

The new Italian Seismic Hazard Model (MPS19)

Carlo Meletti^{*,1}, Warner Marzocchi², Vera D'Amico¹, Giovanni Lanzano¹, Lucia Luzi¹, Francesco Martinelli¹, Bruno Pace³, Andrea Rovida¹, Matteo Taroni¹, Francesco Visini¹ & the MPS19 Working Group⁴

⁽¹⁾ Istituto Nazionale di Geofisica e Vulcanologia, Via di Vigna Murata 605, 00143 Rome, Italy

⁽²⁾ University of Naples, Federico II, Dept. Earth, Environmental, and Resources Sciences, Complesso di Monte Sant'Angelo, Via Cupa Nuova Cintia, 21, 80126 Naples, Italy

⁽³⁾ DiSPUTer Department, Università Degli Studi Gabriele D'Annunzio, Chieti-Pescara, Italy

⁽⁴⁾ The complete list of the participants of the MPS19 Working Group is reported in Appendix A in Supplementary Material

Article history: received September 22, 2020; accepted January 20, 2021

Abstract

We describe the main structure and outcomes of the new probabilistic seismic hazard model for Italy, MPS19 [Modello di Pericolosità Sismica, 2019]. Besides to outline the probabilistic framework adopted, the multitude of new data that have been made available after the preparation of the previous MPS04, and the set of earthquake rate and ground motion models used, we give particular emphasis to the main novelties of the modeling and the MPS19 outcomes. Specifically, we (i) introduce a novel approach to estimate and to visualize the epistemic uncertainty over the whole country; (ii) assign weights to each model components (earthquake rate and ground motion models) according to a quantitative testing phase and structured experts' elicitation sessions; (iii) test (retrospectively) the MPS19 outcomes with the horizontal peak ground acceleration observed in the last decades, and the macroseismic intensities of the last centuries; (iv) introduce a pioneering approach to build MPS19_cluster, which accounts for the effect of earthquakes that have been removed by declustering. Finally, to make the interpretation of MPS19 outcomes easier for a wide range of possible stakeholders, we represent the final result also in terms of probability to exceed 0.15 g in 50 years.

Keywords: Seismic hazard; Epistemic uncertainty; Ensemble modeling; Probabilistic methods; Italy.

1. Introduction

In 2015, the Seismic Hazard Center (Centro di Pericolosità Sismica, CPS) of the Istituto Nazionale di Geofisica e Vulcanologia (INGV) was commissioned by the Italian Civil Protection Department to engage and coordinate the national scientific community with the aim of elaborating a new seismic hazard model for Italy (Modello di Pericolosità Sismica, MPS19 hereafter). MPS19 takes into account the large amount of new data and methods that have been published after the last release of MPS [MPS04; <http://zonesismiche.mi.ingv.it>; Stucchi et al., 2011], and it is developed according to the most advanced international standards in seismic hazard [Gerstenberger et al., 2020].

MPS19 will be taken into consideration to update the Italian building code [NTC, 2018], currently based on MPS04 model. For this reason, since the beginning CPS representatives met the national earthquake engineering community to discuss the most proper outcomes of the new seismic hazard model for the building code purpose. Specifically, the model has to consider only declustered seismicity, to cover the whole national territory, the reference soil is rock (i.e. EC8 site category A, $V_{s,30} > 800\text{m/s}$); the hazard has to be expressed in terms of mean horizontal peak ground acceleration (PGA), peak ground velocity (PGV), and peak ground displacement (PGD), spectral acceleration (SA), velocity, displacement and macroseismic intensity (the first release described here contains only PGA and SA results); the oscillatory period of the spectral ordinates has to be in the range between 0.05 and 4 seconds, and the return periods from 30 to 5000 years.

The definition of MPS19 has engaged about one hundred of researchers from Institutes and Universities through an open call to the Italian scientific community (see Appendix A in Supplementary Material). Such a complex endeavor has been completed in 2019, after a successful interaction with a participatory review panel, which gave a final positive approval to MPS19. At the time of writing this paper, applications for building code purposes, or for any other kind of risk analysis are still under discussion. In this paper, we summarize the scientific background of MPS19, the overall structure of the model, the novelties introduced, and the main outcomes. The specific description of the components is (or will be) covered in different papers.

MPS19 is based on a Probabilistic Seismic Hazard Analysis [PSHA, Cornell, 1968], i.e., MPS19 describes in a probabilistic way the forecast of a variety of ground motion intensity measures (IMs; such as PGA and other quantities) on the Italian territory, and it is rooted in open and transparent procedures that guarantee completely reproducible outcomes; specifically, we use the open-source OpenQuake engine platform [Pagani et al., 2014] and all data will be available through a dedicated website. Before getting into the details, there are some features and novelties introduced in MPS19 that deserve to be highlighted, as they mark a difference with previous similar activities.

- MPS19 honors the hazard/risk separation principle [Jordan et al., 2014], i.e., the hazard is evaluated without being conditioned by the effects on the risk. In other words, MPS19 is only focused on providing the forecast of the IM (i.e., the likelihood of observing different levels of IM) in a time window of 50 years, describing at best the historical seismicity and ground motion; purposely, we do not evaluate the impact on the risk, and/or introduce any additional subjective “precautionary” choices that, according to the hazard/risk separation principle, should be introduced only at the risk level.
- MPS19 relies on a clear and coherent probabilistic framework [Marzocchi and Jordan, 2014; 2017; 2018], which allows a univocal distinction of aleatory variability and epistemic uncertainty, a meaningful testing phase (to find possible ontological errors), and a clear interpretation of the mean and percentiles of the hazard outcomes. Then, we use an ensemble modeling strategy to get a complete description of the seismic hazard [Marzocchi et al., 2015].
- MPS19 includes an extensive testing phase to assign weights to each component of the model, to verify the reliability of the earthquake rate models (ERMs hereafter) and of ground motion models (GMMs hereafter), and of the hazard outcomes in terms of PGA and macroseismic intensity at several sites and, on average, over the whole territory.
- Large aftershocks and foreshocks may affect significantly the seismic hazard and risk, but they are usually removed by classical declustering techniques to get a set of time-independent (Poissonian distributed) earthquakes [Cornell, 1968]. For this reason, we also estimate a version of our model introducing a pioneering attempt to correct the classical PSHA model to account for earthquake clustering [Iervolino et al., 2012; 2019; Marzocchi and Taroni, 2014].

The roadmap followed to build MPS19 consists of few basic steps which will be outlined in this paper: i) define a coherent probabilistic framework to fully describe the uncertainties (ensemble modeling), to check the consistency and relative predictive skill of each component of MPS19, and to test the consistency of the final MPS19 outcome (section 2); ii) improve the quality and accuracy of the input data (e.g. seismic catalog, seismotectonic zonation, etc.; section 3); iii) build ERMs based on these new input data, and test their consistency and relative predictive skill (section 4); iv) select the most proper GMMs for Italy, and estimate their relative predictive skill (section 5); v) weigh ERMs and GMMs through a combination of formal experts’ judgment elicitations with the results of the testing phase (section 6); vi) show the outcomes of MPS19 and the results of the testing phase (section 7);

vii) introduce pilot procedures to build a national seismic hazard model which takes into account the effects of aftershocks and foreshocks that have been removed by the declustering technique (MPS19_cluster; section 8). Finally, in section 9, we summarize the main novelties of the model, the open questions, and future developments.

2. The probabilistic framework and the testing phase

PSHA combines ERMs and GMMs to estimate the hazard curve in each site, i.e., the probabilistic function of the IM in a given time interval. To acknowledge the role of pervasive uncertainties and their importance, SSHAC [1997] states that PSHA should “*represent the center, the body, and the range of technical interpretations that the larger technical community would have if they were to conduct the study.*”

Handling uncertainties of different kinds is far from being trivial. For example, in PSHA it is common to separate aleatory variability and epistemic uncertainty even though a satisfactory definition has never been provided [SSHAC, 1997; NRC, 1997]. In MPS19 we adopt a probabilistic framework which makes this separation possible and external to the model [Marzocchi and Jordan, 2014; 2017]. The framework is rooted in a univocal hierarchy of uncertainties which allows a clear interpretation of what a probability is (e.g., frequency or degree of belief), a meaningful testing of the PSHA model, and a clear interpretation of the mean hazard and of the percentiles that have been matter of intense discussion [Musson, 2005; McGuire et al., 2005; Abrahamson and Bommer, 2005]. The details of the framework can be found in Marzocchi and Jordan [2014; 2017]. Here we just summarize the most important feature in the PSHA context: the probability of exceedance (PoE), which characterizes the hazard in one site, is the frequency of an infinite sequence of yearly IM exceedances that are considered exchangeable (i.e., the sequence of exceedances can be reshuffled without losing information); the true frequency is the aleatory variability of the data-generating process and it is unknown (we could know it only if we have the “true” hazard model). We estimate this unknown aleatory variability using different models (or different parametrizations of the same model); the dispersion of these models describes the epistemic uncertainty and allows bounding where the true frequency (aleatory variability) should be. For example, every single model gives an estimation of the PoE for one selected IM, and the PoE of all models is described by a distribution; the center of this distribution is the best guess of the aleatory variability, and the dispersion mimics the epistemic uncertainty (an example of this distribution will be shown in section 6: “Weighing and merging all MPS19 components”); if the observed frequency of future data does not belong to this distribution, it means that we have found an ontological error, i.e., using the Rumsfeld’s words, “*the presence of unknown unknowns*”; this test is what Marzocchi and Jordan [2014] define as model validation (see Figure 4 in Marzocchi and Jordan [2017], for a visual representation of the problem).

This probabilistic framework allows us to use a meaningful testing phase to check the reliability of MPS19. In past common practice, the reliability of a seismic hazard model is based on the consensus of the scientific community on the model’s outcomes [Marzocchi and Zechar, 2011], even though a full consensus is impossible, and the definition of “scientific consensus” is intrinsically fuzzy [Aspinall, 2010]. The testing phase is a more objective approach that may contribute to check the reliability of a model. Owing to the length of the time window considered (50 years), seismologists cannot usually validate models using independent data, even though some efforts in this direction are ongoing, with some limitations, and usually using earthquakes from small to moderate magnitudes [e.g. Mak et al., 2014; Brooks et al, 2017; Mousavi and Beroza, 2018]. Most of the time, as for MPS19, only past data, which have been used more or less directly to build the hazard model, can be considered for testing [e.g. Albarello and D’Amico, 2008; Stirling and Gerstenberger, 2010; Mak and Schorlemmer, 2016]. Hence, instead of using the term “validation”, we prefer to use the less ambitious term “consistency” of the model’s output with data. In other words, the model should be consistent with the data that were available when the model was built. At the same time, the adopted framework outlines a clear pathway to validate the model with future observations.

We test the MPS19 outcomes, ERMs and GMMs to check their consistency with the available data, and their relative forecasting skill. To this purpose, we adopt some tests that have been developed in the Collaboratory for the Study of Earthquake Predictability (CSEP) framework [Jordan, 2006; Zechar et al., 2010a; Schorlemmer et al., 2018]. Specifically, we check the consistency of a model through the *N*- and *S*-test [Schorlemmer et al., 2007; Zechar et al., 2010b], which have been modified to account for the epistemic uncertainty [Marzocchi and Jordan, 2018]. The *N*-test compares the total number of earthquakes forecast with the observed number over the entire region. The *S*-test is similar to a classical likelihood test, but it is applied to a normalized forecast isolating the spatial

component of the forecast. A model that passes both tests explains satisfactorily the overall number of target earthquakes and their location.

The forecasting skill of the models can be quantified through one or more scoring schemes. We argue that the wide proliferation of scoring schemes in scientific literature is due to the fact that a perfect single scoring system does not exist, but each one of them ranks models according to a specific metric, or loss function. Here we use two proper scoring methods [Scherbaum et al., 2009; Zechar and Zhuang, 2014] that explore different aspects of the models: the parimutuel gambling score which tends to reward a model that produces forecasts often close to the observations, and the Log-likelihood score which rewards models that never failed badly (one model that predicts perfectly all the events but one, at which it assigns zero probability, will have the worst score possible).

To sum up, for the testing phase, we:

- test the consistency of every single ERM, comparing its rates with the observed seismicity rates of the last centuries (taken from the CPTI15 catalog [Rovida et al., 2016; 2020]). An ERM that does not pass both consistency tests is not included in MPS19.
- score ERMs and GMMs, comparing all models against each other using different metrics to rank them according to their retrospective forecasting (hereafter hindcasting) skill.
- evaluate the overall consistency of MPS19 with the past accelerometric data that are available in Italy, accounting for the epistemic uncertainty. In other words, we test if the numbers of exceedances recorded in the last decades are in agreement with what expected by MPS19.
- test MPS19 using macroseismic intensity data. Specifically, we test if the number of the macroseismic intensities 6 MCS (Mercalli–Cancani–Sieberg scale [Sieberg, 1932]) and larger observed in the last centuries is consistent with the MPS19 outcomes.

A full description of the tests is beyond the scope of this paper and has been thoroughly discussed in Gneiting and Raftery [2007], Schorlemmer et al. [2007], Zechar et al. [2010b], and Marzocchi and Jordan [2018]. In this paper we just report the main results of the testing phase.

3. The input data

MPS19 takes advantage of a collection of new datasets (see Table 1) that have been used to build and test ERMs and GMMs. They include, among others, an updated historical-instrumental earthquake catalog [CPTI15 v. 1.5, Rovida et al., 2016; 2020], an updated database of seismogenic sources [DISS 3.2.1, DISS Working Group, 2018], a new harmonized GPS velocity model for the Mediterranean area [Devoti et al., 2017], a new flatfile of strong motion recordings for the moment magnitude $M_w \geq 4.0$ earthquakes for the selection and test of the GMMs [Lanzano et al., 2019; see Section 5] and two assessments of the focal mechanism of expected earthquakes [Roselli et al., 2018; Pondrelli et al., 2020]. To be used in MPS19, all datasets cover the whole Italian territory, with the exception of those referred to volcanic areas. Although it was not prescriptive, these datasets have been used by several modelers to build their ERMs.

