



# Ground motion models for the new seismic hazard model of Italy (MPS19): selection for active shallow crustal regions and subduction zones

Giovanni Lanzano<sup>1</sup> · Lucia Luzi<sup>1</sup> · Vera D'Amico<sup>2</sup> · Francesca Pacor<sup>1</sup> · Carlo Meletti<sup>2</sup> · Warner Marzocchi<sup>3</sup> · Renata Rotondi<sup>4</sup> · Elisa Varini<sup>4</sup>

Received: 25 July 2019 / Accepted: 20 April 2020 / Published online: 27 April 2020  
© Springer Nature B.V. 2020

## Abstract

The aim of this paper is to identify the ground motion models (GMMs), applicable in active shallow crustal regions (ASCRs) and subduction zones (SZs), to be used for the new release of the seismic hazard model of Italy (MPS19) for peak ground acceleration and 12 ordinates of the acceleration response spectra (5% damping) in the period range 0.05–4 s. The steps to achieve such goal are: (1) a pre-selection of the GMMs that takes into account the suitability for the application to the Italian territory and the fulfillment of the new hazard model requirements; (2) the assessment of a proper scoring of the pre-selected GMMs using strong-motion data recorded in Italy; (3) the selection of the GMMs to be used in the hazard calculation. The final set of GMMs describes satisfactorily the epistemic uncertainty of the ground shaking process, privileging the simplicity and flexibility of the functional form. Finally, the weights of the selected GMMs are assigned combining the results of the scoring and the weights obtained through an experts' elicitation.

**Keywords** Ground motion models · Model scoring · Weights · Italy · Seismic hazard model

---

**Electronic supplementary material** The online version of this article (<https://doi.org/10.1007/s10518-020-00850-y>) contains supplementary material, which is available to authorized users.

---

✉ Giovanni Lanzano  
giovanni.lanzano@ingv.it

<sup>1</sup> Istituto Nazionale Di Geofisica e Vulcanologia, Milan, Italy

<sup>2</sup> Istituto Nazionale Di Geofisica e Vulcanologia, Pisa, Italy

<sup>3</sup> University of Naples "Federico II", Naples, Italy

<sup>4</sup> Consiglio Nazionale Delle Ricerche, IMATI, Milan, Italy

## 1 Introduction

The new Italian seismic hazard model, MPS19 (in Italian *Modello di Pericolosità Sismica*, Meletti et al. 2017) is a broad research initiative, launched by the Seismic Hazard Center (*Centro Pericolosità Sismica*, CPS) of the Istituto Nazionale di Geofisica e Vulcanologia (INGV) in 2015 and supported by the Italian Civil Protection Department (*Dipartimento della Protezione Civile*, DPC), that aims to produce a new seismic hazard model for Italy, to update the current one released in the period 2004–2006, named MPS04 (Stucchi et al. 2011). The latter is adopted as input for the seismic actions by the Italian Building Code (NTC08 2008; NTC18 2018) providing the uniform hazard spectra for each location of the national territory across a regular 5 km equally-spaced grid.

The main objective of the update of MPS04 is to exploit, as much as possible, all the recent progress in seismology, seismotectonics, seismic engineering and engineering seismology, in terms of data, models, methods and numerical tools. The requirements of the MPS19 model are: (1) uniform cover of the national territory and the surrounding areas; (2) rock site conditions<sup>1</sup> in flat topographic setting; (3) hazard estimates in terms of horizontal component of Peak Ground Acceleration, PGA, and 12 ordinates of the acceleration response spectra (5% damping), at periods  $T$  0.05, 0.10, 0.15, 0.20, 0.30, 0.40, 0.50, 0.75, 1.0, 2.0, 3.0, 4.0 s; (4) geometrical mean of the horizontal components; (5) range of return periods between 30 and 2475 years.<sup>2</sup>

One of the tasks of MPS19 is focused on the prediction models for strong motion parameters, named Ground Motion Models (GMMs), and for macroseismic intensity to be used in the hazard calculations. In MPS04, the GMMs selected for shallow active crustal seismicity were those proposed for the maximum of the horizontal components by Sabetta and Pugliese (1996) for Italy and by Ambraseys et al. (1996) for Europe (see Montaldo et al. 2005). The selected models were also scaled according to Bommer et al. (2003) to account for different style of faulting and distance metrics. In addition, local GMMs derived from weak motions (Malagnini et al. 2000, 2002; Morasca et al. 2006) were included, in order to account for regional attenuation features in different areas of Italy (see Montaldo et al. 2005).

The GMMs selection for MPS04 was not supported by an objective measurement of the performance of the candidate models against the data, also because the available accelerometric records of strong events in Italy before 2004 are very few and the number of available GMMs was limited. Nowadays, considering the dramatic growth of the seismic networks in Italy and the occurrence of three major sequences in the past last ten years (L'Aquila,  $M_W$  6.1,<sup>3</sup> 2009, Emilia,  $M_W$  6.0, 2012 and Central Italy,  $M_W$  6.0 and 6.5, 2016–2017), the number of recorded data available to test the GMMs has increased

<sup>1</sup> The rock site conditions correspond to the definition of site category A in the NTC08, i.e. rock or rock-like formations with average shear wave velocity  $V_{s,30} > 800$  m/s, with the exception of maximum 3 m of weaker material at the surface. This description is partially modified in the latest release of the Italian Building Code (NTC18), where category A is defined as rock or rock-like formations with shear wave velocity  $V_s > 800$  m/s, with the exception of maximum 3 m of soil with  $V_s$  lower than 800 m/s.

<sup>2</sup> The range of return periods is coherent with the requirements of the Italian Building Code (NTC18) for the seismic design of the structures (reference periods from 50 to 100 years). The corresponding annual probabilities of exceedance are in the range  $3 \times 10^{-2}$ – $4 \times 10^{-5}$ .

<sup>3</sup> All the values of the local and moment magnitudes, reported in this paper, are derived from ITACA (Italian Accelerometric Archive, itaca.mi.ingv.it) database.

exponentially. The availability of this great volume of data allows us to prepare a very large and qualified dataset of accelerometric recordings for the GMMs testing.

The strategy followed in the selection of the GMMs mimics the procedure implemented by Delavaud et al. (2012) to construct the ground-motion logic tree for the Probabilistic Seismic Hazard Assessment (PSHA) in Europe, in the framework of the EU project SHARE (Woessner et al. 2015, <https://www.share-eu.org/>). In particular, the final outcome is given by the combination of both objective models testing and expert judgment: the first approach aims at minimizing the role of subjectivity in PSHA, which has been often the main target for criticism (Roselli et al. 2016) and the second one should compensate some limits derived from the use of the data only and help in the evaluation of the predictive power of the GMMs (Bindi 2017). As a matter of fact, strong motion data are relatively few especially for the seismic scenarios that mainly contribute to the hazard, i.e. strong earthquakes and sites located in near field conditions. We believe that this approach could return a hazard model that captures “the center, the body and the range” of possible ground motions in Italy.

In the following, the procedure to pre-select and rank the GMMs for the Active Shallow Crustal Regions (ASCRs) and Subduction Zones (SZs) in Italy is detailed. On the basis of the scoring results, few models are finally selected to be implemented in the new Italian hazard model. None of the available GMMs for the volcanic areas are able to satisfactorily predict the ground motion amplitudes observed recently in Italy (2018–2019 Mount Etna sequence, mainshock 26/12/2018  $M_W$  4.9, and 2017 Ischia event, 21/08/2017  $M_W$  3.9), especially in near-source conditions (Lanzano and Luzi 2020). For this reason, the selection of the models for the volcanic earthquakes and the relative weights assignment were completely carried out on the basis of the expert judgement.

## 2 Preliminary selection of the GMMs for ASCRs

The candidate GMMs are obtained from the list compiled by Douglas (<https://www.gmpe.org.uk>; the database was accessed in 2016), who reports the main features of more than 300 empirical GMMs, published in literature. As a first criterion, we select models for ASCRs that calibrate GMMs both for PGA and Spectral Acceleration (SA) ordinates, using the same dataset. To restrict the number of empirical GMMs available, we adopt the criteria by Cotton et al. (2006) and, more recently, by Bommer et al. (2010), rejecting models: (1) derived for an inappropriate tectonic regime; (2) not published in a peer-reviewed journal; (3) with insufficient documentation on the calibration dataset; (4) superseded by a more recent publication; (5) with the period range not appropriate for engineering applications; (6) with inappropriate functional form (i.e. lacking either non-linear magnitude dependence or magnitude-dependent decay with distance); (7) obtained with inappropriate regression method. We then add further criteria, that are project-specific, i.e. we reject models: (8) not applicable on the whole national territory (avoiding local models); (9) based on magnitudes different than moment magnitude  $M_W$ ; (10) with the range of applicability too small for the requirements of MPS19 (narrow range of magnitude or distance); (11) with intensity measures (IMs) not defined as the geometrical mean (GM) of the horizontal components or the RotD50 (Boore 2010); (12) with the range of response periods incompatible with the requirement of MPS19, chosen to be from 0.05 to 4 s, or for very few periods.

The candidate models after the pre-selection are given in Table 1, where the characteristics of the GMMs are reported in terms of magnitude and distance ranges, period ranges and site classes.

The proposed 12 rules are easy-to-use for the pre-selection, but the strict application leads to the exclusion of several models. Some caveats, that lead us to allow some exceptions in the pre-selection, are:

- Models versioning (criterion 4): when same (or a part of the) authors decided to publish an update of a GMM, the new one not necessarily replaces and supersedes the previous model. As a matter of fact, the increase of the dataset, a new functional form and more modern procedure to calibrate the equations could lead to results that could not improve the predictions of the previous model;
- Inadequate range of periods (criterion 12, sub-criterion #1): few models (i.e. ITA10, AB10 and BND14) are available for a range of periods which is slightly narrower than that required by MPS19 (0.05–4 s). As a matter of fact, such models were calibrated on a significant amount of analog records, with high pass filter generally higher than 0.3 Hz ( $T=3$  s);
- Inadequate number of periods (criterion 12, sub-criterion #2): some models, especially with innovative functional forms and regression procedures, were proposed in literature for few illustrative spectral ordinates. The model application to other ordinates is possible via linear interpolation of the nearest periods. However, such interpolation could be rough if the period sampling is large. In our study, the lower limit is represented by the KS15 model, calibrated on 7 (+1 outside the MPS19 range) spectral ordinates ( $T=0.01, 0.05, 0.1, 0.5, 1, 2, 3$  and 4 s).

The pre-selection includes GMMs derived at different scales: (a) global, from worldwide datasets; (b) pan-European, from datasets including European and Middle-East records; (c) regional, from national datasets, i.e. Italy. We also include models developed in other regions (e.g. Japan), that could be applied in Italy, because of the similar attenuation crustal properties (Boore et al. 2014). Moreover, we take into account that several so-called global or European models are often constrained by the data of the regions that mostly contribute to the calibration datasets, such as the RESORCE models, which are Italian-biased (especially BND14), and CZ15, which is Japanese-biased.

Some models in Table 1 were developed for different distance metrics and/or for different parameters of the site effects modelling ( $V_{S,30}$  or site categories). In order to limit the number of the investigated models, we decide to include all the GMMs for different distances in the test, but to select only the predictive equations with site effect term that exhibits the best performance against the data (see §4 for the procedure). In particular, the model ASB14 by Akkar et al. (2014b) is proposed in terms of epicentral,  $R_{epi}$ , hypocentral,  $R_{hyp}$ , and Joyner–Boore ( $R_{JB}$ ) distances; BND14 proposed predictive equations both for  $R_{JB}$  and  $R_{hyp}$ .

