid line) within the range of the data set used in this study and linear function (dashed line) in the range of intensity 1 to ∼3 are shown.

Table S1: Regression coefficients ($a$, $b$, $c$, $d$) and their standard deviation ($\sigma_a$, $\sigma_b$, $\sigma_c$, $\sigma_d$) for the cubic equations.

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