N.	Dataset	Description/notes	References
1	Earthquake Catalog CPTI15 v 1.5	Parametric catalog for the Italian area and the period 1000-2014, $M_w \geq 4.0$	Rovida et al. [2016; 2020]
2	Earthquake catalog for the Mt Etna volcanic area	Input for CPTI15 and for the earthquake rate model of the Etna area (added to all ERMs)	Azzaro et al. [2017]
3	Instrumental catalog 1981-2015	Input for CPTI15 ($M_w \geq 4.0$) and for some ERMs for low magnitudes	Gasperini et al. [2016]; Lolli et al. [2020]
4	Statistical completeness of CPTI15	Used only in MA5 ERM	Rotondi and Garavaglia [2002], (method)
5	Focal mechanism determination and deformation model	Used in MA4 and MA5 ERM	Pondrelli et al. [2020]

6	GPS velocity model for the Euro-Mediterranean Area	Main input for MG1 ERM	Devoti et al. [2017]
7	Database of Italian seismogenic sources DISS 3.2.1	Main input for MAs and MFs ERMs	DISS Working Group [2018]
8	Strong motion data flatfile	Input for the GMMs selection (see Section 5)	Lanzano et al. [2019]
9	Conversion between ground motion parameters and macroseismic intensity	Used in the MPS19 testing phase (see Section 7)	Gomez Capera et al. [2020]
10	Forecast of focal mechanism of future earthquakes	Used in MSs, MF2, and MG1	Roselli et al. [2018]

Table 1. Datasets, with references, used as a common input for MPS19.

One important dataset is the CPTI15 catalog, which represents the common input to almost all ERMs, but those that use strain rate field derived from geodesy as main input. CPTI15 contains the parameters of more than 4500 earthquakes, from 1000 CE to 2014, determined with homogeneous criteria merging updated macroseismic intensity data of past earthquakes and instrumental determinations of modern ones, as described in Rovida et al. [2020]. Owing to major improvements with respect to previous versions, we argue that the updated catalog turns out to be one of the most impacting factors to explain the differences of MPS19 with previous Italian PSHA models. Among the major improvements, macroseismic magnitudes (which are meant to estimate the moment magnitude) derived only from epicentral intensity are replaced in CPTI15 with magnitudes that are obtained by a new and updated intensity data distribution. The novelties of the catalog lead to a magnitude-frequency distribution that is more realistic than in previous versions of the catalog (i.e., the Gutenberg-Richter law has the b-value close to 1), and to the removal of spurious temporal patterns which characterize the time evolution of the seismicity in previous versions of the catalog. Figure 1 shows how the earthquakes magnitudes are changed between CPTI04 ([Gruppo di Lavoro, 2004] adopted in MPS04) and CPTI15; in particular, a substantial magnitude reduction is evident for the magnitude range between 4.5 and 5.8.

Furthermore, we provide additional pieces of information which have been used by several ERM modelers: i) a model for the maximum magnitude [Visini et al., 2019], ii) a determination of the seismogenic layer and of the hypocentral depths [D’Amico et al., 2019], and iii) the completeness intervals of the CPTI15 catalog and its declustered version [Meletti et al., 2019]. For the purpose of MPS19, we estimate the maximum magnitude in 18 macro-areas with homogenous tectonic features. The maximum magnitude is evaluated by comparing the observed magnitudes in the catalog (spanning approximately 1000 years) with the magnitudes calculated from geological information in the DISS 3.2.1. Within each macro-area, we initially define the maximum observed magnitude as the magnitude of the strongest event in the CPTI15 catalog, considering a minimum level of uncertainty for the magnitude evaluation of 0.2 and 0.3, respectively for the instrumental (post 1980 AD) and historical (pre 1980 AD) earthquakes. This magnitude is then compared to the magnitudes of the composite seismogenic sources of the DISS 3.2.1, and the largest between the two values is taken as representative of the maximum magnitude of each macro-area. Then, according to the procedure adopted in the European Seismic Hazard Model ESHM13 [Woessner et al., 2015], we define the lower threshold for the maximum magnitude as 6.5 for active shallow crustal regions (ASCR), 6.0 for stable continental regions (SCR), and 5.6 for Etna volcanic area. In lack of better qualitative and quantitative data coverage, we define a second and highest value of maximum magnitude, adding a uniform magnitude increment of 0.3 (except for the Etna area where we defined it to be 0), to account for epistemic uncertainty.

The same macro-areas defined for the maximum magnitude are used to assess the depth of the seismogenic layer. Following Boncio et al. [2009], in each macro-area we first calculate the depth distribution of the earthquakes in the instrumental catalog 1981-2015 (dataset 3 in Table 1; Lolli et al. [2020]). Then, for the events within the 5th and 95th percentiles, we evaluate the parameters of both a unimodal and a bimodal distributions, selecting the best

one according to the Akaike Information Criterion, AIC [Akaike, 1974]. Finally, we check the stability of the results using different magnitude thresholds ($M_w \geq 2.0, 2.5, 3.0, 4.0$).

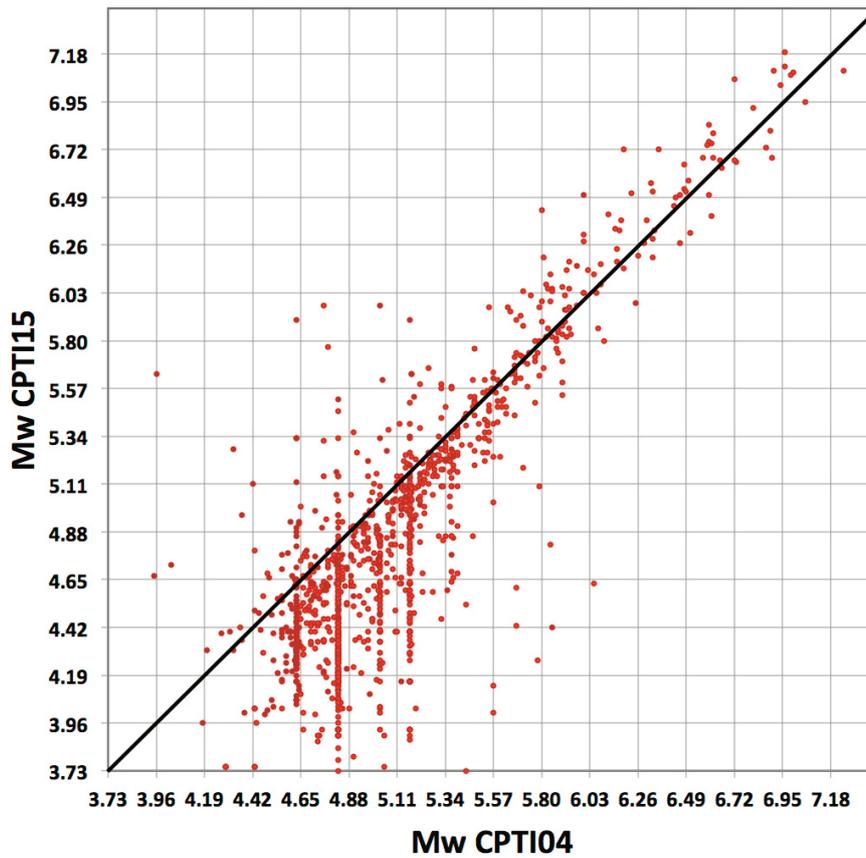


Figure 1. Plot of M_w values in CPTI04 and CPTI15 catalogs.

For seismic hazard analysis purposes, we distributed a version of the catalog in which every earthquake is marked as mainshock or aftershock, with the starting years of the complete periods for its magnitude according to the two approaches described below, and the completeness region it belongs to. Since in PSHA an earthquake catalog is declustered to get the mainshocks distributed in time according to Poisson [Lomnitz, 1966], we use GK74 algorithm [Gardner and Knopoff, 1974] that is the most effective to achieve this goal [van Stiphout et al., 2011; Luen and Stark, 2012]. The use of the general GK74 technique is justified by a recent empirical analysis of worldwide seismic sequences [Stallone and Marzocchi, 2019], which shows that earthquake clustering is similar in different crustal tectonic regions. Independently from the declustering technique used, we underline that its effect on the seismic hazard could be significantly mitigated adding a suitable correction for declustering. We discuss this point in section 8.

As regards the completeness of the catalog we use the same approaches adopted in MPS04 [MPS Working Group, 2004, Stucchi et al., 2011], which are based on the historical and statistical procedures described in Stucchi et al. [2004] and Albarello et al. [2001], respectively. To match the new features of CPTI15 with the previous historical completeness estimates, we i) redefine the M_w bins according to the new empirical conversion relation between epicentral intensity and magnitude used in CPTI15 [Rovida et al., 2016; 2020]; ii) evaluate the completeness intervals for epicentral intensity < 6 , because such intensities were not included in the catalog used for MPS04; iii) revise the geometry of the regions with uniform completeness, and add one more that covers the off-shore areas. As in MPS04, we assess completeness for M_w bins of $0.23 M_w$ units, to avoid the oversampling of some M_w intervals that contain values derived from the conversion of more than one discrete epicentral intensity values [see Stucchi et al., 2011]. As regards the statistical completeness, we apply the statistical analysis of Albarello et al. [2001] in the same regions

defined for the historical completeness with bins of 0.46 M_w units to enhance the stability of the results. In this way we consider in the same class integer and intermediate values of the macroseismic intensity (e.g. epicentral intensity 6-7 together with 6). The results are then applied to bins of 0.23 M_w units, consistently with the historical completeness estimates.

4. The ERMs

The researchers who answered to the open call spontaneously organized themselves in working groups and provided 11 independent ERMs for the whole Italian territory, and one model for the Etna volcano. An ERM forecasts the earthquake rate for different magnitudes, locations, geometry and kinematics. The models were built independently on purpose, to minimize the possibility that single models would be adjusted towards the mean consensus value, which unavoidably leads to underestimate the real uncertainty. The 11 ERMs are grouped according to the main typology of seismogenic sources: 5 area source ERMs (MA1 to MA5) based on the definition of areal sources for which seismicity rates are defined; 2 smoothed, zonation-free ERMs (MS1, MS2) that model the spatial density distributions of earthquakes in seismic catalogs; 2 fault sources plus background ERMs (MF1, MF2), based on the identification of seismogenic sources using tectonic and geophysical data; and 2 geodetically-derived, zonation-free ERMs (MG1, MG2) that model seismicity rates derived from the geodetic strain rate field. The last two models, which are mostly based on ground deformation, provide a different and independent level of information with the respect to the other ERMs because they are almost completely independent from CPTI15.

In table 2 we show the ERMs proposed and their main features. As any part of MPS19, the models have to be reproducible, and so modelers have been encouraged to publish the description of their model in scientific publications. Eventually, modelers provided codes, databases, and the outcomes of their models to make the process fully reproducible. Each ERM describes the aleatory variability of the data-generating (earthquake occurrence) process, and it explores its own epistemic uncertainty, i.e., how the model's outcome varies as a function of the variability of the parameters of the model itself, through a dedicated logic tree. The epistemic uncertainty of the final model includes the uncertainty of each ERM and the dispersion among the ERMs. A brief description of the methodology used to build each ERM and its main references are given in Appendix C of the Supplementary Material.

Model structure	Model code	Input data for earthquakes rate	Epistemic uncertainty explored	Number of branches
Seismotectonic zonation	MA1	CPTI15	Historical/statistical completeness, M_{max} , GR/observed rates	8
	MA2	CPTI15	Historical/statistical completeness, Anchoring/partitioning from Macroarea to source zones	8
	MA3	CPTI15	Historical/statistical completeness, Anchoring/partitioning from Macroarea to source zones	8
	MA4	CPTI15	Historical/statistical completeness, M_{max} , Anchoring/partitioning from Macroarea to source zones	20
	MA5	CPTI15		1
Fault (3D geom) + background (grid)	MF1	Slip rates and composite seismogenic source from DISS geometries	Slip rate (min-mean-max)	3
Fault + background (grid)	MF2	CPTI15, Slip rates and individual seismogenic source from DISS geometries	Historical/statistical completeness, Percentage of smoothed versus faults	6

Smoothed seismicity (grid)	MS1	CPTI15 + INGV seismic bulletin	Historical/statistical completeness, Percentage of fixed versus adaptative ratio	6
	MS2	CPTI15	Historical/statistical completeness, truncated/no-truncated GR, Isotropic/anisotropic	8
Geodetics (grid)	MG1	GPS	Modeling strain rate to seismic moment, seismogenic thickness	14
	MG2	GPS and SHmax	Smoothing factor, coupling coefficient, coupled thickness	12

Table 2. Main features of the ERMs; kind of seismogenic sources, input data, uncertainties explored, number of branches of an internal logic tree.

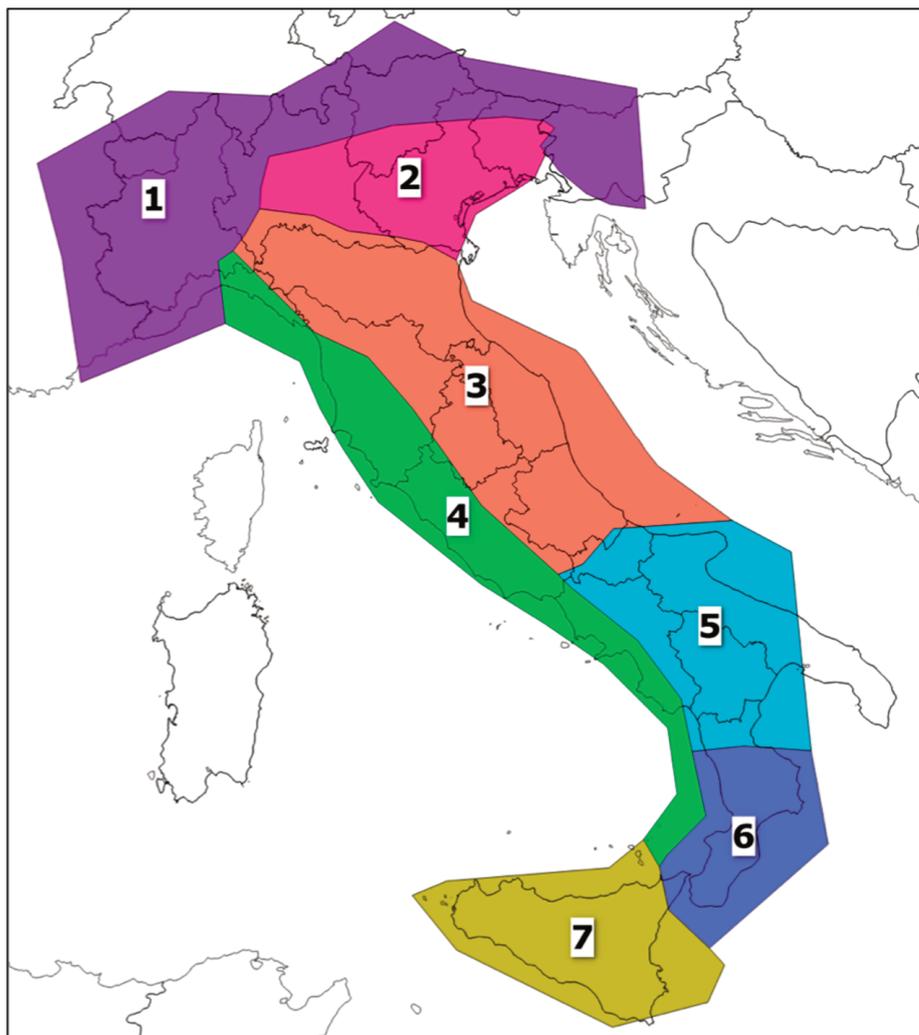


Figure 2. Macro-areas in which the testing phase is carried out. See text for more details.