Figure 1 shows the trellis plots of the spectral predictions of the candidate GMMs, assuming rock site conditions (EC8 site category ‘A’ or  $V_{S,30}=800$  m/s) and normal faulting. The distances metrics ( $R_{JB}$  and  $R_{rup}$ ) are calculated following the procedure by Kaklamanos et al. (2011) and using the well-known empirical relations by Wells and Coppersmith (1994) for the fault geometry with a fixed dip of  $45^\circ$  and focal depth of 10 km. The epicentre is positioned at half-length and half-width of the surface fault projection. The recording site is assumed to be in hanging wall conditions, with  $R_X > 0$  and  $R_Y = 0$  (see Ancheta et al. 2014 for distance notation in NGA-West2). The basin depth is not available for all the recording stations in Italy

**Table 1** List of the candidate GMMs for the application to ASCRs for PGA<sup>a</sup> and acceleration response spectra ordinates

Name	Code	Area	H comp	M range	R type	R range (km)	Period range (s)	Site-effect term	Style of faulting
Zhao (2006)	ZHA06	Japan	GM	5.0–8.3	R <sub>rup</sub>	0–300	0.05–5.0	5 Classes	N, R, SS
Bindi et al (2011) <sup>b</sup>	ITA10	Italy	GM	4.1–6.9	R <sub>JB</sub>	0–200	0.04–2.0	5 EC8 <sup>c</sup> classes	N, R, SS, U
Akkar and Bommer (2010) <sup>d</sup>	AB10	Europe Middle-East	GM	5.0–7.6	R <sub>JB</sub>	0–100	0.05–3.0	3 Classes	N, R, SS
Abrahamson et al (2014)	ASK14	Worldwide	RotD50	3.0–7.9	R <sub>rup</sub>	0–300	0.01–10	V <sub>S,30</sub>	N, R, SS
Boore et al (2014)	BSSA14	Worldwide	RotD50	3.0–7.9	R <sub>JB</sub>	0–400	0.01–10	V <sub>S,30</sub>	N, R, SS, U
Campbell and Bozorgnia (2014)	CB14	Worldwide	RotD50	3.0–7.9	R <sub>rup</sub>	0–300	0.01–10	V <sub>S,30</sub>	N, R, SS
Chiou and Youngs (2014)	CY14	Worldwide	RotD50	3.1–7.9	R <sub>rup</sub>	0–400	0.01–10	V <sub>S,30</sub>	N, R, SS
Idriss (2014)	ID14	Worldwide	RotD50	4.5–7.9	R <sub>rup</sub>	0–175	0.01–10	V <sub>S,30</sub>	N, R, SS
Akkar et al (2014b)	ASB14	Europe and Middle-East	GM	4.7–7.6	R <sub>JB</sub>	0–200	0.01–4.0	V <sub>S,30</sub>	SS, R, N
Bindi et al (2014) <sup>a</sup>	BND14	Europe and Middle-East	GM	4.0–7.6	R <sub>JB</sub> , R <sub>hyp</sub>	0–300	0.02–3.0	4 Classes, V <sub>S,30</sub>	SS, R, N, U
Derras et al (2014)	DBC14	Europe and Middle-East	GM	3.6–7.6	R <sub>JB</sub>	0–550	0.01–4.0	V <sub>S,30</sub>	N, R, SS
Cauzzi et al (2015) <sup>d</sup>	CZ15	Worldwide	GM	4.5–7.9	R <sub>rup</sub>	0–150	0.02–10	V <sub>S,30</sub>	SS, N, R
Kuehn and Scherbaum (2015)	KS15	Europe and Middle-East	GM	4.0–7.6	R <sub>JB</sub>	0–200	0.01–4.0	V <sub>S,30</sub>	SS, N, R

GM geometrical mean, RotD50 median of orientation-independent ground motion amplitudes (Boore 2010), R<sub>epi</sub> epicentral distance, R<sub>JB</sub> Joyner–Boore distance, R<sub>hyp</sub> hypo-central distance, R<sub>rup</sub> distance from the rupture plane, N normal fault, R reverse fault, SS strike-slip fault, U undefined style of faulting

<sup>a</sup>Most of the GMMs are also calibrated for Peak Ground Velocity (PGV), with the exception of the model IDR14 and ZHA06

<sup>b</sup>The GMMs coefficients up to 4 s by ITA10 and BND14 have been provided by the Authors, even if they are not published

<sup>c</sup>The EC8 site categories are defined in the Eurocode 8 part 1 (CEN 2003) and are mainly based on the average shear wave velocity in the uppermost 30 m, V<sub>S,30</sub>

<sup>d</sup>The AB10 predictions at 4 s (not provided by the Authors) are linearly extrapolated from those at 2 and 3 s

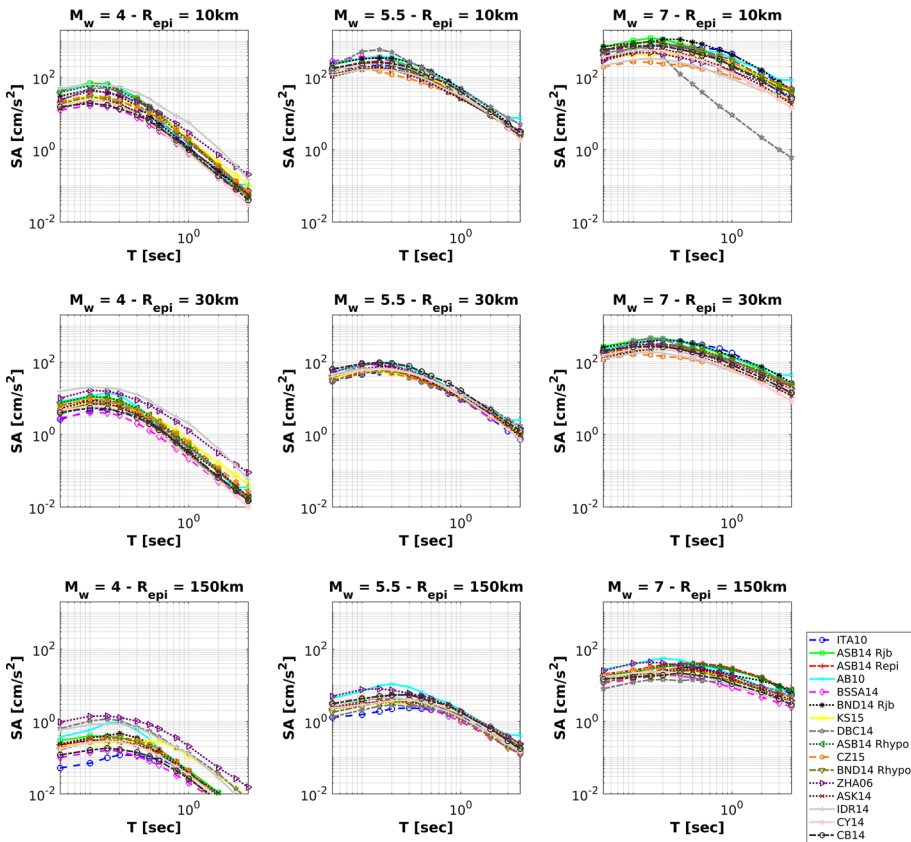
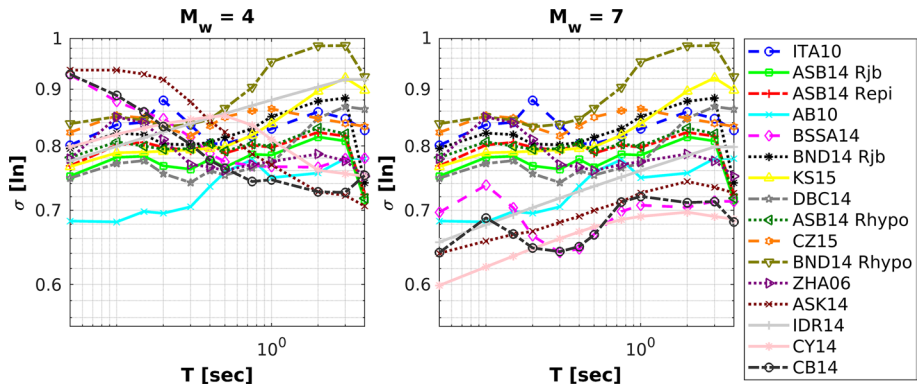


Fig. 1 Trellis plot of the median values of the GMMs pre-selected for ASCRs

located in alluvial plains and, for simplicity, is set as “unknown” for all the predictions of NGA-West2 models.

The trellis charts have been used by Stewart et al. (2015) to display the multidimensional predicted ground motion space to provide insight into the preselected GMMs. The plots of the median values allow the identification of outliers or models with similar behaviour. There are no anomalous predictions among the pre-selected GMMs, except for the model DBC14 (Deraras et al. 2014) for short distances (10 km) and high magnitudes ( $M_w = 7.0$ ).

In order to guide the selection of models to capture epistemic uncertainty, Fig. 2 shows the trend of the total standard deviation,  $\sigma$ , with period, considering two different magnitudes. The NGA-West2 models are heteroskedastic and the corresponding standard deviations decrease with increasing magnitudes; in the other cases,  $\sigma$  is constant. Note that the BND14 model in hypocentral distance exhibits large variabilities at long periods.



**Fig. 2** Standard deviation versus period of the GMMs pre-selected for ASCRs at  $M_w=4.0$  (left) and  $M_w=7.0$  (right)

### 3 ASCRs datasets

The dataset of recordings is derived from ITACA (ITalian ACcelerometric Archive <https://itaca.mi.ingv.it>; Luzi et al. 2008, Pacor et al. 2011a) and ESM (Engineering Strong Motion Database <https://esm.mi.ingv.it>; Luzi et al. 2016) databases. The Active Shallow Crustal Regions (ASCRs) dataset matches the following criteria:

- The events have focal depth shallower than 30 km;
- Style of faulting and  $M_w$  determined by moment tensor inversion, using the Regional Centroid Moment Tensor (Pondrelli et al. 2011), or, when not available, the Time Domain Moment Tensor (Scognamiglio et al. 2009; Krieger and Heimann 2012).

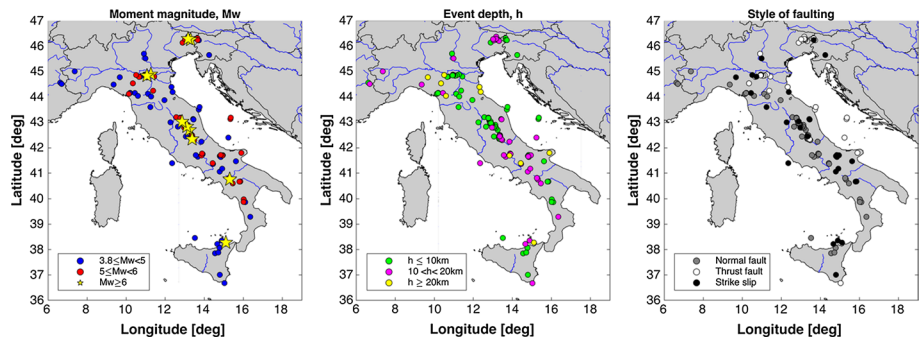
Starting from the dataset used to derive the GMM by Bindi et al. (2011) we exclude several analog records of poorly sampled events or lacking the style of faulting. We also add all the events, since 2009, with  $M_w \geq 5$  and the well-sampled events (more than 15 records) with  $4 \leq M_w \leq 5$ , uniformly distributed on the national territory.

The events having  $M_w \geq 6.5$  in Italy are represented only by the 1980 Irpinia ( $M_w$  6.9) and the Norcia earthquakes, which is the 2nd mainshock of the 2016 Central Italy sequence ( $M_w$  6.5). We increase the dataset in order to extend the maximum magnitude by adding well-sampled foreign events with  $M_w$  up to 7.5 and known fault geometry. In particular, we select three events in Japan and three pan-European earthquakes (one in Greece, one in Turkey and one in Iran). The metadata of the foreign events are derived from the NEar-Source Strong-motion (NESS) flatfile (Pacor et al. 2018).

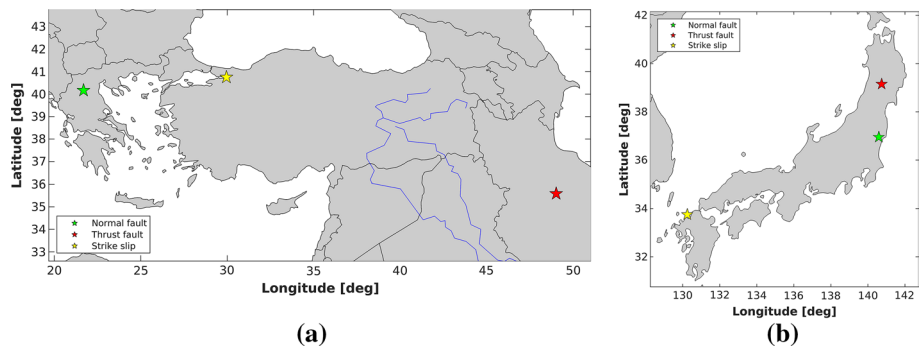
The parametric table (named *flatfile*) of the testing dataset is available in csv format in the electronic supplement ESUPP1. The table is organised following the template of the Engineering Strong Motion (ESM) *flatfile* (<https://esm.mi.ingv.it/flatfile-2018/>; Lanzano et al. 2019) and contains event and station metadata and intensity measures of the records. The meaning of the different fields (columns) of the *flatfile* is given in a supporting pdf document in the ESUPP1.

The location of the Italian events selected for the models scoring is shown in Fig. 3.

The shallow active crustal seismicity in Italy is mainly represented by events located along the Apennine chain, characterized by normal style of faulting and focal depth



**Fig. 3** Maps of the shallow crustal earthquakes in Italy, selected for the scoring of candidate GMMs as a function of moment magnitude, event depth and style of faulting



**Fig. 4** Maps of the pan-European (a) and Japanese (b) shallow crustal earthquakes selected for the scoring of candidate GMMs

lower than 10 km (1980 Irpinia; 1996–1997 Umbria-Marche; 2009 L'Aquila; 2016–2017 Central Italy).

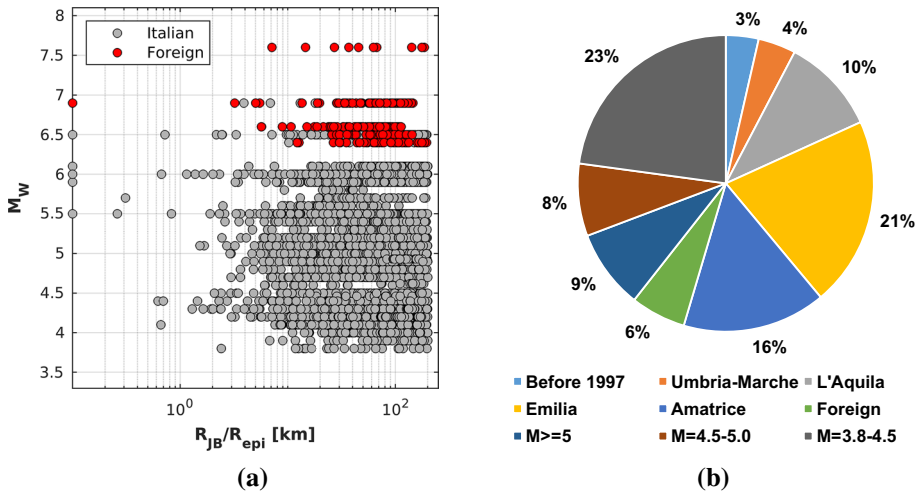
The largest events in Northern Italy (Po Plain and Eastern Alps) are characterized by thrust faults and are related to two major seismic sequences (1976–1977 Friuli; 2012 Emilia). Despite the large hazard in Southern Italy (i.e. Calabria and Eastern Sicily according to the MPS04 model), no events with  $M_W > 6.0$  occurred in the past 50 years, with the exception of the offshore  $M_W$  6.0 1978 Patti earthquake, characterized by strike-slip faulting.

The pan-European and Japanese events, included in the *flatfile*, are reported in Fig. 4.

The largest event is the 1999  $M_W$  7.6 Izmit (Turkey) earthquake, characterized by strike-slip focal mechanism, while the reverse fault event with the largest magnitude is the 2008  $M_W$  6.9 Eastern Honshu earthquake in Japan. The 1980 Irpinia ( $M_W$  6.9) earthquake is still the largest normal fault event in this dataset, as well as in the NGA-West2 dataset.

Figure 5a shows the magnitude-distance distribution of the dataset: the records are 4658, for 137 events and 1187 stations. Figure 5b reports the percentage contributions of different data into the testing dataset. The latest seismic sequences (2009 L'Aquila, 2012 Emilia and 2016 Amatrice-Central Italy) contribute to about half of the dataset





**Fig. 5** **a** Moment magnitude–distance distribution; **b** Pie chart of the composition of the dataset. Distance is  $R_{JB}$  for the events in Table 2, otherwise  $R_{epi}$

(47%). The 6 foreign events have been recorded by 274 stations for a total of 266 waveforms, corresponding to the 6% of the total dataset.

The testing dataset partially overlaps those used to derive the selected GMMs of Table 1. In particular, there is a 17% overlap with Bindi et al. (2011), 10% overlap with the RESORCE dataset (Akkar et al. 2014a), used to derive several models such as Bindi et al. (2014) and Akkar et al. (2014b), and approximately 5% overlap with NGA-West2 database (used to derive Boore et al. 2014, Abrahamson et al. 2014 among others). The percentage is less than 5% for the models by Zhao (2006) and Cauzzi et al. (2015).

### 4 Scoring methods

In decision theory, a scoring rule measures the accuracy of probabilistic predictions. It is applicable to problems in which predictions must assign probabilities to a set of mutually exclusive outcomes. A scoring rule should be also proper, i.e. the forecaster model maximizes the score for the set of data used to generate the model itself. It is also strictly proper if the maximum is unique (Gneiting and Raftery 2007).

For the purpose of measuring the performances of GMMs, the scoring is based on the residuals, calculated as the logarithmic difference between observations and predictions and normalized by the total standard deviations, which are assumed to be normally distributed. Since scoring methods already used for GMMs selection may have limitations (Arroyo et al. 2014; Roselli et al. 2016), we decide to adopt multiple scoring techniques to measure the performance of the models, since they should capture different aspects, relevant for the hazard modelling, and help in capturing the epistemic uncertainty associated to GMMs.

We consider the following scoring methods: (1) Log-likelihood (LLH); (2) parimutuel gambling score; (3) quantile score. The latter two have never been used for this purpose and are detailed in the following sections. Additional details for scoring rules implementation are provided in the Appendix.

## 4.1 Log-likelihood

For the first time, Scherbaum et al. (2004) proposed an objective method to select and weight the GMMs using scores based on likelihoods. This proposal and following adjustments (Scherbaum et al. 2009) have become very popular and have been promptly used for the selection of models in the framework of PSHA (Delavaud et al. 2012).

In particular, Scherbaum et al. (2009) derived the selection procedure from an information-theoretic perspective. In this context, the main issue is to determine the distance in the model space between the unknown model, representing the reality, and the candidate model, based on the concept of likelihood for a set of observations. The likelihood gives the probability of the observed data under a model.

Let us assume that we want to measure the likelihood of a model, described by a continuous probability density function  $g(x)$ , given the set of independent observations  $\mathbf{x} = \{x_i\}$  with  $i = 1 \dots N$ . The log-likelihood of the model, given the sample set  $x$  is:

$$\log_b L(\mathbf{x}) = -\frac{1}{N} \sum_{i=1}^N \log_b(g(x_i)) \quad (1)$$

where  $N$  is the number of observations and the base  $b$  of the logarithm is set to 2 to obtain the results in bit. This method calculates the average log-likelihood for  $x$ , given by a normal density function characterized by its median and corresponding variance. The method is implemented for minimizing the LLH value when the model has the best performance.

The quality of LLH depends on the relationship between the model  $g$  and the set of observations  $x$ . If  $x$  has been used for the derivation of the model (e.g. for determination of the parameters), usually by maximizing the likelihood, the estimator is known to be biased or overfitted (Akaike 1973; Forster and Sober 1994). In this case the log-likelihood will underestimate the true information loss (the model will appear better than it really is) by an amount that depends on the degree of freedom of the model. Overfitting can be corrected by subtracting  $n/N$  to the LLH estimate, where  $n$  is the complexity of the model (here, the degree of the polynomial) and  $N$  the sample size. Moreover, Kale and Akkar (2013) observed that LLH method may lead to assigning an artificially better performance score to the predictive models with larger sigma if the observed data are accumulated away from the median estimations, because models with larger standard deviations can predict outlier observations with higher probabilities. Finally, Arroyo et al. (2014) pointed out that the introduction of  $N$  in Eq. (1), included by Scherbaum et al. (2009) to obtain a measure that is independent of the sample size, is acceptable for model ranking, but it may be critical for model weighting (see also discussion in Delavaud et al. 2012).

## 4.2 Parimutuel gambling score

The comparison of the performance of several GMMs can be considered in terms of gambling or betting. Suppose that there are  $m$  bettors (or  $m$  models) and that at the end of every game they split the total sum of the bets ( $m$ , because each of them plays 1) in a way that reflects their forecast skill. The return of each bettor is the ratio of the amount that the bettor wagered on the outcome to the total amount wagered on the outcome, multiplied by the size of the 'pot'; hence the net return for the  $j$ -th forecast/model is denoted as the parimutuel gambling score (Zechar and Zhuang 2014):

$$G_j = \frac{1}{N} \sum_{i=1}^N \left( -1 + m p_{ij} \frac{1}{\sum_{t=1}^m p_{tj}} \right) \tag{2}$$

where  $N$  is the number of observations,  $p_{ij} = (1 - \Phi)$  is the probability of the normalized residual of being exceeded and is evaluated from the cumulative normal distribution function ( $\Phi$ ).

The original parimutuel gambling score of Zhuang (2010) has been adapted for scoring GMMs in a way that the probabilities of exceedance are calculated for the absolute value of the residuals. Hence, the probability  $p_{ij}$  becomes  $2(1 - \Phi)$ . The gambling scores can be positive or negative and the best performance is reached for the largest positive values. Differently from LLH, gambling scores can be regarded as probabilities, since the results depend on the number of the models (bettors) and the sum of the results of each bet is always 0.

### 4.3 Quantile method

We consider probabilistic forecasts of a continuous quantity (e.g. strong motion variable) that take the form of predictive quantiles. Let assume that  $\alpha$  is the desired quantile,  $r$  is the forecaster quantile and  $\omega$  the observed value, the scoring rule is (Gneiting and Raftery 2007):

$$S(r;\omega) = (\omega - r)(1\{\omega \leq r\} - \alpha) \tag{3}$$

The original quantile method has been adapted for scoring GMMs in a way that we account for the absolute value of the residuals. Hence, the quantile  $\alpha$  becomes  $\alpha_1 = (1 + \alpha)/2$  and  $r_1$  is relative to  $\alpha_1$ . In order to reward the models with positive highest scores, we modified Eq. (3), as:

$$\begin{cases} (\omega - r_1)(1 - \alpha_1) & \text{if } (\omega \leq r_1) \\ (\omega - r_1)(-\alpha_1) & \text{if } (\omega > r_1) \end{cases} \tag{4}$$

The quantile is introduced because it penalizes the residuals on the tails of the distribution to a greater extent than those in the body of the distribution are rewarded. In this analysis, we assume a quantile of level  $\alpha = 95\%$ .

## 5 Results for ASCRs

The three scoring metrics are applied to four different testing datasets in order to verify the stability of the forecasting skill of each GMM and to examine the behavior of the predictive models on the basis of the specific requirements of the MPS19. In particular, the sets of data, derived by initial one, are: (1) the complete dataset (named **All**); (2) a subset of records with stations classified as soil class ‘A’ according to EC8 (CEN 2003), i.e. with  $V_{S,30} > 800$  m/s (named **EC8-A**); (3) a subset of events with magnitude larger than 5.0 and stations located at  $R_{JB} < 50$  km (named **MR**), since these magnitude and distance ranges are those mostly affecting the hazard estimates in Italy; (4) a validation subset without the Italian records preceding 2009 (named **After 2009**), in order to obtain a dataset independent by that used to calibrate the regional model ITA10. The latter subset is included in order to check the overfitting or bias-variance trade-off in the scoring analysis (Forster and Sober 1994). In particular, if a predictive model is tested using the same set of data used to derive

**Table 2** Records, events and stations of the four datasets for the scoring of the candidate GMMs

Dataset	# Records	# Events	# Stations
<i>All</i>	4658	137	1187
<i>EC8-A</i>	891	109	235
<i>MR</i>	1072	66	427
<i>After 2009</i>	3636	61	1051

it (or part of the dataset), its performance tends to be overestimated. Table 2 reports the number of records, stations and events in the testing sets.

The percentage of data of *EC8-A* and *MR* is about the 19% and 23% of the dataset *All*, respectively. In particular, 235 stations are installed at rock sites, according to geophysical measurements or topography proxies, and about 66 events have magnitude larger than 5.0. Moreover, thanks to strong increase of the permanent and temporary monitoring networks in the last years, the records of the *After 2009* subset account for the 78% of *All*, though they are taken from only 61 events out of a total.

The scoring values for each IM (PGA and 12 SA ordinates) and each GMM (see Table 1) are reported in the ESUPP2 for each scoring rule (LLH, gambling and quantile). Figure 6 shows the scoring results as a function of vibration period for the candidate models for *All* and *After 2009* datasets.