Each ERM is deemed appropriate to be included in MPS19 if it i) passes consistency tests described in Section 2, and ii) is considered physically plausible from a group of independent experts through a formal experts' elicitation, which is described in Section 6. At first, we carry out statistical tests on a national scale, to define if an ERM is appropriate to be included in MPS19, and to measure the relative hindcasting skill which is used to weigh each single ERM (see Section 6). Following a suggestion of the participatory review panel (see Appendixes A and B in the Supplementary Material) we carry out the consistency tests of ERMs also on regions identified by a group of experts according to the following criteria: separate regions with high and low number of geological data, earthquakes, and geodetic stations; minimize the overlapping of areas with different completeness; keep the areas as large as possible avoiding marked tectonics inhomogeneity; minimize the off-shore part included in the regions. Following these criteria, we define 7 macro-areas (Figure 2).

Testing ERMs in each region allows us to check if their hindcasting skill varies with space and/or with the number of input data. If an ERM does not pass both consistency tests in one region, its overall earthquake rate is replaced in that subregion by the weighted average of the other consistent ERMs, preserving the geometries, depth, and focal mechanisms of the model. All models of Table 2 passed both consistency tests at national scale, hence they are deemed to be included in MPS19. However, as shown in Table 3, some of them failed to describe satisfactorily the data in some macro-areas of Figure 2.

Model	Macro-area
MA1	None
MA2	7
MA3	7
MA4	None
MA5	7
MF1	1,6
MF2	None
MS1	None
MS2	4
MG1	4
MG2	3,4,6

Table 3. Results of tests conducted on ERMs on 7 macro-areas. For each model, in the second column we list the number of macro-areas (shown in Figure 2) in which it does not pass the test; “None” means that the model passes the test in all macro-areas.

ERMs in Table 2 are defined for the entire Italian territory (excluding Etna area), and do not include seismogenic sources external to the area covered by CPTI15. For this reason, we integrate each ERM with: a) an ERM for the Etna area [adapted from Azzaro et al., 2017 and Peruzza et al., 2017]; b) an ERM for deep seismicity related to the slab activity in the southern Tyrrhenian Sea [adapted from Maesano et al., 2017] and; c) an ERM which models seismogenic sources located out of the area of the CPTI15 catalog [adapted from Woessner et al., 2015]. A brief description of the methodology used to build these 3 additional ERMs is given in Appendix D of the Supplementary Material.

Figure 3 maps the mean rates of occurrence of earthquakes with M_w larger or equal than 4.5 of the 11 national ERMs considered in MPS19 and Figure 4 shows the epistemic uncertainty of the rates. The mean rates of the 11 ERMs lie in a range from ~4 eq/yr to ~8eq/yr. Lowest and highest rates are ~3eq/yr and ~17eq/yr, respectively from a fault-based and an area-source ERM. Considering the overall weighed distribution of the rates, 2.5th - 97.5th and 16th - 84th percentiles are, respectively, ~4.4 eq/yr - ~8.6 eq/yr, and ~4.8 eq/yr - ~6.9 eq/yr.

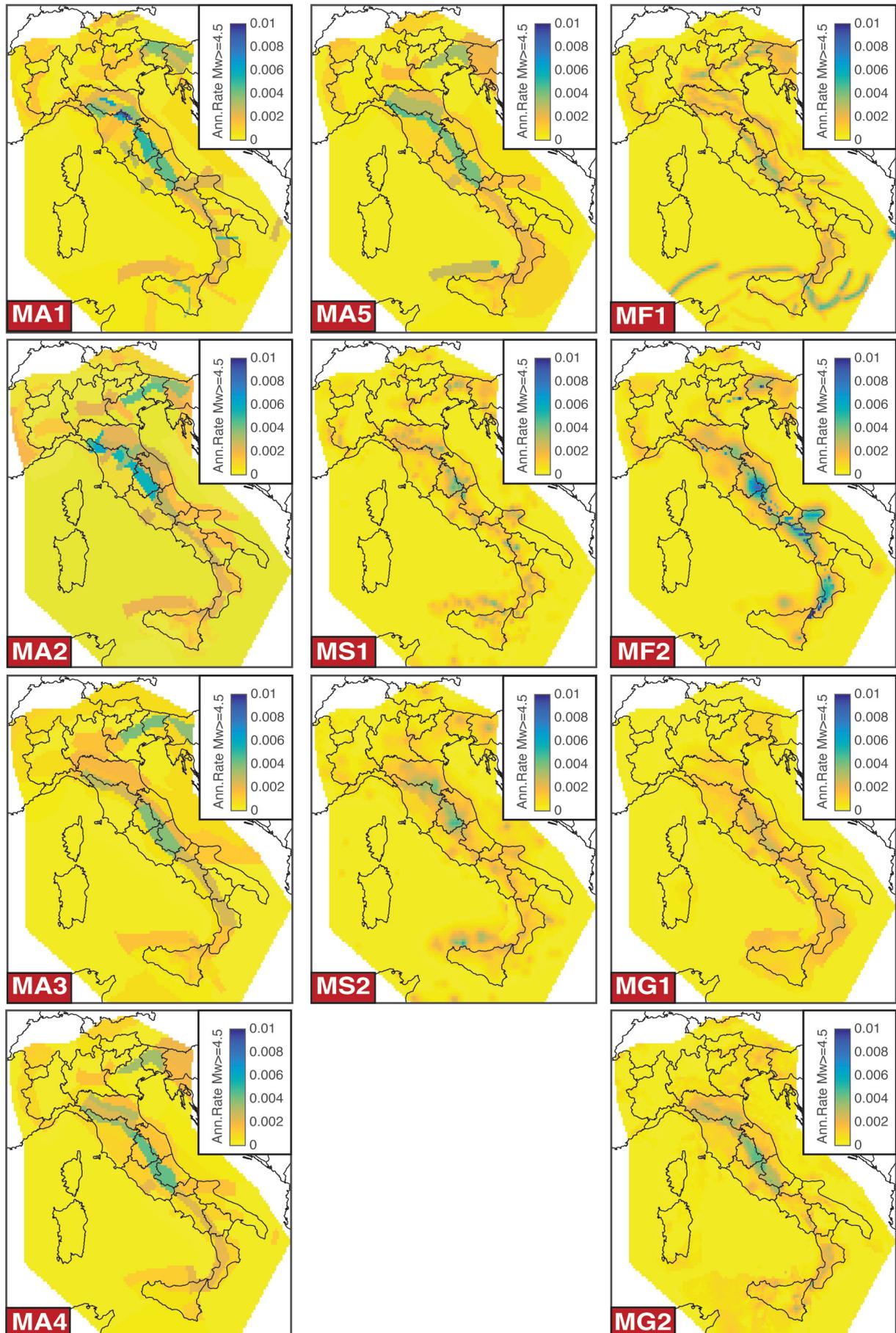


Figure 3. Mean annual earthquake rates for $M_w \geq 4.5$ of each ERM.

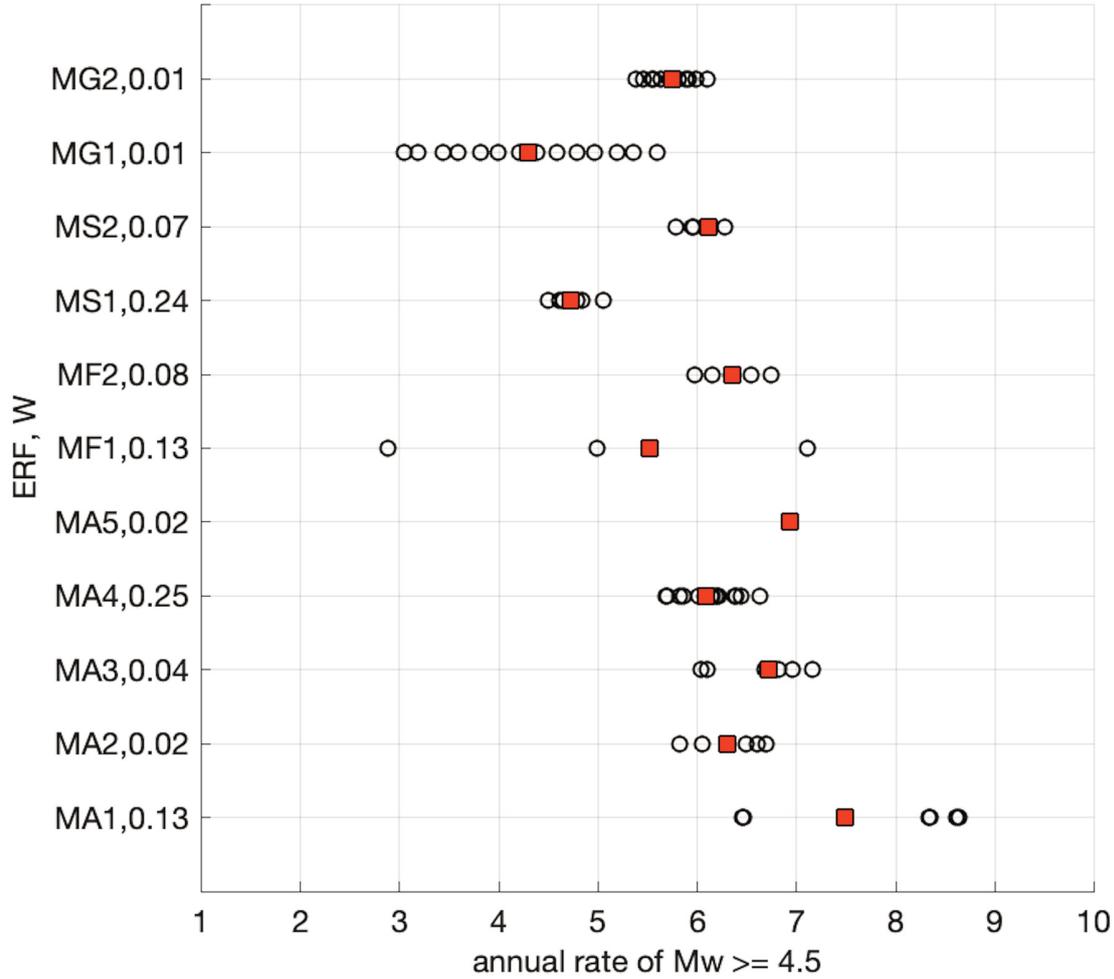


Figure 4. Distribution of rates of occurrence for earthquakes with M_w larger or equal than 4.5, as evaluated by the 11 ERMs and their sub-models (open circle). Red squares are the mean rate of each ERM. The y axis reports the ERMs with their own code and final weight (see Section 6).

5. The GMMs

This section summarizes the activity of the selection and weight of GMMs for MPS19, that has been exhaustively treated in Lanzano et al. [2020]. The candidate GMMs potentially suitable for Italy have to satisfy the following general conditions: i) representation of the characteristics of ground motion for whole Italian territory; ii) calibration for the engineering bedrock (i.e. EC8 site category A, $V_{s,30} > 800$ m/s) and flat topography conditions; iii) prediction of the horizontal PGA and 12 ordinates of the acceleration response spectra (5% damping), at periods T 0.05, 0.1, 0.15, 0.2, 0.3, 0.4, 0.5, 0.75, 1, 2, 3, 4 seconds; iv) horizontal motion in terms of geometrical mean of the horizontal components.

We identify three main tectonic regions, which are characterized by peculiar ground motion attenuation properties: i) Active Shallow Crustal Regions (ASCRs); ii) Volcanic Zones (VZs); and iii) Subduction Zone (SZ), in the southern Tyrrhenian sea. A preliminary selection of candidate GMMs is performed over nearly one thousand models published in scientific literature, which are listed in Douglas [2019]. In order to reduce the number of available models, we adopt the rejecting criteria by Cotton et al. [2006] and Bommer et al. [2010], and then add further project-specific criteria to fulfil the requirements of MPS19 listed above [see Lanzano et al., 2020 for details]. The candidate models for ASCRs after this pre-selection are shown in Table 4, and Figure 5 displays the trellis plots of their spectral acceleration predictions, assuming rock-site conditions (EC8 site category A or $V_{s,30} > 800$ m/s) and normal faulting.