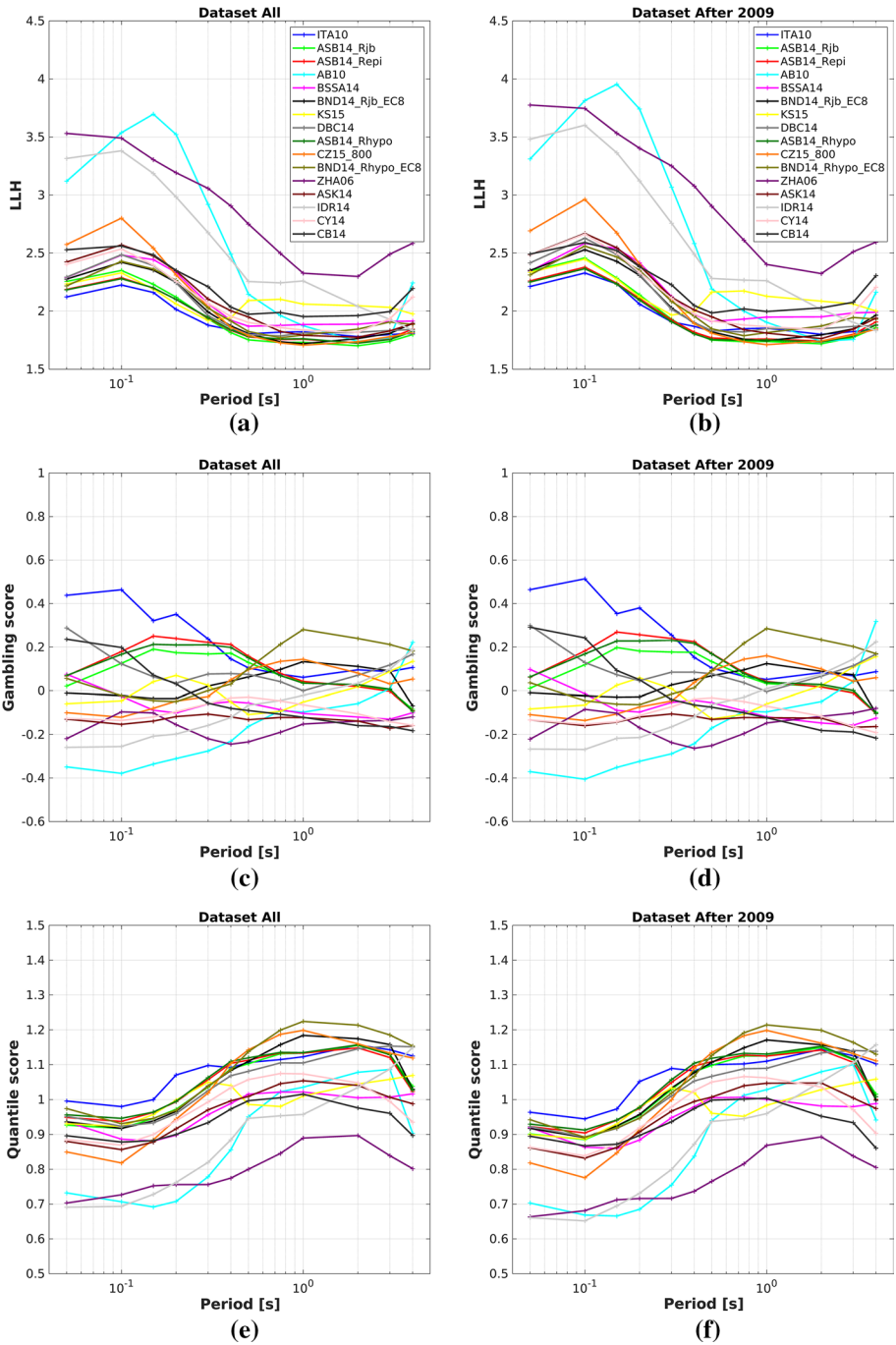
The largest differences in model performance can be observed at short periods, around  $T=0.1-0.2$  s, where the best performing GMM is ITA10. At longer periods the performances are similar among models. Indeed, the dataset of ITA10 comprises a significant amount of analog records (more than 50%) and the high pass filter is generally equal or larger than 0.3 Hz (Pacor et al. 2011b). The results of the subsets *All* and *After 2009* are very similar: this evidence reassures us that the overfitting effects are limited and the results are not significantly affected by the partial overlap between the testing and calibration datasets.

Figure 7 shows the results of the candidate GMMs for *EC8-A* and *MR* datasets. The results of *EC8-A* mirror and emphasize the results of *All* and *After 2009*. In particular, the gap between the best and the worst performing models is greater, especially at longer periods. The scores calculated using *MR* are lower and more similar than those observed for the other subsets: this evidence is not surprising, since the variability of the ground motion is generally lower at large magnitudes, as observed in the heteroskedastic models for aleatory variability (see for example the GMMs of NGA-West2 in Fig. 2). Finally, in relation to the caveats of the exclusion criteria discussed in the pre-selection section (see Sect. 2), KS15 seems to be not influenced by the linear interpolation of the nearest periods, since no bumps are observed in the scoring trend with period.

## 6 Final selection for ASCRs and weights assignment

For the final selection of the GMMs to be used in MPS19, we consider only the results for *EC8-A* and *MR* subsets, since the two subsets highlight the GMMs performance on the basis of the fundamental requirements of the MPS19 model.

To obtain a global estimate of the performance of each model over all IMs and scoring rules, the scores are averaged considering five ordinates of the acceleration response spectra at periods  $T=0.1, 0.2, 0.5, 1$  and  $2$  s. The choice of the ‘control’ periods is similar to



**Fig. 6** Scoring of the candidate GMMs (see Table 1 for the abbreviations). LLH (a, b), gambling (c, d) and quantile (e, f) for All (a, c, e) and After 2009 (b, d, f) datasets

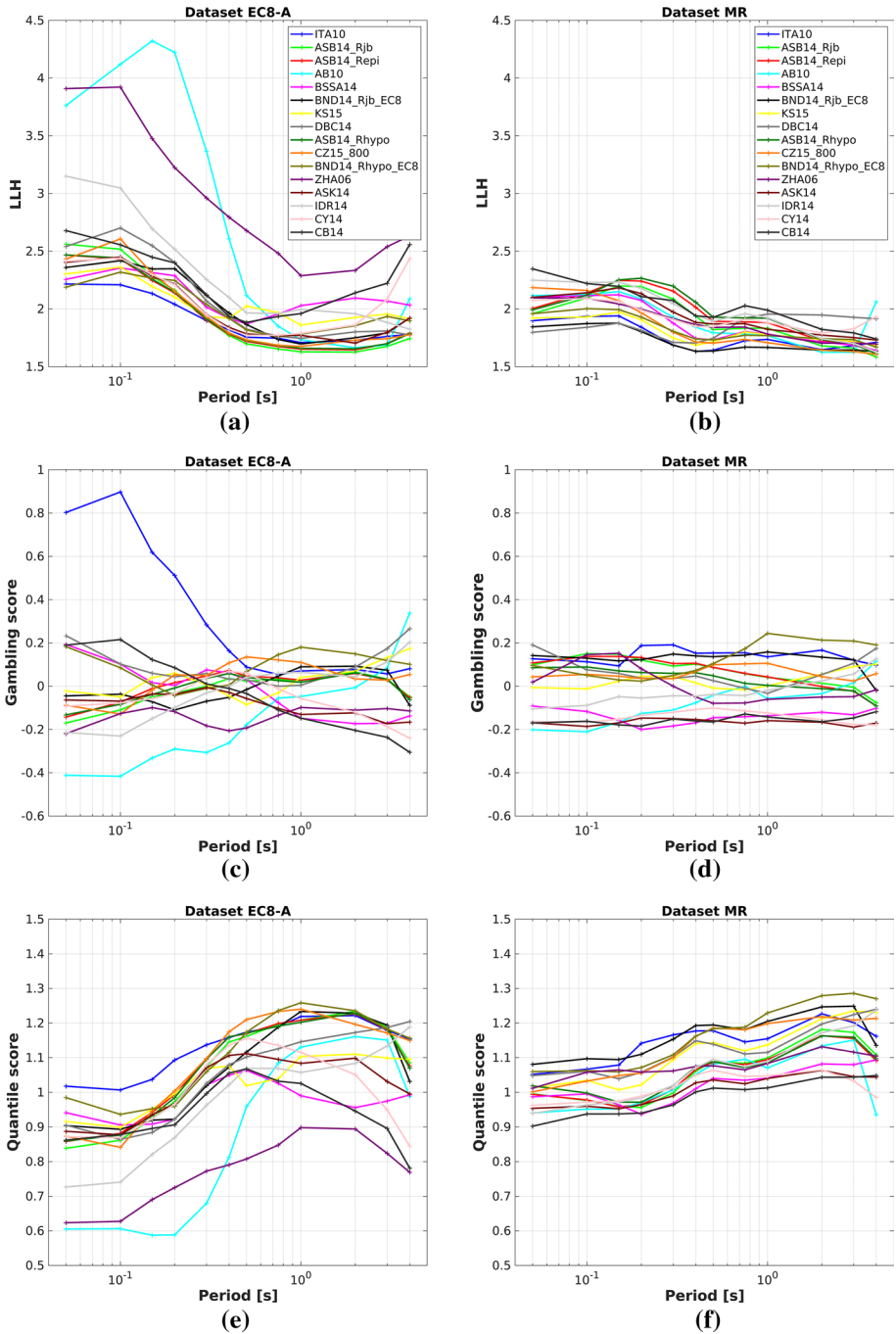


Fig. 7 Scoring of the candidate GMMs (see Table 1 for the abbreviations). LLH (a, b), gambling (c, d) and quantile (e, f) for EC8-A (a, c, e) and MR (b, d, f) datasets

**Table 3** Final rank of candidate GMMs for the dataset *EC8-A*

GMMs	LLH		Gambling		Quantile		Total
	Mean*	UNW	Mean	UNW	Mean	UNW	UNW
<i>ITA10</i>	1.888	16	0.329	16	1.143	16	48
<i>ASB14 Rept</i>	1.918	15	0.015	12	1.096	13	40
<i>BND14 Rhypo</i>	2.004	10	0.088	15	1.113	15	40
<i>CZ15</i>	1.979	12	0.040	13	1.098	14	39
<i>ASB14 Rhypo</i>	1.926	14	0.006	10	1.093	12	36
<i>ASB14 Rjb</i>	1.926	13	-0.005	8	1.088	11	32
<i>DBC14</i>	2.095	6	0.053	14	1.041	9	29
<i>BND14 Rjb</i>	2.015	8	0.004	9	1.079	10	27
<i>KS15</i>	2.055	7	0.007	11	1.025	6	24
<i>ASK14</i>	1.985	11	-0.084	3	1.028	7	21
<i>CY14</i>	2.014	9	-0.046	4	1.032	8	21
<i>BSSA14</i>	2.127	5	-0.037	5	0.967	5	15
<i>CB14</i>	2.167	4	-0.017	7	0.965	4	15
<i>IDR14</i>	2.297	3	-0.036	6	0.964	3	12
<i>AB10</i>	2.771	2	-0.188	1	0.889	2	5
<i>ZHA06</i>	2.889	1	-0.129	2	0.790	1	4

UNW un-normalized weight

\*The scores are averaged over 5 periods of the SA (0.1, 0.2, 0.5, 1 and 2 s)

that proposed by Delavaud et al. (2012) and is necessary to equally account for the contribution of different periods of interest for engineering applications.

In order to combine, in an objective way, the three different scoring metrics, an ordinal un-normalized weight for each scoring metric is used, accounting for their cardinality. Moreover, our goal is to select few well-fitting models that capture the epistemic uncertainty, and sometimes the tested models exhibit very similar scores (few centesimal points of difference), that makes difficult to express a preference. In such sense, although a weighting scheme based on the cardinality enlarges the differences between the models, it is useful to produce a global rank and to support the final selection.

In particular, we assign 16 points to the best performing GMM and 1 point to the worst, on the basis of the scoring results for each method. The total un-normalized weight for each GMM (and for each subset) is then the sum of these three ordinal un-normalized weights. Tables 3 and 4 provide the average values of the scores, the un-normalized weights and the sum of each GMM for *EC8-A* and *MR* subsets, respectively.

ITA10 has a much better performance than the other models for *EC8-A* sites and it is particularly evident from the results of the gambling score. The GMMs that are mostly penalized are the European AB10 and the Japanese ZHA06. AB10 was calibrated only up to 100 km and for magnitude larger than 5.0 and, in addition, is characterized by a total  $\sigma$  lower than the other models. The performance of ASB14 (supposed to be an update of AB10) is generally better than AB10, with the exception of dataset *MR*. It suggests that ASB14 better captures the data attenuation with distance (distances larger than 50 km), while AB10 predicts well the large magnitude earthquakes in near fault conditions. In the case of ZHA06, the poor performance is probably related to the characterization of the sites: ZHA06 is based on a simplification of the NEHRP classification (BSSC 2001), i.e. it

**Table 4** Final rank of candidate GMMs for the dataset *MR*

GMMs	LLH		Gambling		Quantile		Total UNW
	Mean*	UNW	Mean	UNW	Mean	UNW	
<i>BND14 Rjb</i>	1.724	16	0.136	15	1.171	16	47
<i>ITA10</i>	1.760	15	0.151	16	1.154	14	45
<i>BND14 Rhypo</i>	1.835	13	0.127	14	1.165	15	42
<i>CZ15</i>	1.836	12	0.068	11	1.138	13	36
<i>KS15</i>	1.812	14	0.016	8	1.110	11	33
<i>DBC14</i>	1.861	11	0.032	9	1.114	12	32
<i>ASB14 Rjb</i>	1.925	6	0.082	13	1.060	7	26
<i>ZHA06</i>	1.883	8	0.007	7	1.082	10	25
<i>ASB14 Repi</i>	1.975	4	0.079	12	1.057	6	22
<i>ASB14 Rhypo</i>	1.994	2	0.038	10	1.061	9	21
<i>AB10</i>	1.880	10	-0.094	5	1.042	5	20
<i>IDR14</i>	1.993	3	-0.035	6	1.061	8	17
<i>BSSA14</i>	1.883	9	-0.144	3	1.019	3	15
<i>CY14</i>	1.920	7	-0.136	4	1.027	4	15
<i>ASK14</i>	1.948	5	-0.164	1	1.013	2	8
<i>CB14</i>	2.021	1	-0.164	2	0.989	1	4

UNW un-normalized weight

\*The scores are averaged over 5 periods of the SA (0.1, 0.2, 0.5, 1 and 2 s)

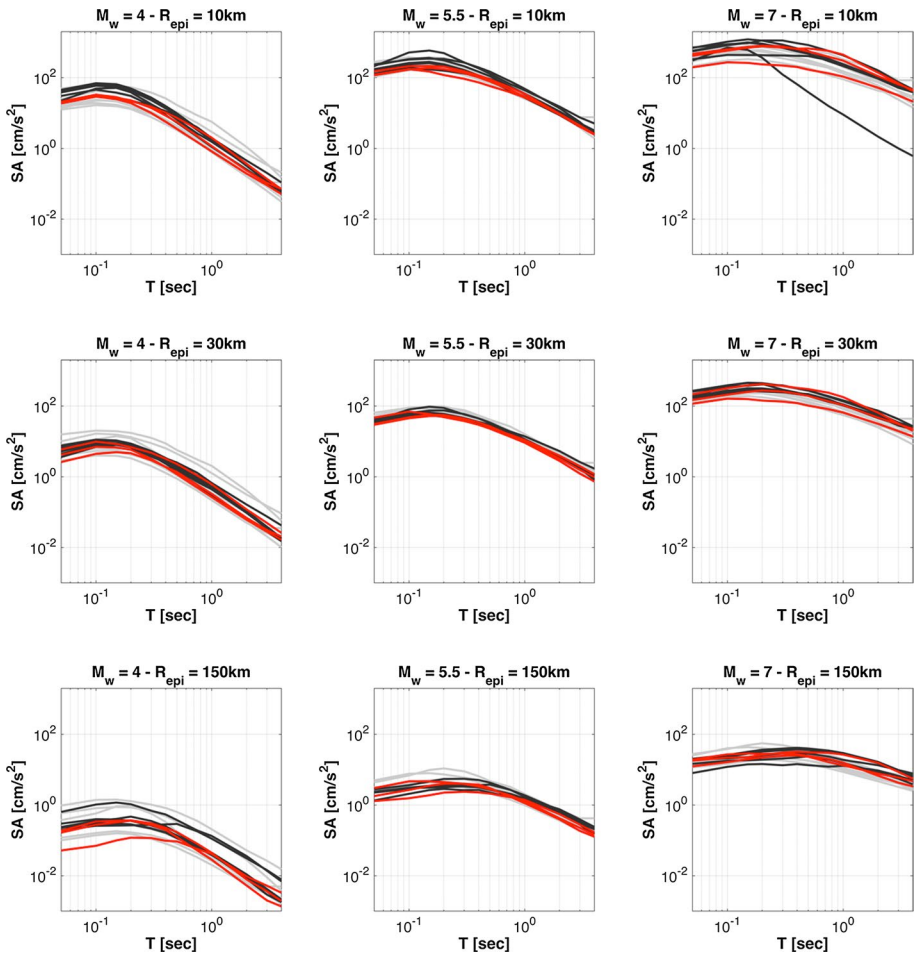
merges classes that in the other models are separated (EC8-A up to 1100 m/s and EC8-B are in the same class). Conversely, the CZ15 classification is coherent with Italian and European classification schemes and the model exhibits a good performance, despite the calibration dataset is composed by many Japanese earthquakes.

As previously highlighted, the performance of the different models against *MR* subset is almost equivalent and the differences are small. This is also expected, since the near fault data are less affected by regionalization (Pacor et al. 2018). The relatively poor performance of the NGA-West2 GMMs (ASK14, BSSA14, CY14, CB14 and IDR14) for *MR*, emphasized by the score's cardinal assignment, is mainly due to the lower heteroskedastic  $\sigma$  at largest magnitudes (see Fig. 2), but it could be also related to the fact that several parameters (fault geometries, basin depths) of such models are not available in Italy and have been set as unknown or inferred from other parameters (according to Kaklamanos et al. 2011).

Summing up the total un-normalized scores of *EC8-A* and *MR* subsets, the grand-total score is obtained: (#1) **ITA10** (93 points); (#2) **BND14 Rhypo** (82 points); (#3) **CZ15** (75 points); (#4) BND14 RJB (74 points); (#5) ASB14 Repi (66 points); (#6) DBC14 (61 points); (#7) ASB14 RJB (58 points); (#8) ASB14 Rhypo and KS15 (57 points); (#9) CY14 (36 points); (#10) BSSA14 (30 points); (#11) ASK14, IDR14 and ZHA06 (29 points); (#12) AB10 (25 points); (#13) CB14 (19 points).

The first three models (in bold) are selected for the seismic hazard estimation: worthy of note, without any a priori imposition, the selected GMMs are calibrated on different datasets, since the first one is regional (Italian), the second one is European, and the third one is global. The fourth-scoring GMM is a version close to the BND14 Rhypo, but considering a different distance metric, and the fifth one (ASB14 Repi) has a nonlinear description of the





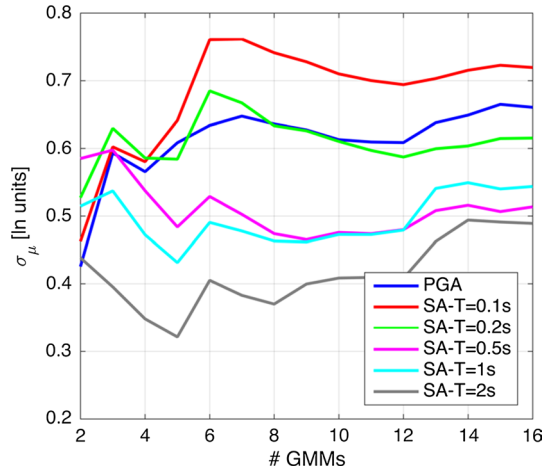
**Fig. 8** Trellis plot of the median predictions of the pre-selected models (Table 1) for EC8 class A and normal faulting. Red curves are the selected GMMs; the dark gray curves are associated to the GMMs with grand-total score from 57 to 74; the light gray lines are the remaining GMMs with grand-total score from 19 to 36

site effects but is calibrated on the same dataset of BND14. We propose to select the first three GMMs because, besides to show the best forecasting skill, they use different metrics for the distance.

Figure 8 shows the three selected GMMs with respect to the unselected models: for epicentral distances of 30 and 150 km, the median prediction of the selected GMMs (red lines) describes quite well the variability of all GMMs except for the worst performing ones (light gray lines). The variability of ground shaking assuming M<sub>w</sub> = 4.0 and epicentral distances of 150 km is not well described, but these magnitude-distance values are of scarce interest for PSHA purposes.

For epicentral distances of 10 km, we note that the dispersion among the selected GMMs describes reasonably well the variability of all the other models for spectral periods above about 0.5 s; exceptions are associated with the few worst performing

**Fig. 9** Model-to-model variability (Eq. (5)) as a function of the number of the GMMs ( $np$ ) for different intensity measures. #GMMs = 3 corresponds to the models selected for hazard calculation



GMMs (light gray lines) for  $M_W=4.0$  earthquakes. The most remarkable difference is for spectral periods below 0.5 s; in this case, the selected GMMs yield lower ground shaking with respect to most of the other ones, in particular for the small and intermediate magnitude earthquakes ( $M_W=4.0$  and 5.5).

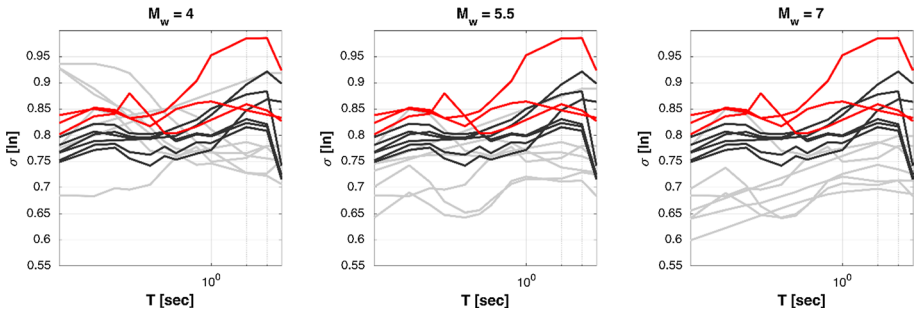
In order to quantify the uncertainty related to the model predictions, we compute the model-to-model variability,  $\sigma_\mu$ , following the approach proposed by Al Atik and Youngs (2014). The best performing model, i.e. ITA10, is considered as a reference model and  $\sigma_\mu$  is calculated as:

$$\sigma_\mu = \sqrt{\frac{\sum_{M=1}^{nm} \sum_{sof=1}^{ns} \sum_{R=1}^{nr} \sum_{i=1}^{np} (\mu_{i(M,R,sof)} - \mu_{ref(M,R,sof)})^2}{nm * nr * ns * np}} \tag{5}$$

where  $\mu$  is the median prediction of the GMMs, the subscript  $i$  refers to the models from 2nd to 16th place of the final rank and the subscript  $ref$  to ITA10;  $nm=9$  is the number of magnitude bins, in the range 4.0–8.0;  $nr=34$  are the distance bins, with a logarithmic spacing in the range 0–200 km;  $ns=3$  is the number of focal mechanisms, i.e. normal, reverse and strike-slip styles of faulting. The predictions for the EC8-A site category or, alternatively,  $V_{S,30}=800$  m/s are only considered. The number of the GMMs ( $np$ ) varies from 2 (first two ranked models) to 16 (all the preselected models): as  $np$  increases, the models are added following strictly the ranking results (Fig. 9).

The ratio between the  $\sigma_\mu$  obtained for three selected models ( $np=3$ ) and the maximum value obtained from the pre-selected models varies from 79% (SA-T=0.1 s) to 100% (SA-T=0.5 s). Moreover,  $\sigma_\mu$  does not grow monotonously as increases by  $np$  and reaches maximum variability at different  $np$  values for each intensity measure considered.

An interesting aspect is how the three selected GMMs describe the aleatory variability: in Fig. 10 we show the standard deviation,  $\sigma$ , for each GMM. The three selected models (red lines) have nearly always the largest  $\sigma$ , which may explain the good forecasting skill. This result is in agreement with Roselli et al. (2016) that analyzed the performance of several GMMs applied to the Italian territory and found that the forecasting performance is correlated with  $\sigma$  as shown in Fig. 10.



**Fig. 10** Total standard deviation  $\sigma$  as a function of the spectral periods for different magnitudes. Red curves are the selected GMMs; the dark gray curves are associated to the GMMs with grand-total score from 57 to 74; the light gray lines are the remaining GMMs with grand-total score from 19 to 36

The weight of each selected GMM is a combination of two different evaluations: (1) a weight based on the testing phase described above,  $W_1$ , and (2) a weight derived by expert judgment,  $W_2$ . The normalized weight  $W_1$  for each selected model  $i$ ,  $W_{1,i}$ , is calculated as:

$$W_{1,i} = \frac{UNW_i}{\sum_{i=1}^3 UNW_i} \tag{6}$$

where  $UNW_i$  is the grand-total score of the  $i$ th model and  $i = 1, 2$  and  $3$  corresponds to ITA10, BND14 and CZ15, respectively.

In order to preserve homogeneity, an experts’ elicitation is carried out following the same procedures adopted for the scoring of the seismicity rate models in MPS19 and for the probabilistic tsunami hazard map in the Mediterranean area (TSUMAPS-NEAM; Basili et al. 2019), which are both rooted in a procedure that is named Analytic Hierarchy Process (AHP). AHP is a multi-criteria decision-making method that is useful to make decisions under complex problems (Saaty 1980), breaking down into a series of pairwise comparisons and taking also into account for the subjectivity in the decision. Moreover, the process includes techniques for consistency checks the of the expert opinions, thus reducing the bias in the process. To this purpose, seven experts have been selected among the scientific coordinators of the project not involved in any of the selected GMMs. Specifically, experts have been elicited on scoring the GMMs on their reliability to describe the ground shaking in Italy via a questionnaire (Basili et al. 2019). The outcomes of the elicitation process are expressed in terms of weights,  $W_2$ , similarly to those obtained from the objective scoring  $W_1$ .