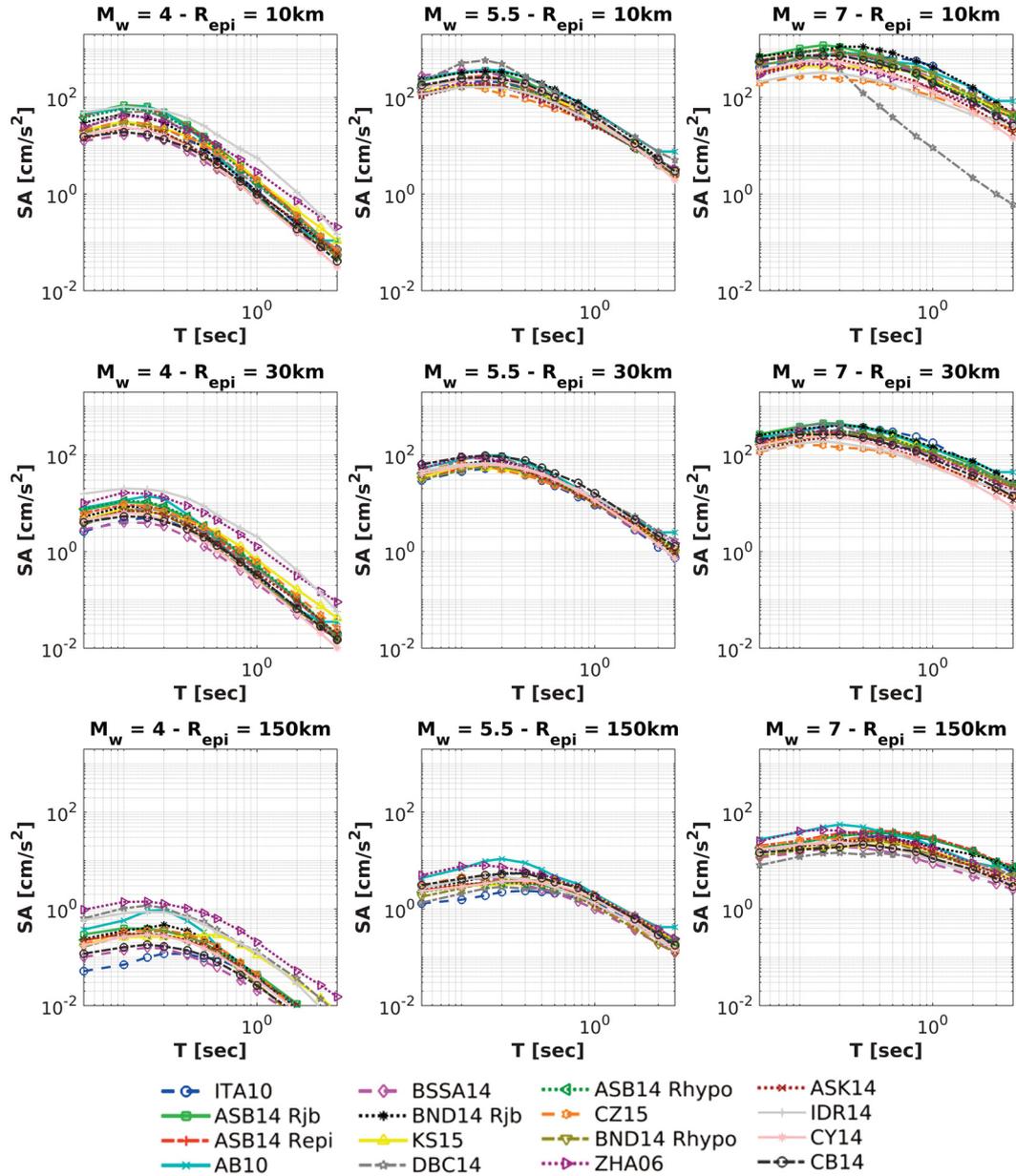


Figure 5. Trellis plot of the median values of the GMMs pre-selected for ASCRs (see Table 4 for the abbreviations).

Name	Code	Area	H comp	M range	R type	R range (km)	Period range (s)	Site-effect term	Style of faulting
Zhao et al. [2006]	ZHA06	Japan	GM	5.0 - 8.3	R_{rup}	0 - 300	0.05 - 5.0	5 Classes	N, R, SS
Bindi et al. [2011] ¹	ITA10	Italy	GM	4.1 - 6.9	R_{JB}	0 - 200	0.04 - 2.0	5 EC8 ² Classes	N, R, SS, U
Akkar and Bommer [2010] ³	AB10	Europe and Middle- East	GM	5.0 - 7.6	R_{JB}	0 - 100	0.05 - 3.0	3 Classes	N, R, SS

¹ The GMMs coefficients up to 4 s by ITA10 and BND14 have been provided by the Authors, even if they are not published.

² 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, $V_{S,30}$.

Abrahamson et al. [2014]	ASK14	Worldwide	RotD50	3.0 - 7.9	R_{rup}	0 - 300	0.01 - 10	$V_{S,30}$	N, R, SS
Boore et al. [2014]	BSSA14	Worldwide	RotD50	3.0 - 7.9	R_{JB}	0 - 400	0.01 - 10	$V_{S,30}$	N, R, SS, U
Campbell & Bozorgnia [2014]	CB14	Worldwide	RotD50	3.0 - 7.9	R_{rup}	0 - 300	0.01 - 10	$V_{S,30}$	N, R, SS
Chiou & Youngs [2014]	CY14	Worldwide	RotD50	3.1 - 7.9	R_{rup}	0 - 400	0.01 - 10	$V_{S,30}$	N, R, SS
Idriss [2014]	ID14	Worldwide	RotD50	4.5 - 7.9	R_{rup}	0 - 175	0.01 - 10	$V_{S,30}$	N, R, SS
Akkar et al. [2014]	ASB14	Europe and Middle -East	GM	4.7 - 7.6	R_{JB} R_{epi} R_{hyp}	0 - 200	0.01 - 4.0	$V_{S,30}$	SS, R, N
Bindi et al. [2014] ¹	BND14	Europe and Middle -East	GM	4.0 - 7.6	R_{JB} R_{hyp}	0 - 300	0.02 - 3.0	⁴ Classes, $V_{S,30}$	SS, R, N, U
Derras et al. [2014]	DBC14	Europe and Middle -East	GM	3.6 - 7.6	R_{JB}	0 - 550	0.01 - 4.0	$V_{S,30}$	N, R, SS
Cauzzi et al. [2015] ⁴	CZ15	Worldwide	GM	4.5 - 7.9	R_{rup}	0 - 150	0.02 - 10	$V_{S,30}$	SS, N, R
Kuehn & Scherbaum [2015]	KS15	Europe and Middle -East	GM	4.0 - 7.6	R_{JB}	0 - 200	0.01 - 4.0	$V_{S,30}$	SS, N, R

Table 4. List of the candidate GMMs for the application to ASCRs for PGA⁵ and acceleration response spectra ordinates. GM: geometrical mean; RotD50: median of orientation-independent ground motion amplitudes [Boore, 2010]. R_{epi} : epicentral distance; R_{JB} : Joyner-Boore distance; R_{hyp} : hypocentral distance; R_{rup} : distance from the rupture plane. N: normal fault; R: reverse fault; SS: strike-slip fault; U: undefined style of faulting.

We evaluate the relative hindcasting skill of the pre-selected GMMs through the comparison with accelerometric records from the strong-motion databases available for Italy [ITACA, Italian ACcelerometric Archive, <http://itaca.mi.ingv.it>; Luzi et al., 2008; Pacor et al., 2011] and Europe [ESM, Engineering Strong Motion Database, <https://esm.mi.ingv.it>; Luzi et al., 2016]. The testing datasets for ASCRs and SZ are provided and described in Lanzano et al. [2020]. We verify the stability of the hindcasting skill of candidate GMMs, and examine their behavior on the basis of the requirements of MPS19, using 4 different subsets: i) the whole dataset (named All); ii) a subset of records with stations classified as EC8 site class A (named EC8-A); iii) a subset of events with $M_w \geq 5$ and stations located at $R_{JB} < 50$ km (named MR), since these magnitude and distance ranges are those mostly affecting the hazard in Italy; iv) a validation subset without the Italian records preceding 2009 (named After 2009), to obtain a dataset independent of that used to calibrate the ITA10 [Bindi et al., 2011] model. The last subset is the only one which allows a truly independent comparison of the GMMs; this is essential to quantify possible overfitting effects, which could favor those GMMs built using, at a different extent, the data from ITACA.

For the purpose of measuring the relative GMM hindcasting skill, we consider the proper scoring methods described in Section 2, the Log-likelihood (LLH), proposed by Scherbaum et al. [2009], and the parimutuel gambling score [Zechar and Zhuang, 2014]; we also consider an additional method, i.e. the quantile score [Gneiting and Raftery, 2007], to check the consistency of the results. The details of the scoring procedure and the scoring values for each GMM are reported and discussed in Lanzano et al. [2020]; here in Table 5 we show the average values of the scores, and the un-normalized weight for each GMM using the EC8-A subset; the un-normalized weight is the ordinal number of the score. Summing up the total un-normalized scores of EC8-A and MR subsets, the grand-total

³ The AB10 predictions at 4 s (not provided by the Authors) are linearly extrapolated from those at 2 and 3 s.

⁴ CZ15 is provided in terms of ordinates of the displacement response spectra (SD), and the predictions have been converted to SA by multiplying for $(2\pi/T)^2$.

⁵ Most of the GMMs are also calibrated for PGV, with the exception of the model IDR14 and ZHA06.

score is obtained, providing a final rank of the candidate GMMs. The 3 best performing GMMs are: #1) Bindi et al. [2011], ITA10 (93 points); #2) Bindi et al. [2014], BND14 R_{hyp} (82 points); #3) Cauzzi et al. [2015], CZ15 (75 points). Worthy of note, these GMMs, besides showing the best relative hindcasting skill, use different metrics for the distance (i.e., R_B , R_{hyp} , R_{rup} , respectively) and are calibrated on different datasets (i.e., Italian, European, and global, respectively).

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

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

Table 5. Rank of candidate GMMs for the ASCR dataset EC8-A. UNW: un-normalized weight.

For the Italian subduction zone, which corresponds to the Calabrian Arc in the southern Tyrrhenian sea, we select two GMMs after the pre-selection and scoring procedures: i) Skarlatoudis et al. [2013], HEL13, and ii) Abrahamson et al. [2016], HYD15 [for details, see Lanzano et al., 2020]. For volcanic zones (i.e., Etna, Aeolian islands, and Campanian volcanoes) we use a single GMM that is the model calibrated on purpose by Lanzano and Luzi [2019], since no GMM available in the previous literature [e.g. Faccioli et al. 2010; Tusa and Langer, 2016] is able to describe satisfactorily the ground motion amplitudes observed recently in Italy (2018-2019 Mount Etna sequence, mainshock M_w 4.9, and 2017 Ischia event, M_w 3.9), especially in near-source conditions.

6. Weighing and merging all MPS19 components

It is essential to weigh each ERM and GMM to properly merge them in the final MPS19. The weighing procedure is often based on unstructured experts' judgment [Gerstenberger et al., 2020]. Here we aim at providing a more objective procedure that could make the weighing scheme more sound, transparent and reproducible.

The final weight of each component is assigned accounting for its statistical scoring (described in the sections 4 and 5), and the outcomes of an experts' elicitation session with independent national and international experts. As a matter of fact, the weight of each model (either ERM or GMM) is a combination of two different evaluations:

i) a weight based on the testing phase of the model, W_1 , and ii) a weight derived by expert judgment, W_2 . The final weight of that model is

$$W = W_1 \times W_2 / C \tag{1}$$

where C is a normalization factor that sets W in the range (0,1) and that the sum of all weights is 1; specifically, C is the sum of $W_1 \times W_2$ for all models. The need of merging these two different scores is due to the fact that the number of data for the testing phase is limited and the test is made retrospectively (so models that tend to overfit the data are more rewarded than others in the testing phase), and that the earthquake catalog could not cover the whole seismic cycle; in this case, the experts' opinion is expected to provide an independent additional information for weighing models [Delavaud et al., 2012]. Note that equation (1) provides zero weight to models that are either not able to explain the data, or unrealistic, e.g. based on assumptions that are blatantly wrong.

In Tables 6 and 7 we show the weights of the ERMs and GMMs, which take into account the statistical scoring and the experts' judgment. In Appendix E of the Supplementary Material we report the description of the experts' elicitation sessions.

Model	Testing phase weight W_1	Experts' elicitation weight W_2	Final weight W
MA1	0.1439	0.0955	0.13
MA2	0.0303	0.0539	0.02
MA3	0.0833	0.0539	0.04
MA4	0.1136	0.2240	0.25
MA5	0.0833	0.0198	0.02
MF1	0.1136	0.1172	0.13
MF2	0.1136	0.0756	0.08
MS1	0.1136	0.2193	0.24
MS2	0.1439	0.0501	0.07
MG1	0.0303	0.0529	0.01
MG2	0.0303	0.0378	0.01

Table 6. ERMs weights.

GMMs	W_1	W_2	W
ASCRs			
ITA10 [Bindi et al., 2011]	0.37	0.423	0.45
BND14 Rhyp [Bindi et al., 2014]	0.33	0.328	0.32
CZ15 [Cauzzi et al., 2015]	0.3	0.259	0.23
SZ			
HEL13 [Skarlatoudis et al., 2013]	0.53	0.5	0.53
HYD15 [Abrahamson et al., 2016]	0.47	0.5	0.47
VZs (depth \leq 5 km)			
LL19 [Lanzano & Luzi, 2019]	-	1.0	1.0

Table 7. GMMs weights.

To summarize, the weighing procedure adopted is as much objective and auditable as possible; it considers independent information, such as the hindcasting performance of the models and the experts' judgment on the scientific reliability of the models. Nonetheless, the procedure has also some limitations, which are mostly due to the finite number of data, and the impossibility to carry out more specific tests at regional scale, or tests that depend on the earthquake magnitude. Here we report a list of these limitations.

- i. The statistical procedure to select and to score ERMs and partially GMMs does not take into account that some models were calibrated with the same data used for testing, while others do not. This means that the latter are probably underscored (and underweighted).
- ii. The weight of each ERM and GMM is constant for the whole country, whereas some ERMs are not used in some sub-regions where they failed the tests.
- iii. Owing to the limited number of data, the testing procedure is limited only to some specific aspects of the earthquake occurrence and ground motion processes. For example, the data available do not allow us to test ERMs for M7+ earthquakes, or to compare GMMs performance in the very near field (less than 10 km) which are very relevant for seismic hazard.
- iv. The procedure carries on an unavoidable subjectivity in merging the different scoring metrics of the testing phase and the scoring of the experts' judgment. Other procedures to merge these scores are possible.
- v. The weights based on the testing phase do not take into account that some ERMs may perform poorly because of uncertainties in the oldest earthquakes of the seismic catalog, such as large uncertain location of off-shore and/or very old earthquakes, and the magnitude of historical events which may be affected by cumulative damages caused by seismic sequences. This is unavoidable, but it highlights again the need to not weigh a model only using its hindcasting performance.

Once the ERMs and GMMs weights have been defined, the seismic hazard at each site is estimated by \mathcal{N} hazard curves, where \mathcal{N} is the number of ERMs and GMMs combinations. Worthy of mention, this set of hazard curves is expected to sample the epistemic uncertainty, and a higher number of models cannot reduce it. In fact, even though models are developed independently from one another, almost all of them use the same amount of information [Marzocchi and Jordan, 2014].

If we dissect the bunch of curves for one specific IM value, we get a set of estimations of the PoE for that specific IM value (Figure 6). We model them using a Beta distribution [Marzocchi et al., 2015], which describes where the true PoE is expected to be [Marzocchi and Jordan, 2014; 2017]. The choice of using the Beta distribution is subjective and other ones may be used [see Gelman et al., 2003], keeping the same probabilistic framework adopted here. That said, we argue that this choice is certainly less subjective than describing the hazard using the mean alone, which would imply the use of a Dirac distribution, or using the percentiles of the distribution, which implicitly means to impose a nonparametric distribution. The Beta distribution usually fits well aleatory variables with a unimodal distribution and bounded between 0 and 1; in case the data show more modes [for examples 2 modes as in Figure 6 of Marzocchi et al., 2015], we can still use the Beta distribution if we think the bimodality is only due to the fact that the models used are just exploring two extreme scenarios, and a proper epistemic uncertainty has to fill the gap between the modes. To sum up, no matter what distribution we decide to use, this choice is certainly less critical than using one single value for the hazard.

7. The MPS19 outcomes and consistency test

Regional seismic hazard is usually presented as a map showing the PGA values that have 10% PoE in 50 years. This representation is of particular interest to the engineering community which establishes this hazard level to be of interest for the building code. However, this representation is often misinterpreted by laymen and some end-users, which assume that this map is reporting the maximum PGA that can occur in each area. In fact, although it is quite obvious to observe PGA exceedances just close to the epicenter of a large earthquake, this observation is often deemed as a proof that the probabilistic model is wrong. For this reason, we propose here an additional graphical representation of MPS19 that may be easier interpreted by the non-engineering community. Specifically, we show the PoE of PGA = 0.15 g in the whole Italian territory for a time interval of 50 years (Figure 7a). The choice

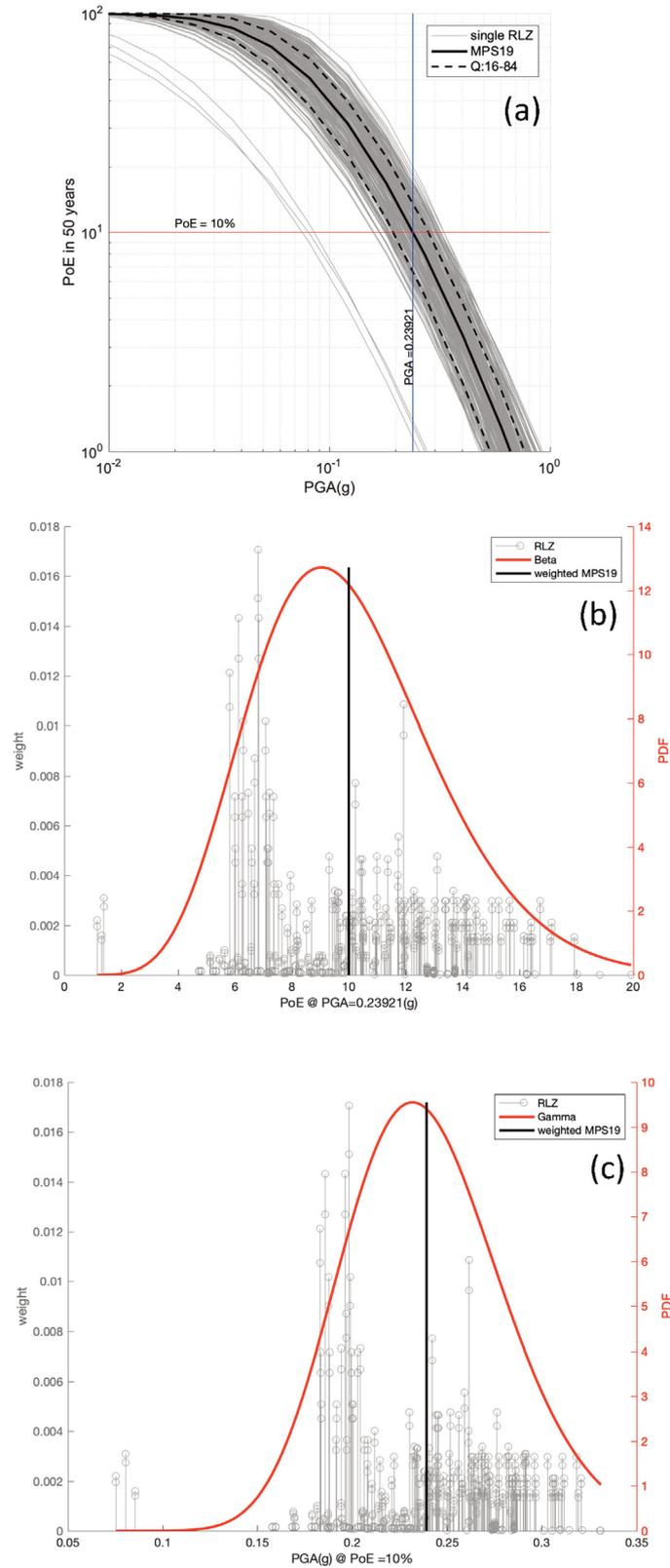


Figure 6. Seismic hazard for L'Aquila. (a) Hazard curves of all the realizations of MPS19 and weighted mean hazard curve along with 16th and 84th percentiles; the horizontal red line marks the 10% PoE in 50 years, and the blue vertical line identifies the PGA corresponding to the mean hazard at 10% PoE in 50 years (PGA*). (b) The Beta distribution (red line) of the PoE corresponding to PGA*; each vertical line shows the weight of the outcome of one hazard model. (c) The Gamma distribution (red line) of the PGA corresponding to PoE of 10% in 50 years.

of 0.15 g has been adopted as a criterion for significant shaking, and other values can be used [WGCEP, 1995]. Figures 7b and 7c show the more classical representation of MPS19, in terms of PGA with 10% and 2% PoE in 50 years.

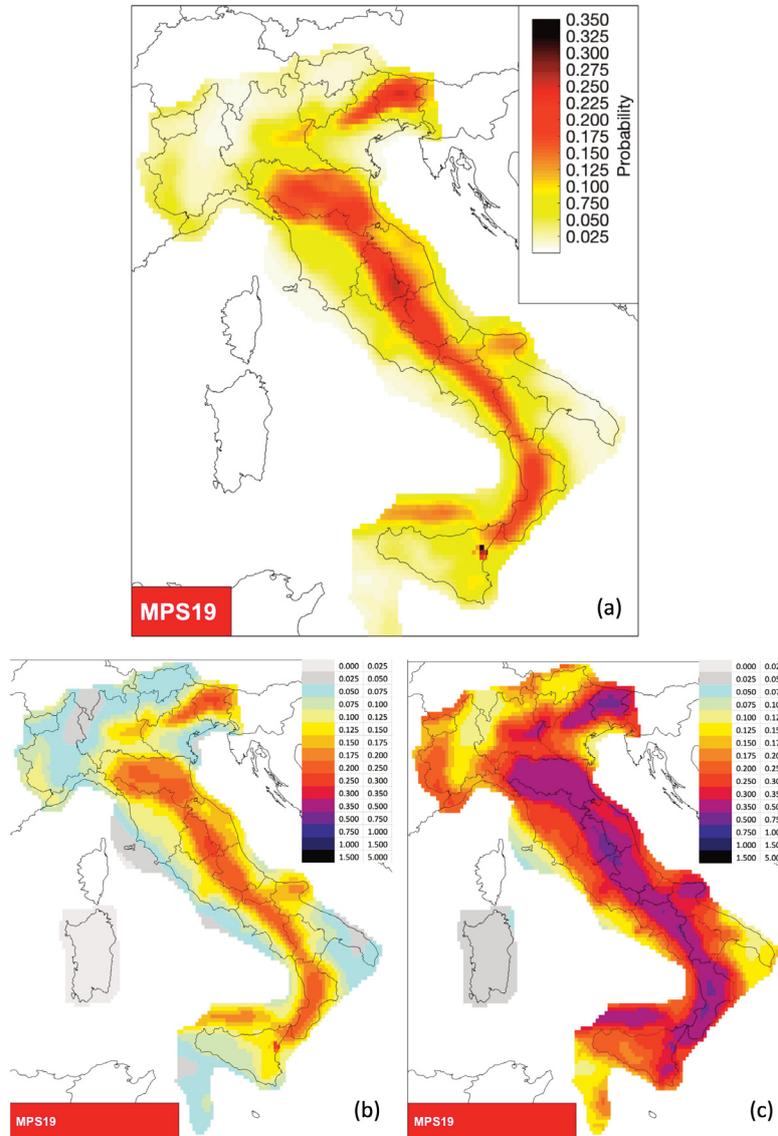


Figure 7. The outcomes of MPS19: (a) PoE for PGA=0.15 g in 50 years; (b) PGA with 10% probability to be exceeded in 50 years; (c) PGA with 2% probability to be exceeded in 50 years.

One of the most innovative aspects of MPS19 concerns the quantification of the epistemic uncertainties of the final hazard estimates. As mentioned above, for the probabilistic scheme used, the exceedance probability at one site for one specific IM is not a single number, but it is a Beta distribution, the dispersion of which describes the epistemic uncertainty on the true exceedance frequency (random variability). One way to represent epistemic uncertainties on the whole national territory is given by the mapping of the CoV, i.e. coefficient of variation. CoV is given by

$$\text{CoV} = \frac{\sigma}{\mu} \quad (2)$$

where σ is the weighted standard deviation of all the hazard estimates, and μ is the weighted average value among all estimates (these parameters are used to calculate the parameters of the Beta distribution). A CoV of 0.3 indicates that the standard deviation between all the models used (including the uncertainties within each model) is 30% of the average. Figure 8 shows the CoV map for the whole national territory relating to PGA values which have the exceedance probability of 10% in 50 years (Figure 7b). From the figure it can be seen that the highest values of the CoV are for areas with low seismic hazard. When we consider only areas with high hazard (i.e. those with a higher μ , panel (c)), the CoV mainly varies between 0.15 and 0.30. The fact that the CoV is higher for low hazard areas is due to the fact that in those areas the information available are relatively scarce, and so it is understandable to have greater uncertainties.

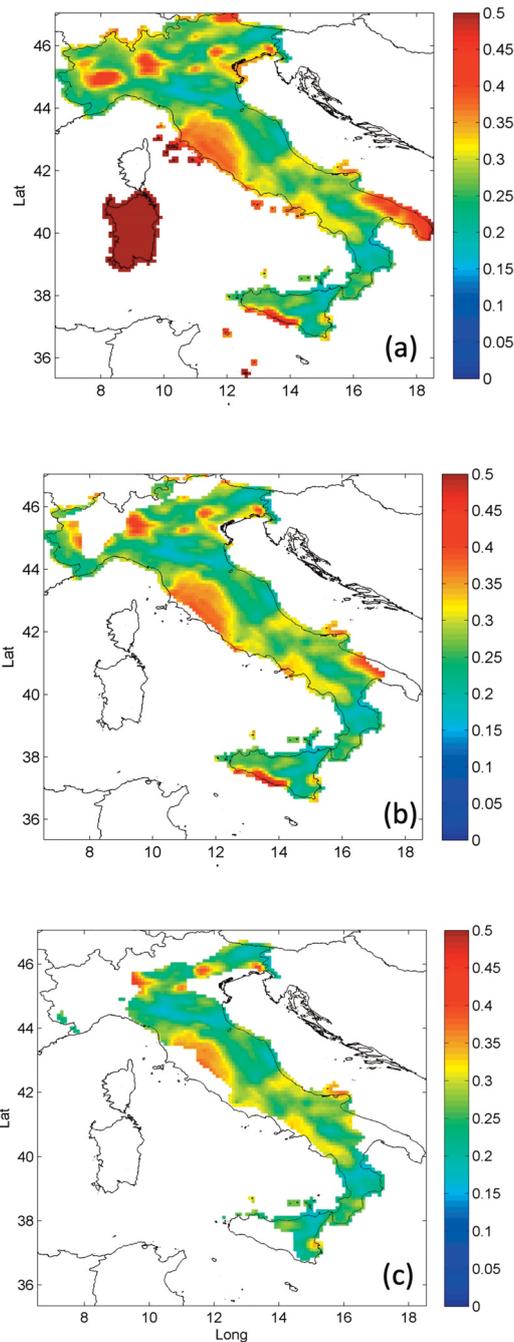


Figure 8. CoV (equation 2) map for the 10% PoE in 50 years PGA (Figure 7b). (a) CoV for the whole territory; (b) CoV for the sites of Figure 7b with PGA above the 25th percentile; (c) CoV for the sites of Figure 7b with PGA above the 50th percentile.

It is worth remarking that this result poses a great challenge to the use of MPS19 (as well as of any other model built on the same amount of information) for the definition of the building code. As a matter of fact, the current Italian building code is based on one hazard input having 4 decimal digits and it does not consider epistemic uncertainty; *de facto*, the latter is handled by engineers introducing some arbitrary conservative parameters. Figure 8 shows that this request of precision does not match the current epistemic uncertainty in PSHA. In practice this means that even in the most optimistic (and likely unrealistic) option where MPS19 is “right”, the mean values of future PSHA models considering more data are expected to vary according to the range given by the CoV. In essence the CoV represents the lower bound of the expected future variations of the mean hazard.

As regards the test of MPS19 against observations, in Figure 9 we show the results of the consistency test of MPS19 compared to the PGA recorded ed by different accelerometric stations in recent decades (from the ITACA database). In particular, we consider accelerometric stations that have been operating for at least 30 years. If the period is longer, the last 30 years are considered (until 31 December 2018). In this way all stations have observations for the same length of the time interval considered. The fact that the time intervals are temporally staggered is irrelevant because of the stationarity of the Poisson process. Owing to the PGA values of MPS19 refer to a 50-yr time window, we rescale the PoE to 30-yr time window. This is done maintaining the same annual exceedance rate of PGA; the 10% threshold of PoE in 50 years (which becomes about 6% in 30 years) is chosen for the obvious practical interest in terms of building code; the 30% threshold in 50 years (which becomes about 19% in 30 years) is used to lower the expected PGA values in order to have more exceedances and therefore to empower the test.