The final weight of the selected GMMs is calculated as:

$$W_i = W_{1,i} \times W_{2,i} / C (i = 1, 2, 3) \tag{7}$$

where  $C$  is a normalization factor that sets  $W$  in the range  $[0,1]$ . The normalized weights from scoring,  $W_1$ , the experts’ elicitation weights,  $W_2$  and the final weights,  $W$ , are given in Table 5: the weights obtained from scoring are quite similar and ITA10 is slightly favored. Instead, it appears clear that experts tend to decisively favor the regional model with respect to models calibrated on a wider set of data.

**Table 5**  $W_1$ : normalized weight from scoring;  $W_2$ : experts' elicitation weight;  $W$ : final weight according to Eq. (7)

<i>GMMs</i>	$W_1$	$W_2$	$W$
ITA10	0.37	0.423	0.45
BND14 Rhypo	0.33	0.328	0.32
CZ15	0.3	0.259	0.23

## 7 Selection and weighing of GMMs for the subduction zone (SZ)

The subduction zone (SZ) in Italy corresponds to the Calabrian Arc, located in the southern Tyrrhenian sea. The geometry of the subduction surface has been reconstructed by Maesano et al. (2017) for TSUMAPS-NEAM and has been also adopted in MPS19. The scoring procedure, detailed in the previous sections, is repeated for selecting the GMMs to predict the ground motion in the subduction zones. The SZs candidate models are reported in Table 6.

The list of GMMs, after the application of the rejection criteria (see Sect. 2), includes models for Hellenic Arc (Skarlatoudis et al. 2013), Japan (Zhao 2006) and for worldwide events (Abrahamson et al. 2016), while there are no GMMs specifically derived for Italy. Several existing models are disregarded, mainly for the minimum magnitude of the GMM that is generally larger than 6.0. Although the functional form does not include a non-linear scaling with magnitude, we keep in the list HEL13, because it is the only model calibrated for the Mediterranean area and we want to check the similarities between the subduction of the Hellenic and Calabrian arc in terms of ground motion.

Trellis plots of the median predictions and standard deviations,  $\sigma$ , of the candidate GMMs are reported in the electronic supplement (ESUPP3): the total  $\sigma$  of each candidate model are generally higher with respect to those observed for ASCRs. Note that HYDR15 has a lowest constant  $\sigma$  over all periods (0.74 ln units).

For the strong motion testing dataset, we collect the records of the events with depth larger than 40 km in Ionian Sea and Calabria and the deepest events (focal depths larger than 80 km) in the Southern Tyrrhenian Sea. Figure 11 shows the location of the events and the magnitude-distance distribution.

The depth distribution of subduction events shows a lack of seismicity in the range 80–150 km, also confirmed by the earthquakes distribution analysed in Chiarabba et al. (2008) and Maesano et al. (2017), and considered in the seismogenic source model of MPS19. The maximum (moment) magnitude of the dataset is 5.8 and corresponds to the 26/10/2006 earthquake, located in the Southern Tyrrhenian Sea with focal depth  $h=220$  km.

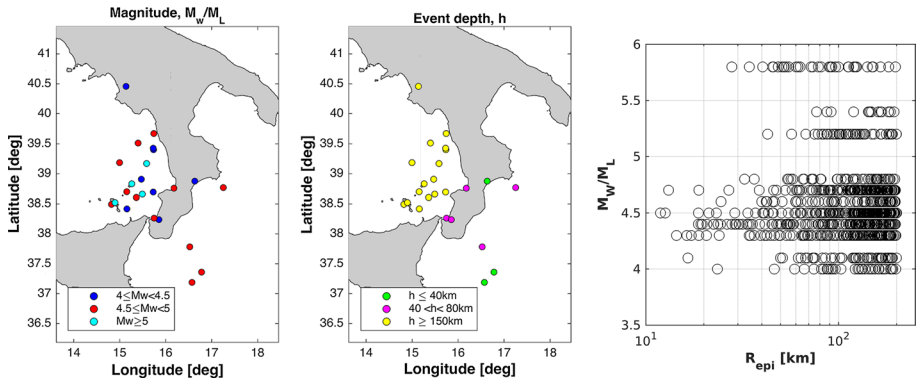
For the residual calculation, the events of the dataset must be classified in *interface* and *in-slab*: “*interface* earthquakes are shallow angle thrust events that occur at the interface between the subducting and overriding plates; *in-slab* earthquakes occur within the subducting oceanic plate and are typically high-angle, normal-faulting events responding to downdip tension in the subducting plate” (Youngs et al. 1997). From engineering point of view, “*in-slab* events produce on average larger near-source amplitudes than *interface* events, but these amplitudes attenuate much faster” (Beauval et al. 2012).

The classification schemes to discern between *interface* and *in-slab* events, proposed by Zhao et al. (2015) and Garcia et al. (2012) among others, are not applicable to the SZs dataset, because several event metadata (e.g. focal mechanism) are lacking or of poor quality. We prefer to adopt the simplified classification by Skarlatoudis et al. (2013), mainly

**Table 6** List of the candidate GMMs for the application to SZs for PGA and acceleration response spectra ordinates

Name	Code	Area	H comp	M range	R type	R range [km]	Period range [s]	Site-effect term
Zhao (2006)	ZHA06	Japan	GM	5.0–8.3	$R_{rup}$	0–300	0.05–5.0	5 classes
Skarlatoudis et al (2013)	HEL13	Hellenic Arc (Greece)	RotD50	4.4–6.7	$R_{hyp}$	65–850	0.01–4.0	3 classes
Abrahamson et al (2016)	HYDR15	Worldwide	GM	5.0–7.9 6.0–8.4	$R_{rup}$	10–300	0.02–10.0	$V_{s,30}$

GM geometrical mean, RotD50 median of orientation-independent ground motion amplitudes (Boore 2010),  $R_{hyp}$  hypocentral distance,  $R_{rup}$  distance from the rupture plane



**Fig. 11** Maps of the location of the events selected for the scoring of GMMs for SZs. Left: magnitude; centre: focal depth. Right: magnitude–distance distribution of the testing dataset

based on depth and spatial location, considering the geometry of the model proposed by Maesano et al. (2017): *interface* events are located in Calabria and Ionian Sea with hypocentral depths in the range 20–40 km; *in-slab* events have focal depth larger than 30 km and are located in the Southern Tyrrhenian Sea.

Adopting this classification scheme, the *interface* seismicity is poorly represented in the dataset (only 5 events) and we prefer to restrict the ranking analysis only to the *in-slab* events, consistently with MPS19 choice to consider only the slab surface as a seismicogenic source in the hazard model configuration.

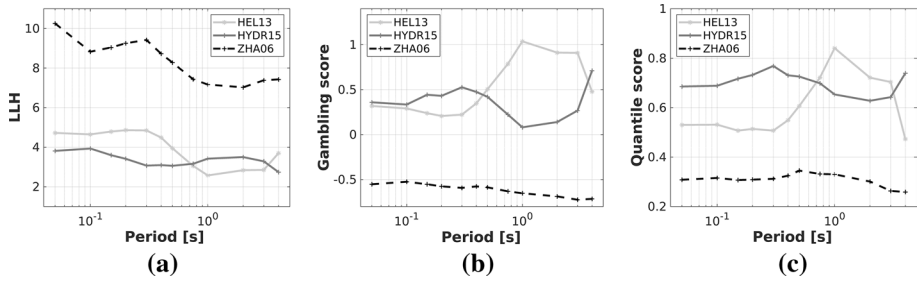
HYDR15 and HEL13 also require, as input parameter, a flag to identify stations in *back-arc* or *fore-arc* position, since they are characterized by different rate and characteristics of ground motion attenuation (Abrahamson et al. 2016): the *fore-arc* region is located between the subduction trench axis and the axis of volcanic front, while the *back-arc* region is on the opposite side of the volcanic front. In our case, the stations located in the islands of the Tyrrhenian sea (i.e. Stromboli, Lipari, Ustica) are classified as *back-arc*, while the remaining sites are in *fore-arc* region.

The SZs dataset is composed by 923 waveforms of 24 *in-slab* events, recorded by 100 stations. The *flatfile* is available in the electronic supplement ESUPP4 and is arranged according to the same structure of ASCRs *flatfile* (ESUPP1). The data selection is carried out considering waveforms from accelerometers and broadband instruments of events with local magnitude larger than 4.0. In order to avoid oversampling, we prefer to use the records of broadband instruments rather than those of co-located accelerometers.

Figure 12 shows the trend of the scoring values with period for the candidate models of Table 7. The scoring values for each IM (PGA and 12 SA ordinates) and each GMM are also reported in the ESUPP5 for each scoring rule.

As a general comment, the values of LLH and quantile scores show a worse performance of GMMs for SZs compared to those observed for ASCRs and this evidence can be caused by the absence of a model tailored for the area under investigation. The performance of HEL13 and HYDR15 are quite equivalent: HYDR15 has a better performance at short periods, while HEL13 at long periods. ZHA06 has instead a poor performance at all periods.

Table 7 shows the mean values and the un-normalized weights of the candidate models for each scoring rule. The scores are summed up in order to obtain a final rank and



**Fig. 12** Scores of the models selected for SZs as function of period. **a** LLH; **b** gambling parimutuel score; **c** quantile score

**Table 7** Final rank of GMMs for SZs

GMMs	LLH		Gambling		Quantile		Total UNW	W <sub>1</sub>	W <sub>2</sub>	W
	Mean*	UNW	Mean	UNW	Mean	UNW				
<i>HEL13</i>	3.465	3	0.282	2	0.686	3	8	0.44	0.50	0.53
<i>HYD15</i>	3.773	2	0.591	3	0.643	2	7	0.39	0.50	0.47
<i>ZHA06</i>	8.113	1	-0.605	1	0.320	1	3	0.17	-	-

UNW un-normalized weight, W<sub>1</sub> normalized weight from scoring, W<sub>2</sub> experts’ elicitation weight, W final weight according to Eq. (7)

\*The scores are averaged over 5 periods of the SA (0.1, 0.2, 0.5, 1 and 2 s)

the normalized weights, W<sub>1</sub>, according to Eq. (6). The latter have been combined with the weights obtained by expert judgment, W<sub>2</sub>, using Eq. (7), in order to set the weights, W, to be adopted in the hazard calculation.

Considering that the final weight W<sub>1</sub> of the ZHA06 model is very small, we finally decide to disregard this model and re-calculate the final weights, W, only for the two remaining models (HEL13 and HYD15).

### 8 Conclusions

A strong effort has been spent by the Italian scientific community in the last years to derive the new probabilistic seismic hazard model, named MPS19. In this framework, particular attention is devoted in the selection of the GMMs, testing the reliability of the median prediction and properly assessing the contribution of aleatory and epistemic uncertainties. The general philosophy adopted in MPS19 for the ground motion prediction equations is to take advantage of the many recent models published in literature, without developing specific new models. The procedure is, as much as possible, transparent and repeatable.

We first apply several exclusion criteria, modified after Cotton et al. (2006) and Bommer et al. (2010), to the long list of GMMs provided by John Douglas (<https://www.gmpe.org.uk/>, last access 2016). We add some models that do not strictly meet all the exclusion criteria also to test the robustness of the pre-selection process. A shorter list of 16 models is then obtained.

We use three proper scoring methods, two of them introduced in the framework of MPS19 to test the pre-selected GMMs. The use of multiple scoring rules lies in the aim of climbing some caveats of the most popular methods, such as LLH (Arroyo et al. 2014; Roselli et al. 2016), and helping in the best representation of the epistemic uncertainty related to ground motion predictions.

The testing methods are based on the residuals, calculated as the logarithm difference between the observations and the model predictions, normalized w.r.t. the standard deviation. The experimental data are derived by the Italian and European accelerometric archives (ITACA and ESM) for the shallow active crustal regions (ASCRs) and subduction zones, using, as much as possible, qualified and well-referenced metadata associated. The dataset for ASCRs is mainly composed by events occurred in Italy, plus few foreign events, selected to extend the magnitude of the dataset to 7.5, close to the maximum magnitude used for the PSHA calculation.

In order to highlight the performance of the GMMs as a function of the main requirements of the MPS19 model, the scoring results are obtained on the basis of two subsets: the first one is composed of strong motion records of rock sites (EC8 category A); the second one is composed by records of large magnitude events ( $M_W > 5.0$ ) in near source conditions ( $R_{JB} < 50$  km). The scoring results are still affected by aleatory sigma of the models, picking up the performances of models with larger standard deviations. In such way, GMMs with heteroskedastic sigma, such as NGA-West2 models, may be disadvantaged more than necessary.