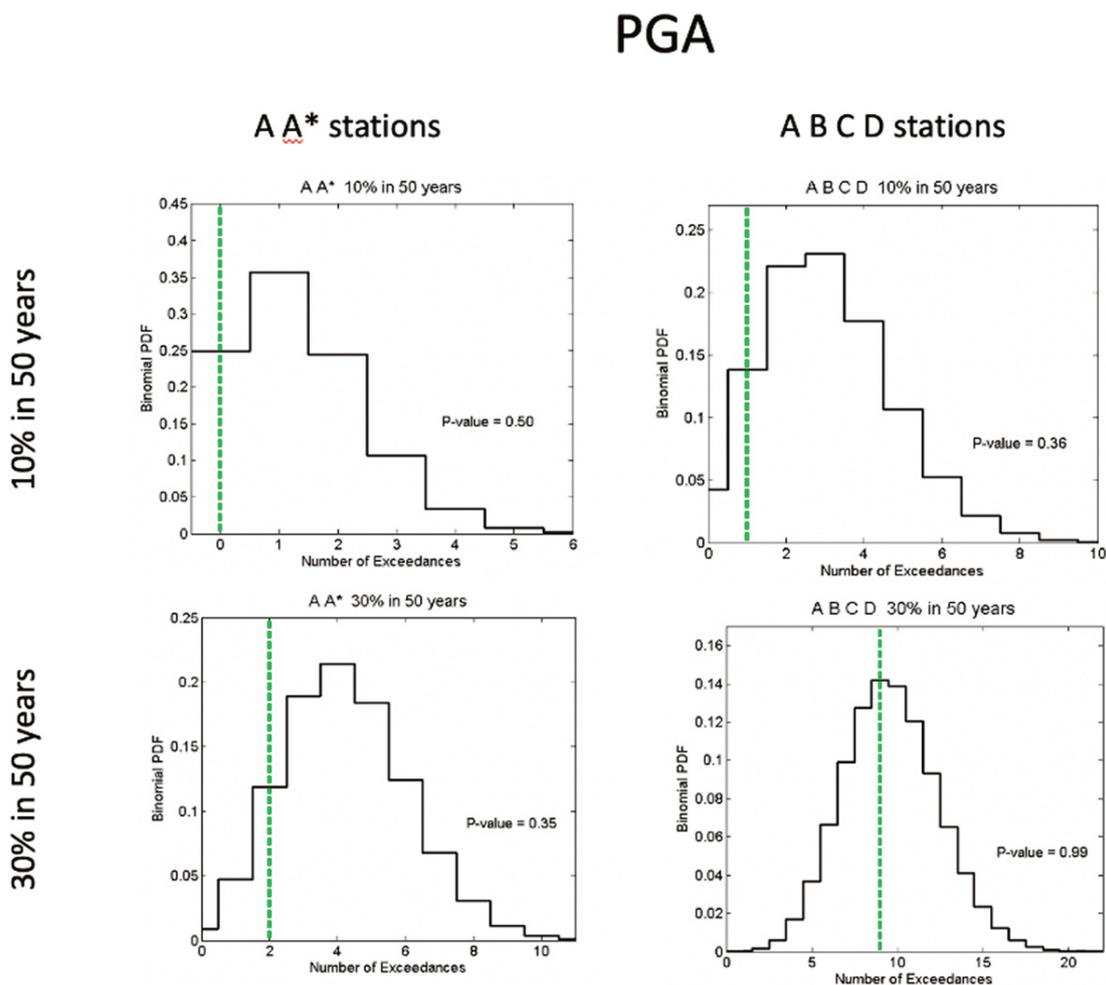


Figure 9. Results of the consistency test of MPS19 with the observed PGA values in the last decades for different PoE (10% and 30% in 50 years). The meaning of sites A, A*, B, C, and D is reported in the text.

We consider accelerometric stations on rock (EC8 A site category) and on all types of terrain (EC8 A, B, C and D) with flat topographical surface (i.e., T1 category according to NTC18 Italian building code; NTC [2018]). In case the station is not located on rock, the PGA observation is scaled down by the weighted average of the site coefficients of the 3 GMMs selected for ASCRs according to their weights. For rock sites we consider all stations, either with a good site characterization (type A), or with more uncertain site characterization (named type A*). When considering all types of terrain, we use only the stations with a good site characterization. The IM observed at different sites is correlated [Park et al., 2007], and this may be an issue for testing [Iervolino et al., 2017]. We adopt a pragmatic approach choosing sites that have a minimum distance of 30 km from one another. In this case, although part of the IM correlation remains independently from the distance [Park et al., 2007], we may assume that, for small PoE, exceedances are independent, i.e., the observation of an exceedance E_Y in one site Y does not modify the probability to observe an exceedance E_X in the nearby site X , $P(E_X|E_Y) = P(E_X) \equiv \text{PoE}$. Hence, the exceedances follow a binomial distribution with the parameter equal to the selected PoE. Worthy of note, this assumption makes the test of PGA conservative: if the null hypothesis (the exceedances are generated by a process which is well described by MPS19) is not rejected, it will not be rejected also relaxing the assumption, i.e., using a more exact distribution for the number of exceedances, which is overdispersed with respect to the binomial distribution [see Figure 2 in Iervolino et al., 2019]. In the analysis reported in Figure 9, we show the results of the test when the epistemic uncertainty of the PoE at each site is not included, which means to not consider possible common mode errors [Marzocchi and Jordan, 2018]. This also makes the test very conservative for our purposes, i.e., a non-rejection means that it will be also a non-rejection if uncertainties are included.

MPS19 is also tested against the macroseismic data observed in Italy as reported in the database DBMI15 v1.5 [Locati et al., 2016]. In this way, we can use a much longer time period of observations, i.e., several centuries. Since MPS19 does not currently produce estimates in macroseismic intensity, the first step is to convert the PGA exceedance rates $\lambda(x)$ where x is the PGA, in terms of rates of exceedance of macroseismic intensities $\lambda(I)$. This is done using the procedure suggested by D'Amico and Albarello [2008]

$$\lambda(I) = \sum_{i=1}^N \lambda(x_i) P(I|x_i) \quad (3)$$

where $P(I|x_i)$ is the empirical function for converting PGA into macroseismic intensity, and N is the number of PGA levels considered in the hazard curves of the model. Obviously, the passage introduces a further source of possible ontological error, as the empirical function for transforming PGA into macroseismic intensity is subject to high uncertainty. For this test, we account for these uncertainties using two functions estimated by Faenza and Michelini [2010; 2011], and Gomez Capera et al. [2020].

We select 27 sites that have a long seismic history for the analysis. The number of observations at each site takes into account the completeness of the data and considers only macroseismic intensities equal to 6 or higher which have been generated by mainshocks (macroseismic intensities associated to foreshocks and aftershocks are removed). To avoid the spatial correlation of macroseismic intensities, the distance between any couple of sites is always larger than 30 km (see the discussion above for testing the PGA exceedances), and we keep only one intensity (the maximum observed) for each earthquake; being generated by mainshocks following the Poisson distribution, we assume that these macroseismic observations are independent and follow the same distribution. This assumption facilitates the consistency test phase since we can test with aggregate observations. The procedure is similar to that described for the PGA exceedances. In this case, the epistemic uncertainty is taken into account, that is, $P(\lambda_{tot})$ is the pdf Gamma of the parameter λ_{tot} , which is estimated by the different (weighted) rates of all the models. Since there is considerable uncertainty and variability on the type of soil where macroseismic intensities have been observed, the test is carried out for two different types of soil: A and C.

Since all the GMMs used in MPS19 assume a linear behavior of soils, to estimate the PGA for type C soils the hazard curve calculated for type A soils is translated by a factor derived for Italy from the GMM by Bindi et al. [2011]. Note that the test for type A soils represents an underestimation of the number of exceedances because it does not take into account any site amplification.

Figure 10 displays the test plots for each macroseismic intensity value. These results show that MPS19 is compatible with the overall number of macroseismic intensities observed at the selected localities in the last centuries. As for the testing of PGA, we underline that any violation of the assumption on the independence of observations does not affect the conclusions of this test (MPS19 is not rejected). In Appendix F of the Supplementary

Material we add the same test for individual sites. The results show a general good agreement in all sites, with some exceptions (mostly when considering type A soils) where the number of observations for intensity 7 or less are not in agreement with the model.

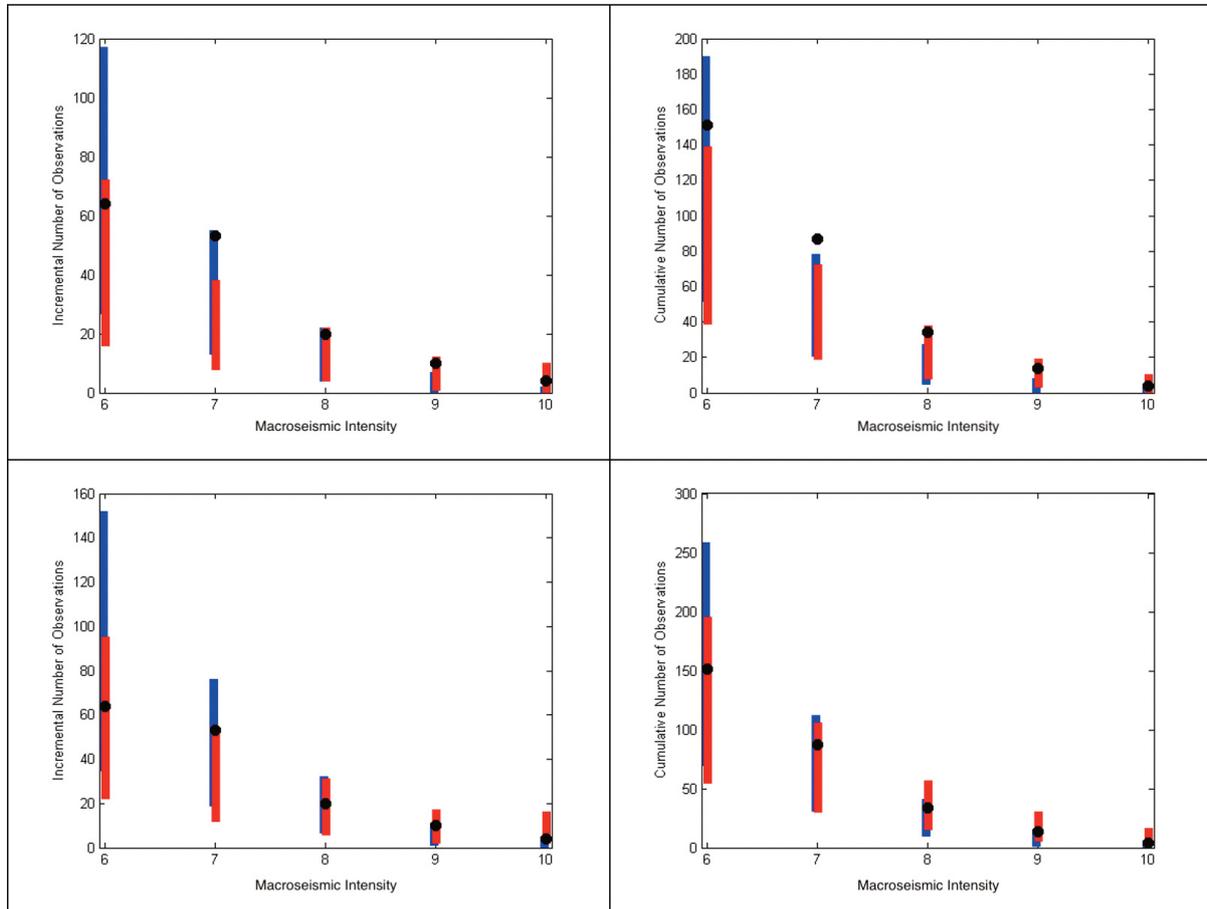


Figure 10. Results of the consistency test of MPS19 with the observed macroseismic intensities (black dots) of the last centuries for different soils. The upper row is for A sites; the lower row is for C sites. The left panels report the incremental number of observations for different macroseismic intensity, and the right panels the cumulative versions; the blue [Faenza and Michellini, 2010] and red bars [Gomez Capera et al., 2018] are the forecasts of MPS19 using two different transformation laws between PGA and macroseismic intensities (see text for more details).

8. A pilot modeling: MPS19_cluster

One of the basic assumptions in PSHA is that the earthquake occurrence has to be stationary in time and Poisson distributed [Cornell, 1968]. To satisfy this assumption seismic catalogs are commonly declustered, i.e., clusters of seismicity, which represent a clear deviation from the Poisson distribution, have to be removed. The basic rationale behind this choice is that by keeping the largest events we remove earthquakes that have a negligible contribution in terms of IMs and seismic hazard [Lomnitz, 1966]. However, recent earthquakes sequences, such as the Canterbury sequence 2010-2011 in New Zealand, Kumamoto sequence 2016 in Japan, the central Italy sequence 2016-2017, have raised doubts on the validity of this assumption, in particular for a complete risk analysis. For this reason, scientists are beginning to estimate the effects of the declustering procedure on PSHA.

As previously said, different declustering procedures exist, which may lead to important variability of the declustered seismicity rate [vanStiphout et al., 2011]. We think that such a variability cannot be considered as epistemic, because most of the declustering techniques do not do what they were planned for, i.e., leaving a mainshocks distribution similar to Poisson. So, even assuming that a true Poisson mainshocks distribution really exists (yet we are not aware of any convincing physical reason that allows us to postulate the existence of such unknown sharp distinction between mainshocks and all other earthquakes), the epistemic uncertainty cannot be estimated by the variability generated by the current declustering techniques.

In MPS19 we use GK74 declustering technique, because it is the most efficient procedure to get a mainshocks distribution similar to Poisson distribution (see Section 3). No matter of the declustering method used, this operation implies that strong ground shaking caused by aftershocks and foreshocks is not considered in classical PSHA [Boyd, 2012; Iervolino et al., 2012; 2019; Marzocchi and Taroni, 2014] and risk [Papadopoulos et al., 2020]. For this reason, we build a version of MPS19 which includes the effects of earthquake clustering (hereafter MPS19_cluster) and allows us to quantify its impact in terms of seismic hazard. MPS19_cluster is rooted in two different strategies, which have been described by Iervolino et al. [2012] and Marzocchi and Taroni [2014]. For the sake of example, in Figure 11 we show the MPS19_cluster obtained by applying only the procedure described by Marzocchi and Taroni [2014] on the mean ERM used in MPS19. The correction is magnitude-dependent in order to restore the magnitude-frequency distribution of the earthquakes before declustering, and it is applied to each spatial bin.

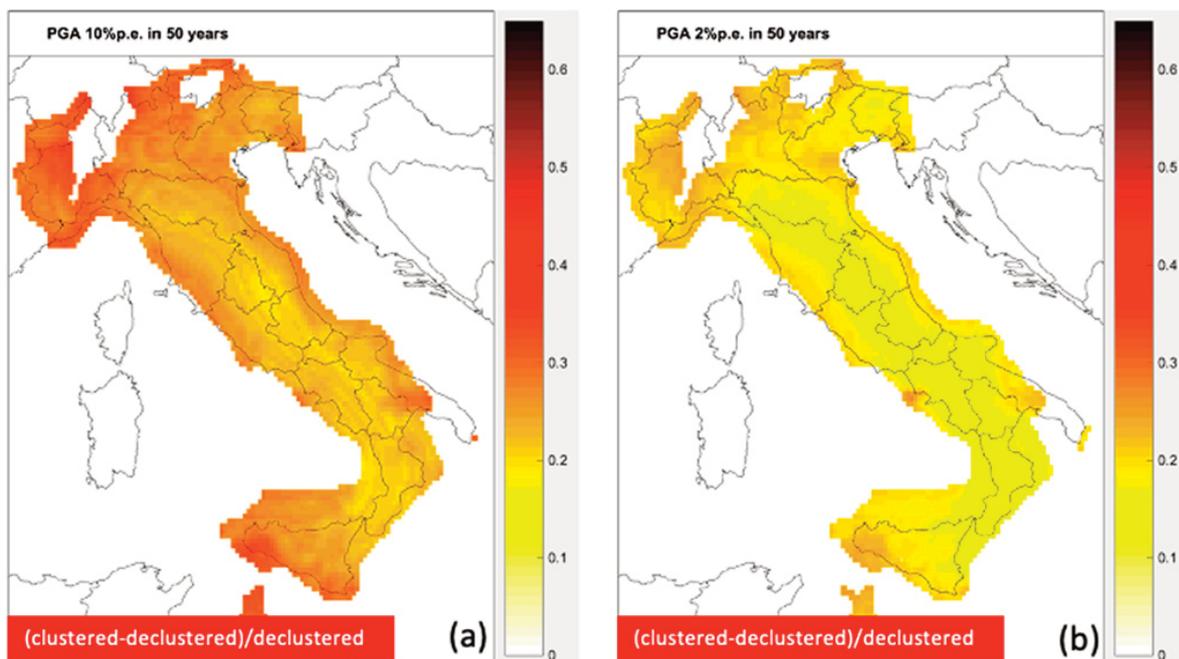


Figure 11. Variations of the (a) 10% PoE in 50 years, and (b) 2% PoE in 50 years PGA when the effect of clustered earthquakes are included to the maps of Figure 7b and 7c, respectively, according to the procedure described by Marzocchi and Taroni [2014]. The maps show only the sites with a PGA above the 25th percentile.

This figure intends to give a flavor of what the impact may be in the PGA estimated for 10% and 2% PoE in 50 years of MPS19 (Figure 7b and 7c respectively); from the figure we notice that the mean hazard could increase of about 20-25% in the regions of higher hazard for 10% PoE in 50 years with some increases of more than 30%. For the 2% PoE in 50 years, the increase is less strong being about 10-15% with some increases of about 25%.

A similar work is carried out also using the procedure described by Iervolino et al. [2012]; the outcomes will be described in a separate paper [Chioccarelli et al., 2021].

9. Final remarks, highlights, and open questions

MPS19 uses all new available data at the best and implements the most advanced practice introducing several novelties that are worth being remarked:

- MPS19 honors the hazard/risk separation principle, i.e., it does not include any subjective “precautionary” choices that should be introduced only at the risk level. In other words, all choices made in MPS19 are based on scientific thoughts, independently from the impact on the seismic hazard.
- MPS19 is modular, transparent and reproducible.
- MPS19 shows the aleatory variability and the epistemic uncertainty in the final output, making the interpretation of the dispersion of the hazard curves more clear, and providing the interpretation for any possible end-user.
- MPS19 has been checked through an extensive testing phase to assign weights to each component of the model, and to verify the reliability of the ERMs, GMMs and of the hazard outcomes in terms of PGA and macroseismic intensity at several sites and, on average, over the whole territory.
- We introduce MPS19_cluster, which is a pioneering attempt to correct the classical PSHA model to account for the large foreshocks and aftershocks that have been removed by the declustering technique.

At the same time, the work that has been made for MPS19 have highlighted also some challenges that should be addressed in the future. Some of them are general, others are specific for the Italian territory:

- MPS19_cluster shows the impact of removing earthquake clusters in terms of seismic hazard. Other methods based on synthetic catalogs may offer some additional interesting new perspectives in this field [e.g., UCERF3; Field et al., 2017].
- One of the basic assumptions in PSHA is that seismicity is stationary in time once the short-term clustering has been removed by declustering. In other words, time dependency on longer time scales is not included in this kind of modeling. This assumption is often taken for granted and not carefully tested. We think that in the future more efforts will be necessary to verify this assumption.
- The magnitude-frequency distribution has a large impact on PSHA. The common assumption of using a Gutenberg-Richter-like distribution over wide areas may be questionable, and currently we do not have enough data to test this assumption, at least for what concerns long-term PSHA purposes.
- GMMs are still mostly empirical and very likely not able to capture the details of IM variability at the spatial scale of interest for PSHA. Moving from general GMMs to more specific models which account for the regional geology may lead to significant variations of PSHA.

As a final remark, we argue that the procedures adopted in MPS19 make the comparison with the previous model MPS04 challenging. In particular, the elements of substantial difference are: i) MPS04 represents the median of the branches, MPS19 the average; ii) MPS04 represents the maximum horizontal component of the shaking, whereas MPS19 the geometric mean; iii) MPS04 made extensive use of expert judgment essentially aimed at introducing precautionary factors, whereas MPS19 keeps this judgment to a minimum, avoiding the use of precautionary components; iv) the basic data are completely different and up to date (more than 15 years have passed); v) MPS19 is based on a fully revised seismic catalog and magnitude estimated with more advanced procedures; vi) some earthquakes in MPS04 (based on the CPTI04 catalog) were aftershocks that have been removed in MPS19 by the declustering method; vii) MPS04 and MPS19 use completely different GMMs; viii) accelerometric databases used to define GMMs contain marked differences on the soil classification for the seismic stations.

Owing to these reasons, a detailed appraisal of the differences will be addressed in a separate future paper.

Acknowledgment. This study has benefited from funding provided by the Italian Presidenza del Consiglio dei Ministri - Dipartimento della Protezione Civile (DPC). This paper does not represent DPC official opinion and policies.

The authors want to thank all the participants to the project for the massive contribution at the final result. Being too many to be cited here, their names are listed in Appendix A of Supplementary Material. We want to thank INGV for supporting the MPS19 initiative, both for administrative issues and scientific contribution.

Finally, we want to thank the reviewers Paolo Bazzurro and Matt Gerstemberger for their comments which allowed us to improve significantly the readability and understandability of the paper.

References

- Abrahamson N.A., and J. J. Bommer (2005). Probability and uncertainty in seismic hazard analysis, *Earthq. Spectra*, 21, 603–607.
- Abrahamson N.A., W.J. Silva, R. Kamai (2014). Summary of the ASK14 Ground Motion Relation for Active Crustal Regions, *Earthq Spectra*, 30(3), 1025–1055.
- Abrahamson N.A., N. Gregor, K. Addo K (2016). BC Hydro Ground Motion Prediction Equations For Subduction Earthquakes, *Earthq Spectra*, 32(1), 23–44.
- Akaike H. (1974). A new look at the statistical model identification, *IEEE Transactions on Automatic Control*, 19 (6): 716–723.
- Akkar S., J.J. Bommer (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., M.A. Sandikkaya, J.J. Bommer (2014). 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.
- Albarelo D., R. Camassi, A. Rebez (2001). Detection of space and time heterogeneity in the completeness level of a seismic catalogue by a “robust” statistical approach: An application to the Italian area, *Bull. Seismol. Soc. Am.*, 91, 6, 1694–1703.
- Albarelo D., and V. D’Amico (2008). Testing probabilistic seismic hazard estimates by comparison with observations: An example in Italy, *Geophys. J. Int.*, 175, 1088–1094.
- Aspinall W.P. (2010). A route to more tractable expert advice, *Nature*, 436, 294–295.
- Azzaro R., Barberi G., S. D’Amico, B. Pace, L. Peruzza, T. Tuvè (2017). When probabilistic seismic hazard climbs volcanoes: the Mt Etna case, Italy. Part I: model components for sources parametrization, *Nat. Hazards Earth Syst. Sci.*, 17, 1981–2017.
- Bindi D., F. Pacor, L. Luzi, R. Puglia, M. Massa, G. Ameri, R. Paolucci (2011). Ground motion prediction equations derived from the Italian strong motion database, *Bull. Earthq. Eng.*, 9(6), 1899–1920.
- Bindi D., M. Massa, L. Luzi, G. Ameri, F. Pacor, R. Puglia, P. Augliera (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.
- Bommer J.J., J. Douglas, F. Scherbaum, F. Cotton, H. Bungum, D. Fah (2010). On the Selection of Ground-Motion Prediction Equations for Seismic Hazard Analysis, *Seismol. Res. Lett.*, 81(5), 783–793.
- Boncio P., D.P. Tinari, G. Lavecchia, F. Visini, G. Milana (2009). The instrumental seismicity of the Abruzzo Region in Central Italy (1981–2003): seismotectonic implications, *Ital. J. Geosci. (Boll.Soc.Geol.It.)*, 128, 367–380.
- Boore D.M. (2010). Orientation-Independent, Nongeometric-Mean Measures of Seismic Intensity from Two Horizontal Components of Motion, *Bull. Seismol. Soc. Am.*, 100(4), 1830–1835.
- Brooks E.M., S. Stein, B.D. Spencer, L. Salditch, M.D. Petersen, and D.E. McNamara (2017). Assessing earthquake hazard map performance for natural and induced seismicity in the central and eastern United States, *Seismol. Res. Lett.*, 89(1), 118–126.
- Campbell K.W., Y. Bozorgnia (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., E. Faccioli, M. Vanini, A. Bianchini (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.*, doi:10.1007/s10518-014-9685-y.
- Chiou B.S.J., R.R. Youngs (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.
- Cornell C.A. (1968). Engineering seismic risk analysis, *Bull. Seismol. Soc. Am.*, 58 (5), 1583–1606.
- Cotton F., F. Scherbaum, J.J. Bommer, H. Bungum (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.
- Chioccarelli E., P. Cito, F. Visini, and I. Iervolino (2021). Sequence-based hazard analysis based on a recent seismic

- source model for Italy, *Ann. Geophys.*
- D'Amico V., C. Meletti, B. Pace, A. Rovida, F. Visini (2019). Prodotto 2.13: Determinazione dello spessore sismogenetico e delle profondità ipocentrali. In: Meletti C. and Marzocchi W. (Eds.): Il modello di pericolosità sismica MPS19, Final Report, CPS-INGV, Roma, 168, 2 Appendices (in Italian).
- Delavaud E., F. Cotton, S. Akkar, F. Scherbaum, L. Danciu, C. Beauval, S. Drouet, J. Douglas, R. Basili, M. Abdullah Sandikkaya, M. Segou, E. Faccioli, N. Theodoulidis (2012). Toward a ground-motion logic tree for probabilistic seismic hazard assessment in Europe, *J. Seismol.*, 16, 451–473.
- Derras B., P.Y. Bard, F. Cotton (2014). Towards fully data driven ground-motion prediction models for Europe, *Bull Earthquake Eng.*, 12(1): 495–516.
- Devoti R., N. D'Agostino, E. Serpelloni, G. Pietrantonio, F. Riguzzi, A. Avallone, A. Cavaliere, C. Cheloni, C. Cecere, C. D'Ambrosio, L. Franco, G. Selvaggi, M. Metois, A. Esposito, V. Sepe A. Galvani, M. Anzidei (2017). A Combined Velocity Field of the Mediterranean Region, *Annals of Geophysics*, 60 (2), <https://doi.org/10.4401/ag-7059>.
- DISS Working Group (2018). Database of Individual Seismogenic Sources (DISS), Version 3.2.1: A compilation of potential sources for earthquakes larger than M 5.5 in Italy and surrounding areas, Istituto Nazionale di Geofisica e Vulcanologia, doi: <https://doi.org/10.6092/INGV.IT-DISS3.2.1>.
- Douglas J. (2019). Ground motion prediction compendium, <http://www.gmpe.org.uk> (last update 19 December 2019).
- Esposito S., and I. Iervolino (2011). PGA and PGV spatial correlation models based on European multievent datasets, *Bull. Seismol. Soc. Am.*, 101(5), 2532–2541.
- Faccioli E., A. Bianchini, M. Villani (2010). New ground motion prediction equations for $T > 1$ s and their influence on seismic hazard assessment. In: The University of Tokyo Symposium on Long-Period Ground Motion and Urban Disaster Mitigation, 23-29.
- Faenza L., and A. Michelini (2010). Regression analysis of MCS intensity and ground motion parameters in Italy and its application in ShakeMap, *Geophys. J. Int.* 180, 1138-1152.
- Faenza L., and A. Michelini (2011). Regression analysis of MCS intensity and ground motion spectral accelerations (SAs) in Italy. *Geophys. J. Int.* 186, 1415–1430. doi: 10.1111/j.1365-246X.2011.05125.x.
- Field E.H., K.R. Milner, J.L. Hardebeck, M.T. Page, N. van der Elst, T.H. Jordan, A.J. Michael, B.E. Shaw, and M.J. Werner (2017). A Spatiotemporal Clustering Model for the Third Uniform California Earthquake Rupture Forecast (UCERF3-ETAS): Toward an Operational Earthquake Forecast. *Bull. Seismol. Soc. Am.*, 107, 1049-1081.
- Gardner J.K., L. Knopoff (1974). Is the sequence of earthquakes in Southern California, with aftershocks removed, Poissonian? *Bull. Seismol. Soc. Am.* 64, 5, 1363-1367.
- Gasperini P., B. Lolli, G. Vannucci (2016). Catalogue of Mw magnitudes for the Italian area, 1981-2015, MPS16 Project Internal report.
- Gelman A., J.B. Carlin, H.S. Stern, and D.B. Rubin (2003). *Bayesian Data Analysis*, Second Ed., Chapman and Hall, London, United Kingdom.
- Gerstenberger M.C., W. Marzocchi, T. Allen, M. Pagani, J. Adams, L. Danciu, E. Field, H. Fujiwara, N. Luco, K-F Ma, C. Meletti, M. Petersen (2020). Probabilistic Seismic Hazard Analysis at Regional and National Scale: State of the Art and Future Challenges, *Rev. Geophys.*, 58, e2019RG000653.
- Gneiting T., and A.E. Raftery (2007). Strictly proper scoring rules, prediction, and estimation, *J. Am. Stat. Assoc.*, 102, 477, 359–378.
- Gomez Capera A.A., M. D'Amico, G. Lanzano, M. Locati, M. Santullin (2020). Relationships between ground motion parameters and macroseismic intensity for Italy, *Bull Earthq. Eng.*, 18, 5143-5164.
- Gruppo di Lavoro CPTI (2004). Catalogo Parametrico dei Terremoti Italiani, versione 2004 (CPTI04), INGV, Bologna, Italy, available at <http://emidius.mi.ingv.it/CPTI/>.
- Idriss I.M. (2014). An NGA-West2 Empirical Model for Estimating the Horizontal Spectral Values Generated by Shallow Crustal Earthquakes, *Earthq Spectra*, 30(3), 1155–1177, doi:10.1193/070613EQS195M.
- Iervolino I., M. Giorgio, and B. Polidoro (2012). Probabilistic seismic hazard analysis for seismic sequences, in Vienna Congress on Recent Advances in Earthquake Engineering and Structural Dynamics 2013, C. Adam, R. Heuer, W. Lenhardt, and C. Schranz (Editors), 28–30 August 2013, Vienna, Austria, Paper No. 66.
- Iervolino I. (2019). Generalized earthquake occurrence processes for sequence-based hazard, *Bull. Seismol. Soc. Am.*, 109, 1435-1450.
- Iervolino I., M. Giorgio, and P. Cito (2017). The effect of spatial dependence on hazard validation, *Geophys. J. Int.*, 209, 1363-1368.