A procedure is developed to take into account the results of the various scoring methods using un-normalized weights and to emphasize the differences among the scoring results allowing us to express a preference in the model selection. Three best performing models are finally selected for shallow active crustal regions, that are the regional model for Italy in Joyner–Boore distance (Bindi et al. 2011), a pan-European model in hypocentral distance (Bindi et al. 2014) and a global model in rupture distance (Cauzzi et al. 2015). The weights derived from scoring procedure are combined with those assigned by experts during an elicitation process, in order to have the final weight for each model.

The procedure is repeated for subduction zones in Italy, which correspond to the Calabrian Arc in the Southern Tyrrhenian Sea. The models, selected for the *in-slab* seismogenic source, are those proposed for the Hellenic Arc by Skarlatoudis et al. (2013) and for worldwide subduction events by Abrahamson et al. (2016). However, in this case, the tested GMMs are not able to satisfactorily describe the ground motion in such zone, as returned by the scoring results. Future efforts should be paid in calibrate a regional model or introduce a proper regionalization in existing global models for Calabrian Arc.

**Acknowledgements** The analysis described in this paper was carried out in the framework of the activities of INGV Seismic Hazard Center (*Centro Pericolosità Sismica*, CPS) for producing the MPS19 model. This study has benefited from funding provided by the Italian Presidenza del Consiglio dei Ministri-Dipartimento della Protezione Civile (DPC). This paper does not necessarily represent DPC official opinion and policies. The authors would also like to acknowledge the European Plate Observing System (EPOS) EU project (Grant Agreement 676564) and its coordinator Massimo Cocco to support the activities relative to the Engineering Strong-Motion (ESM) database. This study has also been partially supported by the H2020 project SERA (Seismology and Earthquake Engineering Research Infrastructure Alliance for Europe, Grant Agreement Number 730900). The maintenance and the development of ITACA (ITalian ACcelerometric Archive) database is also funded by the Italian Department of Civil Protection (DPC). A special thank to the expert's panel for the revision of the MPS19 model (Paolo Bazzurro, Domenico Giardini, Francesco Mulargia, Claudio Modena and Silvio Seno) that strongly contributed to improve this study. The authors of the paper of the INGV section in Milan would like to thank their colleagues (Rodolfo Puglia, Chiara Felicetta, Maria D'Amico, Sara Sgobba, and Emiliano Russo) for their daily commitment to the maintenance and



development of the accelerometric databases ITACA and ESM, which constitute the irreplaceable basis of this research work. Finally, we wish to thank Graeme Weatherill and one anonymous reviewer that provided useful comments to this work, that hopefully improved the quality of the paper.

## Appendix

### Gambling score

The residuals are calculated for each model and are normalised to their standard deviation. The probability of the normalised residual not to be exceeded is evaluated from the normal cumulative distribution function ( $\Phi$ ) and the corresponding probability of being exceeded is  $p_{ij} = (1 - \Phi)$ , where  $i$  are the observations and  $j$  the models, as shown in Fig. 13, top panel.

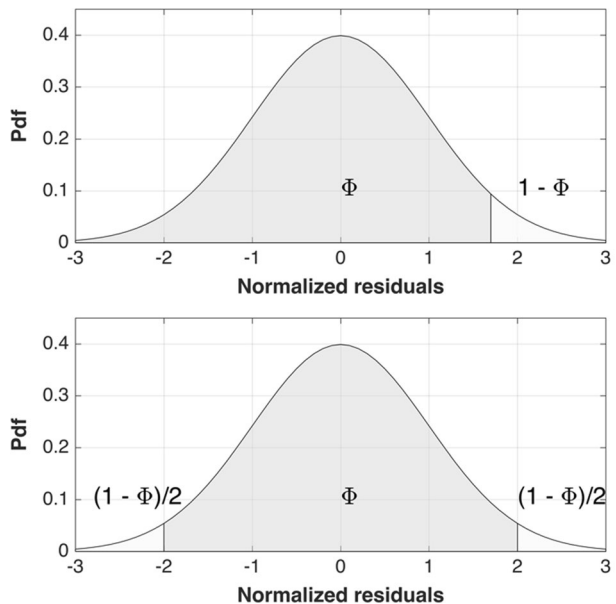
As for GMMs we are interested in the distance from the median of the model, the gambling score model has been modified in a way that the probabilities of exceedance are calculated for the absolute value of the residuals. The probability of not being exceeded is therefore  $p_{ij} = 1 - \Phi$  (Fig. 13, bottom panel). A matrix  $i \times j$  of  $i = 1, \dots, R$  rows ( $R =$  number of observations) and  $j = 1, \dots, m$  columns ( $m =$  number of models) is created (Fig. 14).

The probabilities of not being exceeded of each residual ( $p_{ij}$ ) for each model are normalised to the sum of probabilities for all models:

$$R_{ij} = p_{ij} \frac{1}{\sum_{j=1}^m p_{ij}} \tag{8}$$

Normalised probabilities sum up to 0 as:

**Fig. 13** Normal probability density function of the residuals for single tail (top panel) or double tails test (bottom panel)



**Fig. 14** Matrix of the  $p_{ij} = 1 - \Phi$

<b>P<sub>11</sub></b>	....	...	<b>P<sub>1j</sub></b>	...	....	<b>P<sub>1m</sub></b>
<b>P<sub>i1</sub></b>			<b>P<sub>ij</sub></b>			<b>P<sub>im</sub></b>
<b>P<sub>R1</sub></b>			<b>P<sub>Rj</sub></b>			<b>P<sub>Rm</sub></b>

$$S_{ij} = -1 + mR_{ij} \tag{9}$$

The average score of each model is:

$$G_j = \frac{1}{R} \sum_{i=1}^R S_{ij} \tag{10}$$

All G's sum up to 0 and the maximum G value indicates the best performing model.

### Quantile method

As for the parimutuel gambling score, the quantile score is applied to the normalized residuals. Given a quantile  $q$  of order  $\alpha$  ( $\text{Prob}\{Z \geq q\} = \alpha$ ), the Z-score associated to the quantile is obtained ( $q_\alpha$ ), through the normal inverse cumulative distribution function. The quantile method has been adapted for scoring GMMs in a way that the quantile  $\alpha$  becomes  $\alpha_1 = (1 + \alpha)/2$  and  $r_j$  is relative to  $\alpha_j$  (two tails test in Fig. 13).

If  $z_{\alpha 1}$  is the absolute value of the normalised residual, the score  $g(z_{\alpha 1})$ :

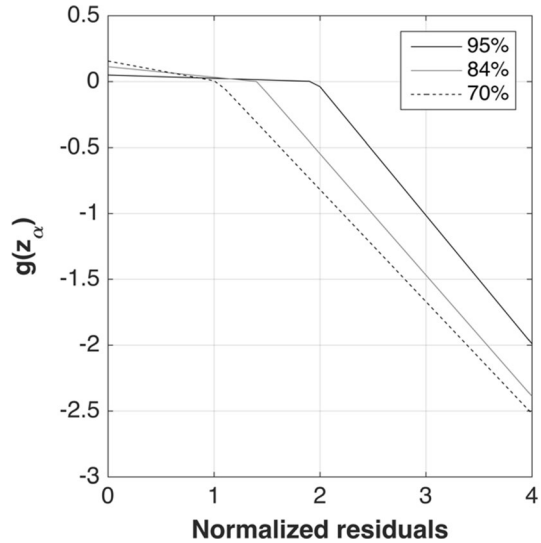
$$g(z_{\alpha 1}) = \begin{cases} (z_{\alpha 1} - r_1)(1 - \alpha_1) & \text{if } (z_{\alpha 1} \leq r_1) \\ (z_{\alpha 1} - r_1)(-\alpha_1) & \text{if } (z_{\alpha 1} > r_1) \end{cases} \tag{11}$$

In Fig. 15, the score  $g(z_{\alpha 1})$  is plotted as a function of the normalized residual  $z_{\alpha 1}$  for different values of the quantile  $q$ . Reducing the  $q$  value, the positive gain becomes larger, increasing the score of the best fitting models (with lowest absolute values of residuals).

The total score of a model  $G_j$  for the R observations is:

$$G_j = \sum_{i=1}^R g_{ij}(z_{\alpha 1}) \tag{12}$$

**Fig. 15** Quantile score  $g(z_{\alpha 1})$  in function of the normalized residuals  $z_{\alpha 1}$



## References

- Abrahamson NA, Silva WJ, Kamai R (2014) Summary of the ASK14 ground motion re-lation for active crustal regions. *Earthq Spectra* 30(3):1025–1055
- Abrahamson NA, Gregor N, Addo K (2016) BC hydro ground motion prediction equations for subduction earthquakes. *Earth Spectra* 32(1):23–44
- Akaike H (1973) Information theory and the maximum likelihood principle in 2nd International Symposium on Information Theory
- Akkar S, Bommer JJ (2010) Empirical equations for the prediction of PGA, PGV, and spectral accelerations in Europe, the Mediterranean region, and the Middle East. *Seismol Res Lett* 81(2):195–206
- Akkar S, Sandikkaya MA, Şenyurt M, Sisi AA, Ay BÖ, Traversa P et al (2014a) Reference database for seismic ground-motion in Europe (RESORCE). *Bull Earthq Eng* 12(1):311–339
- Akkar S, Sandikkaya MA, Bommer JJ (2014b) Empirical ground-motion models for point- and extended-source crustal earthquake scenarios in Europe and the Middle East. *Bull Earthq Eng* 12(1):359–387
- Al Atik L, Youngs RR (2014) Epistemic uncertainty for NGA-West2 models. *Earthq Spectra* 30(3):1301–1318
- Ambraseys NN, Simpson KU, Bommer JJ (1996) Prediction of horizontal response spectra in Europe. *Earthq Eng Struct Dyn* 25(4):371–400
- Ancheta TD, Darragh RB, Stewart JP, Seyhan E, Silva WJ, Chiou BSJ, Wooddell KE, Graves RW, Kottke AR, Boore DM, Kishida T, Donahue JL (2014) NGA-West2 database. *Earthq Spectra* 30(3):989–1005
- Arroyo D, Ordaz M, Rueda R (2014) On the selection of ground-motion prediction equations for probabilistic seismic-hazard analysis. *Bull Seismol Soc Am* 104(4):1860–1875
- Basili R, Brizuela B, Herrero A, Iqbal S, Lorito S, Maesano FE, Murphy S, Perfetti P, Romano F, Scala A, Selva J, Taroni M, Thio HK, Tiberti MM, Tonini R, Volpe M, Glimsdal S, Harbitz CB, Løvholt F, Baptista MA, Carrilho F, Matias LM, Omira R, Babeyko A, Hoechner A, Gurbuz M, Pekcan O, Yalçınar A, Canals M, Lastras G, Agalos A, Papadopoulos G, Triantafyllou I, Bencheikroun S, Agrebi Jaouadi H, Attafi K, Ben Abdallah S, Bouallegue A, Hamdi H, Oueslati F (2019) NEAMTHM18 Documentation: the making of the TSUMAPS-NEAM Tsunami Hazard Model 2018. Istituto Nazionale di Geofisica e Vulcanologia (INGV). <https://doi.org/10.5281/zenodo.3406625>
- Beauval C, Cotton F, Abrahamson N, Theodulidis N, Delavaud E, Rodriguez L, Scherbaum F, Haendel A (2012) Regional differences in subduction ground motions. 15WCEE Lisbon
- 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 Earthq Eng* 9(6):1899–1920

- Bindi D, Massa M, Luzi L, Ameri G, Pacor F, Puglia R, Augliera P (2014) Pan-European ground-motion prediction equations for the average horizontal component of PGA, PGV, and 5 %-damped PSA at spectral periods up to 3.0 s using the RESORCE dataset. *Bull Earthq Eng* 12(1):391–430
- Bindi D (2017) The predictive power of ground-motion prediction equations. *Bull Seismol Soc Am* 107(2):1005–1011
- Bindi et al. (personal communication). Regression coefficients of Bindi et al. (2011) for 5% spectral acceleration (SA) at  $T=2.5, 2.75$  and  $4.0$  s and of Bindi et al (2014) for SA at  $T=4.0$  s.
- Bommer JJ, Douglas J, Strasser FO (2003) Style-of-faulting in ground-motion prediction equations. *Bull Earthq Eng* 1:171–203
- Bommer JJ, Douglas J, Scherbaum F, Cotton F, Bungum H, Fah D (2010) On the selection of ground-motion prediction equations for seismic hazard analysis. *Seismol Res Lett* 81(5):783–793
- Boore DM, Stewart JP, Seyhan E, Atkinson GM (2014) NGA-West2 equations for predicting PGA, PGV, and 5% damped PSA for shallow crustal earthquakes. *Earthq Spectra* 30(3):1057–1085
- Boore DM (2010) Orientation-independent, nongeometric-mean measures of seismic intensity from two horizontal components of motion. *Bull Seismol Soc Am* 100(4):1830–1835
- Bozorgnia Y et al (2014) NGA-West2 research project. *Earthq Spectra* 30(3):973–987. <https://doi.org/10.1193/072113EQS209M>
- Building Seismic Safety Council (BSSC) (2001) 2000 Edition, NEHRP recommended provisions for seismic regulations for new buildings and other structures, FEMA-368, Part 1 (Provisions): developed for the Federal Emergency Management Agency, Washington, DC
- Campbell KW, Bozorgnia Y (2014) NGA-West2 ground motion model for the average horizontal components of PGA, PGV, and 5% damped linear acceleration response spectra. *Earthq Spectra* 30(3):1087–1115
- Cauzzi C, Faccioli E, Vanini M, Bianchini A (2015) Updated predictive equations for broadband (0.01–10 s) horizontal response spectra and peak ground motions, based on a global dataset of digital acceleration records. *Bull Earthq Eng*. <https://doi.org/10.1007/s10518-014-9685-y>
- CEN (2003) EuroCode 8: design of structures for earthquake resistance—part 1: general rules, seismic actions and rules for buildings. European Committee for Standardization, Bruxelles
- Chiarabba C, De Gori P, Speranza F (2008) The southern Tyrrhenian subduction zone: deep geometry, magmatism and Plio-Pleistocene evolution. *Earth Planet Sci Lett* 268(3–4):408–423
- Chiou BSJ, Youngs RR (2014) Update of the Chiou and Youngs NGA model for the average horizontal component of peak ground motion and response spectra. *Earthq Spectra* 30(3):1117–1153
- Chiou B, Darragh R, Gregor N, Silva W (2008) NGA project strong-motion database. *Earthq Spectra* 24(1):23–44
- Cotton F, Scherbaum F, Bommer JJ, 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(2):137–156
- Delavaud E et al (2012) Toward a ground-motion logic tree for probabilistic seismic hazard assessment in Europe. *J Seismol* 16(3):451–473. <https://doi.org/10.1007/s10950-012-9281-z>
- Derras B, Bard P-Y, Cotton F (2014) Towards fully data driven ground-motion prediction models for Europe. *Bull Earthq Eng* 12(1):495–516
- Douglas J, Akkar S, Ameri G, Bard P-Y, Bindi D, Bommer JJ, Bora SS, Cotton F, Derras B, Hermkes M, Kuehn NM, Luzi L, Massa M, Pacor F, Riggelsen C, Sandikkaya MA, Scherbaum F, Stafford PJ, Traversa P (2014) Comparisons among the five ground-motion models developed using RESORCE for the prediction of response spectral accelerations due to earthquakes in Europe and the Middle East. *Bull Earthq Eng* 12(1):341–358
- Douglas J (2015) Ground motion prediction equations 1964–2015. <https://www.gmpe.org.uk>.
- Forster M, Sober E (1994) How to tell when simpler, more unified, or less ad hoc theories will provide more accurate predictions. *Br J Philos Sci* 45(1):1–35
- Garcia D, Wald DJ, Hearne MG (2012) A global earthquake discrimination scheme to optimize ground-motion prediction equation selection. *Bull Seismol Soc Am* 102(1):185–203
- Gneiting T, Raftery AE (2007) Strictly proper scoring rules, prediction, and estimation. *J Am Stat Assoc* 102:359–378
- Idriss IM (2014) An NGA-West2 empirical model for estimating the horizontal spectral values generated by shallow crustal earthquakes. *Earthq Spectra* 30(3):1155–1177. <https://doi.org/10.1193/070613EQS195M>
- Kaklamanos J, Baise LG, Boore DM (2011) Estimating unknown input parameters when implementing the NGA ground-motion prediction equations in engineering practice. *Earthq Spectra* 27(4):1219–1235

- Kale Ö, Akkar S (2013) A new procedure for selecting and ranking ground-motion prediction equations (GMPes): the Euclidean distance-based ranking (EDR) method. *Bull Seismol Soc Am* 103(2A):1069–1084
- Krieger L, Heimann S (2012) MoPaD—moment tensor plotting and decomposition: a tool for graphical and numerical analysis of seismic moment tensors. *Seismol Res Lett Electron Seismol* 83(3):589–595
- Kuehn NM, Scherbaum F (2015) Ground-motion prediction model building: a multilevel approach. *Bull Earthq Eng*. <https://doi.org/10.1007/s10518-015-9732-3>
- Lanzano G, Sgobba S, Luzi L, Puglia R, Pacor F, Felicetta C, D'Amico M, Cotton F, Bindi D (2019) The pan-european engineering strong motion (ESM) flatfile: compilation criteria and data statistics. *Bull Earthq Eng*. <https://doi.org/10.1007/s10518-018-0480-z>
- Lanzano G, Luzi L (2020) A ground motion model for volcanic areas in Italy. *Bull Earthq Eng*. <https://doi.org/10.1007/s10518-019-00735-9>
- Luzi L, Hailemichael 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. <https://doi.org/10.1785/gssrl.79.5.716>
- Luzi L, Puglia R, Russo E, D'Amico M, Felicetta C, Pacor F, Lanzano G, Çeken U, Clinton J, Costa G, Duni L, Farzanegan E, Gueguen P, Ionescu C, Kalogeras I, Özener H, Pesaresi D, Sleeman R, Strollo A, Zare M (2016) The engineering strong-motion database: a platform to access Pan-European accelerometric data. *Seismol Res Lett* 87(4):987–997. <https://doi.org/10.1785/0220150278>
- Maesano FE, Tiberti MM, Basili R (2017) The Calabrian Arc: three-dimensional modelling of the subduction interface. *Sci Rep* 7(1):8887. <https://doi.org/10.1038/s41598-017-09074-8>
- Malagnini L, Akinci A, Herrmann RB, Pino NA, Scognamiglio L (2002) Characteristics of the ground motion in northeastern Italy. *Bull Seismol Soc Am* 92(6):2186–2204
- Malagnini L, Herrmann RB, Di Bona M (2000) Ground-motion scaling in the Apennines (Italy). *Bull Seismol Soc Am* 90(4):1062–1081
- Meletti C, Marzocchi W, Albarello D, D'Amico V, Luzi L, Martinelli F, Pace B, Pignone M, Rovida A, Visini F, the MPS16 Working Group (2017) The 2016 Italian seismic hazard model. In: 16th world conference on earthquake engineering (WCEE), Santiago, Chile, Jan 9–13.
- Montaldo V, Faccioli E, Zonno G, Akinci A, Malagnini L (2005) Treatment of ground-motion predictive relationships for the reference seismic hazard map of Italy. *J Seismol* 9:295–316
- Morasca P, Malagnini L, Akinci A, Spallarossa D, Herrmann RB (2006) Ground-motion scaling in the western Alps. *J Seismol* 10(3):315–333
- NTC (2008) Norme Tecniche per le Costruzioni (2008). Norme Tecniche per le Costruzioni, Decreto Ministeriale 14/01/2008, Gazzetta Ufficiale, N. 29, 4 febbraio 2008
- NTC (2018) Norme Tecniche per le Costruzioni (2018). Norme Tecniche per le Costruzioni, Decreto Ministeriale 17/01/2018, Gazzetta Ufficiale, N. 42, 20 febbraio 2018
- Pacor F, Paolucci R, Ameri G, Massa M, Puglia R (2011a) Italian strong motion records in ITACA: overview and record processing. *Bull Earthq Eng* 9(6):1741–1759
- Pacor F, Paolucci R, Luzi L, Sabetta F, Spinelli A, Gorini A, Nicoletti M, Marcucci S, Filippi L, Dolce M (2011b) Overview of the Italian strong motion database ITACA 1.0. *Bull Earthq Eng* 9(6):1723–1739. <https://doi.org/10.1007/s10518-011-9327-6>
- Pacor F, Felicetta C, Lanzano G, Sgobba S, Puglia R, D'Amico M, Russo E, Baltzopoulos G, Iervolino I (2018) NESS v1.0: a worldwide collection of strong-motion data to investigate near source effects. *Seismol Res Lett* 89(6):2299–2313. <https://doi.org/10.1785/0220180149>
- Pondrelli S, Salimbeni S, Morelli A, Ekström G, Postpischl L, Vannucci G, Boschi E (2011) Euro-pean-Mediterranean regional centroid moment tensor catalog: solutions for 2005–2008. *Phys Earth Planet Int* 185(3):74–81
- Roselli P, Marzocchi W, Faenza L (2016) Toward a new probabilistic framework to score and merge ground-motion prediction equations: the case of the Italian region. *Bull Seismol Soc Am* 106(2):720–733
- Saaty TL (1980) The analytic hierarchy process: planning, priority setting, resource allocation. McGraw-Hill, New York
- Sabetta F, Pugliese A (1996) Estimation of response spectra and simulation of nonstationary earthquake ground motions. *Bull Seismol Soc Am* 86(2):337–352
- Scherbaum F, Cotton F, Smit P (2004) On the use of response spectral-reference data for the selection and ranking of ground-motion models for seismic-hazard analysis in regions of moderate seismicity: the case of rock motion. *Bull Seismol Soc Am* 94(6):2164–2185
- Scherbaum F, Delavaud E, Riggelsen C (2009) Model selection in seismic hazard analysis: an information-theoretic perspective. *Bull Seismol Soc Am* 99(6):3234–3247

- Scognamiglio L, Tinti E, Michelini A (2009) Real-time determination of seismic moment tensor for the Italian region. *Bull Seismol Soc Am* 99(4):2223–2242
- Skarlatoudis AA, Papazachos CB, Margaris BN, Ventouzi C, Kalogeras I, the EGELADOS group (2013) Ground-Motion prediction equations of intermediate-depth earthquakes in the Hellenic Arc, Southern Aegean Subduction Area. *Bull Seismol Soc Am* 103(3):1952–1968. <https://doi.org/10.1785/0120120265>
- Stewart JP, Douglas J, Javanbarg M, Bozorgnia Y, Abrahamson NA, Boore DM, Campbell KW, Delavaud E, Erdik M, Stafford PJ (2015) Selection of ground motion prediction equations for the global earthquake model. *Earthq Spectra* 31(1):19–45
- Stucchi M, Meletti C, Montaldo V, Crowley H, Calvi GM, Boschi E (2011) Seismic hazard assessment (2003–2009) for the Italian building code. *Bull Seismol Soc Am* 101(4):1885–1911. <https://doi.org/10.1785/0120100130>
- Youngs RR, Chiou SJ, Silva WJ, Humphrey JR (1997) Strong ground motion attenuation relationships for subduction zone earthquakes. *Seismol Res Lett* 68(1):58–73
- Wells DL, Coppersmith KJ (1994) New empirical relationships among magnitude, rupture length, rupture width, rupture area, and surface displacement. *Bull Seismol Soc Am* 84:974–1002
- Woessner J, Danciu L, Giardini D, Crowley H, Cotton F, Grünthal G, Valensise G, Arvidsson R, Basili R, Demircioglu MB, Hiemer S, Meletti C, Musson R, Rovida A, Sesetyan K, Stucchi M, the SHARE consortium (2015) The 2013 European seismic hazard model: key components and results. *Bull Earthq Eng* 13:3553–3596
- Zechar JD, Zhuang J (2014) A parimutuel gambling perspective to compare probabilistic seismicity forecasts. *Geophys J Int* 199:60–68. <https://doi.org/10.1093/gji/ggu137>
- Zhao B (2006) Attenuation relations of strong ground motion in Japan using site classification based on predominant period. *Bull Seismol Soc Am* 96(3):898–913. <https://doi.org/10.1785/0120050122>
- Zhao JX, Zhou S, Gao P, Long T, Zhang Y, Thio HK, Lu M, Rhoades DA (2015) An earthquake classification scheme adapted for Japan determined by the goodness of fit for ground-motion prediction equations. *Bull Seismol Soc Am* 105(5):2750–2763
- Zhao JX, Zhou S, Zhou J, Zhao C, Zhang H, Zhang Y, Gao P, Lan X, Rhoades DA, Fukushima Y, Somerville PG, Irikura K (2016) Ground-motion prediction equations for shallow crustal and upper-mantle earthquakes in Japan using site class and simple geometric attenuation functions. *Bull Seismol Soc Am* 106(4):1552–1569
- Zhuang J (2010) Gambling scores for earthquake predictions and forecasts. *Geophys J Int* 181(1):382–390

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.