- Jordan T.H. (2006). Earthquake predictability, brick by brick. *Seismol. Res. Lett.*, 77(1), 3–6.
- Jordan T.H., W. Marzocchi, A. Michael, M. Gerstenberger (2014). Operational earthquake forecasting can enhance earthquake preparedness, *Seismol. Res. Lett.*, 85(5), 955–959.
- Kuehn N.M., and F. Scherbaum (2015). Ground-motion prediction model building: a multilevel approach, *Bull. Earthquake Eng.*, 1–11, doi:10.1007/s10518-015-9732-3.
- Lanzano G., S. Sgobba, L. Luzi, R. Puglia, F. Pacor, C. Felicetta, M. D’Amico, F. Cotton, D. Bindi (2019). The pan-European Engineering Strong Motion (ESM) flatfile: compilation criteria and data statistics, *Bull. Earthquake Eng.*, 17, 561–582.
- Lanzano G., and L. Luzi (2019). A ground motion model for volcanic areas in Italy, *Bull. Earthquake Eng.*, 18(1).
- Lanzano G., L. Luzi, V. D’Amico, F. Pacor, C. Meletti, W. Marzocchi, R. Rotondi, E. Varini (2020). Ground Motion Models for the new seismic hazard model of Italy (MPS19): selection for active shallow crustal regions and subduction zones, *Bull. Earthq. Eng.*, 18, 3487–3516, doi:10.1007/s10518-020-00850-y.
- Locati M., R. Camassi, A. Rovida, E. Ercolani, F. Bernardini, V. Castelli, C.H. Caracciolo, A. Tertulliani, A. Rossi, R. Azzaro, S. D’Amico, S. Conte, E. Rocchetti (2016). Database Macrosismico Italiano (DBMI15), versione 1.5. Istituto Nazionale di Geofisica e Vulcanologia (INGV), <https://doi.org/10.6092/INGV.IT-DBMI15>.
- Lolli B., D. Randazzo, G. Vannucci, P. Gasperini (2020). The Homogenized Instrumental Seismic Catalog (HORUS) of Italy from 1960 to Present, *Seismol. Res. Lett.* 91 3208–3222, doi:10.1785/0220200148.
- Lomnitz C. (1966). Statistical prediction of earthquakes, *Rev. Geophys.* 4, 377–393.
- Luen B., and P.B. Stark (2012). Poisson tests of declustered catalogues, *Geophys. J. Int.* 189, 1, 691–700.
- Luzi L., S. Hailemikael, D. Bindi, F. Pacor, F. Mele and F. Sabetta (2008). ITACA (ITalian ACcelerometric Archive): a web portal for the dissemination of Italian strong-motion data. *Seismol. Res. Lett.*, 79 (5), 716–722.
- Luzi L., R. Puglia, E. Russo, M. D’Amico, C. Felicetta, F. Pacor, G. Lanzano, U. Çeken, J. Clinton, G. Costa, L. Duni, E. Farzanegan, P. Gueguen, C. Ionescu, I. Kalogeras, H. Özener, D. Pesaresi, R. Sleeman, A. Strollo, M. Zare (2016). The Engineering Strong-Motion database: a platform to access Pan-European accelerometric data, *Seismol. Res. Lett.*, 87(4), 987–997.
- Mak S., R.A. Clements and D. Schorlemmer (2014). The statistical power of testing probabilistic seismic-hazard assessments, *Seismol. Res. Lett.*, 85(4), 781–783.
- Mak S., and D. Schorlemmer (2016). A comparison between the forecast by the United States National Seismic Hazard Maps with recent ground motion records, *Bull. Seismol. Soc. Am.*, 106(4), 1817–1831.
- Marzocchi W., and J. D. Zechar (2011). Earthquake forecasting and earthquake prediction: Different approaches for obtaining the best model, *Seismol. Res. Lett.*, 82, 442–448.
- Marzocchi W., T.H. Jordan (2014). Testing for ontological errors in probabilistic forecasting models of natural systems, *Proc. Natl. Acad. Sci.*, 111(33), 11973–11978.
- Marzocchi W., M. Taroni (2014). Some thoughts on declustering in probabilistic seismic hazard analysis, *Bull. Seismol. Soc. Am.*, 104(4), 1838–1845.
- Marzocchi W., M. Taroni, J. Selva (2015). Accounting for epistemic uncertainty in PSHA: logic tree and ensemble modeling, *Bull. Seismol. Soc. Am.*, 105(4), 2151–2159.
- Marzocchi W., T.H. Jordan (2017). A unified probabilistic framework for seismic hazard analysis, *Bull. Seismol. Soc. Am.*, 107(6), 2738–2744.
- Marzocchi W., T.H. Jordan (2018). Experimental concepts for testing probabilistic earthquake forecasting and seismic hazard models, *Geophys. J. Int.*, 215(2), 780–798.
- McGuire, R. K., C. A. Cornell, and G. R. Toro (2005). The case for using mean seismic hazard, *Earthq. Spectra* 21, 879–886.
- Meletti C., V. D’Amico, F. Martinelli, A. Rovida (2019). Prodotto 2.15: Stima delle completezze di CPTI15 e suo declustering. In: Meletti C. and Marzocchi W. (Eds.): Il modello di pericolosità sismica MPS19. Final Report, CPS-INGV, Roma, 168 pp + 2 Appendices.
- Mousavi S.M. and G.C. Beroza (2018). Evaluating the 2016 one-year seismic hazard model for the central and eastern United States using instrumental ground-motion data, *Seismol. Res. Lett.*, 89(3), 1185–1196.
- MPS Working Group (2004). Redazione della mappa di pericolosità sismica prevista dall’Ordinanza PCM 3274 del 20 marzo 2003, Final Report, INGV, Milano-Roma 2004, 65 pp + 5 appendixes.
- Musson R.M.W. (2005). Against fractiles, *Earthq. Spectra* 21, 887–891.
- National Research Council (NRC) (1997). Review of Recommendations for Probabilistic Seismic Hazard Analysis:

- Guidance on Uncertainty and Use of Experts, National Research Council Panel on Seismic Hazard Evaluation, National Academy of Sciences, Washington, D.C.
- NTC (2018). Norme Tecniche per le Costruzioni NTC18, Ministerial Decree 17/01/2018, Italian Official Gazzette, 42, 20 February 2018.
- Pacor F., R. Paolucci, L. Luzi, F. Sabetta, A. Spinelli, A. Gorini, M. Nicoletti, S. Marcucci, L. Filippi, M. Dolce (2011). Overview of the Italian strong motion database ITACA 1.0, *Bull Earthquake Eng*, 9(6), 1723–1739. Doi:10.1007/s10518-011-9327-6.
- Pagani M., D. Monelli, G. Weatherill, L. Danciu, H. Crowley, V. Silva, P. Henshaw, L. Butler, M. Nastasi, L. Panzeri, M. Simionato and D. Vigano (2014). OpenQuake Engine: an open hazard (and risk) software for the Global Earthquake Model, *Seismol. Res. Lett.*, 85, 692-702.
- Papadopoulos A.N., P. Bazzurro, W. Marzocchi (2020). Exploring Probabilistic Seismic Risk Assessment Accounting for Seismicity Clustering and Damage Accumulation: Part I. Hazard Analysis, *Earthq. Spectra*, In press.
- Park J., P. Bazzurro, J.W. Baker (2007). Modeling spatial correlation of ground motion intensity measures for regional seismic hazard and portfolio loss estimation, In: *Applications of Statistics and Probability in Civil Engineering – Kanda, Takada & Furuta* (eds), Taylor & Francis Group, London, ISBN 978-0-415-45211-3.
- Pondrelli S., F. Visini, A. Rovida, V. D'Amico, B. Pace, C. Meletti (2020). Tectonic styles of expected earthquakes in Italy as an input for seismic hazard modeling, *Nat. Haz. Earth Syst. Sci.* <https://doi.org/10.5194/nhess-2020-70> (submitted).
- Roselli P., W. Marzocchi W., M.T. Mariucci, P. Montone (2018). Earthquake Focal Mechanism Forecasting in Italy for PSHA purposes, *Geoph. J. Int.* 212, 491-508.
- Rotondi R., E. Garavaglia (2002). Statistical analysis of the completeness of a seismic catalogue, *Nat. Hazards*, 25, 3, 245 - 258, doi10.1023/A:1014855822358.
- Rovida A., M. Locati, R. Camassi, B. Lolli, P. Gasperini (2016). CPTI15, the 2015 version of the Parametric Catalogue of Italian Earthquakes, Istituto Nazionale di Geofisica e Vulcanologia. doi: <https://doi.org/10.6092/INGV.IT-CPTI15>.
- Rovida A., M. Locati, R. Camassi, B. Lolli, P. Gasperini (2020). The Italian earthquake catalogue CPTI15, *Bull. Earthquake Engin.*, accepted.
- Scherbaum F., E. Delavaud, and C. Riggelsen (2009). Model selection in seismic hazard analysis: An information-theoretic perspective, *Bull. Seismol. Soc. Am.*, 99, 6, 3234–3247.
- Schorlemmer D., and M. Gerstenberger (2007). RELM Testing Center, *Seismol. Res. Lett.* 78, 1, 30–36.
- Schorlemmer D., M.J. Werner, W. Marzocchi, T.H. Jordan, Y. Ogata, D.D. Jackson, S. Mak, D.A. Rhoades, M.C. Gerstenberger, N. Hirata, M. Liukis, P. Maechling, A. Strader, M. Taroni, S. Wiemer, J.D. Zechar, J. Zhuang (2018). The collaboratory for the study of earthquake predictability: achievements and priorities, *Seismol. Res. Lett.*, 89(4), 1305-1313.
- Sieberg A. (1932). *Geologie der Erdbeben*, Handbuch der Geophysik, Vol. 2, Gebr. Bornträger, Berlin, Germany, 550–555 (in German).
- Skarlatoudis A.A., C.B. Papazachos, B.N. Margaris, C. Ventouzi, I. Kalogeras, 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, doi:10.1785/0120120265.
- SSHAC (Senior Seismic Hazard Analysis Committee) (1997). Recommendations for probabilistic seismic hazard analysis: guidance on uncertainties and use of experts, Report NUREG-CR-6372, U.S. Nuclear Regulatory Commission, Washington D.C.
- Stallone A., W. Marzocchi (2019). Features of seismic sequences are similar in different crustal tectonic regions, *Bull. Seismol. Soc. Am.*, 109(5), 1594-1604.
- Stirling M., and M. Gerstenberger (2010). Ground motion-based testing of seismic hazard models in New Zealand, *Bull Seismol. Soc. Am.*, 100, 1407–1414.
- Stucchi M., P. Albini, C. Mirto, A. Rebez (2004). Assessing the completeness of Italian historical earthquake data, *Ann. Geophys.*, 47, 659-674.
- Stucchi M., C. Meletti, V. Montaldo, H. Crowley, G.M. Calvi and E. Boschi (2011). Seismic hazard assessment (2003-2009) for the Italian building code, *Bull. Seismol. Soc. Am.*, 101, 1885-1911.
- Tusa G., H. Langer (2016). Prediction of ground motion parameters for the volcanic area of Mount Etna, *J. Seismol.*, doi:10.1007/s10950-015-9508-x.

- van Stiphout T., D. Schorlemmer, and S. Wiemer (2011). The Effect of Uncertainties on Estimates of Background Seismicity Rate, *Bull. Seismol. Soc. Am.*, 101, 482–494.
- Visini F., V. D’Amico, C. Meletti, B. Pace, A. Rovida, M. Santulin (2019). Prodotto 2.14: Modello della massima magnitudo. In: Meletti C. and Marzocchi W. (Eds.): *Il modello di pericolosità sismica MPS19. Final Report*, CPS-INGV, Roma, 168 pp + 2 Appendices.
- Woessner J., L. Danciu, D. Giardini, H. Crowley, F. Cotton, G. Grunthal, G. Valensise, R. Arvidsson, R. Basili, M. B. Demircioglu, S. Hiemer, C. Meletti, R. Musson, A. Rovida, K. Sesetyan, M. Stucchi, the SHARE consortium (2015). The 2013 European Seismic hazard model: key components and results, *Bull. Earthq. Eng.*, 13, 3553–3596.
- Working Group on California Earthquake Probabilities, WGCEP (1995). Seismic Hazards in Southern California: Probable Earthquakes, 1994 to 2024, *Bull. Seismol. Soc. Am.*, 85, 379–439.
- Zechar J.D., D. Schorlemmer, M. Liukis, J. Yu, F. Euchner, P.J. Maechling, and T.H. Jordan (2010a). The Collaboratory for the Study of Earthquake Predictability perspective on computational earthquake science, *Concurrency Comput. Pract. Ex.* 22, 12, 1836–1847.
- Zechar J.D., M.C. Gerstenberger, and D.A. Rhoades (2010b). Likelihood-based tests for evaluating space-rate-magnitude earthquake forecasts, *Bull. Seismol. Soc. Am.* 100, 3, 1184–1195.
- Zechar J.D., and J. Zhuang (2014). A parimutuel gambling perspective to compare probabilistic seismicity forecasts, *Geophys. J. Int.*, 199, 1, 60–68.
- Zhao B., J. Zhang, A. Asano, Y. Ohno, T. Oouchi, T. Takahashi, Ogawa H., Irikura K., Thio H.K., Somerville P.G., Fukushima Y., Fukushima Y. (2006). Attenuation Relations of Strong Ground Motion in Japan Using Site Classification Based on Predominant Period, *Bull. Seismol. Soc. Am.*, 96(3), 898.

***CORRESPONDING AUTHOR: Carlo MELETTI**

Istituto Nazionale di Geofisica e Vulcanologia

Sezione di Pisa, Via Cesare Battisti 53, 56125 Pisa, Italy

e-mail: carlo.meletti@ingv.it

© 2021 the Istituto Nazionale di Geofisica e Vulcanologia.

All rights reserved