Originally published as:

Mak, S., Clements, R., Schorlemmer, D. (2015): Validating Intensity Prediction Equations for Italy by Observations. - *Bulletin of the Seismological Society of America*, 105, 6, pp. 2942–2954.

DOI: <http://doi.org/10.1785/0120150070>

Validating Intensity Prediction Equations for Italy by Observations

by Sum Mak, Robert Alan Clements, and Danijel Schorlemmer

Abstract The engineering seismology community has recently recognized the importance of validating the performance of predictive models for seismic hazard by independent observations, yielding a number of studies on the relative performance of ground-motion prediction equations. The validation of intensity prediction equations (IPEs) has attracted less attention. We fill this gap by validating eight Italian IPEs plus one global IPE using five sets of Italian macroseismic intensity data, of which three are prospective and two retrospective to the models. We implemented multiple scoring methods to validate the models and found that the simple score of mean absolute error is sufficient to measure the general model performance. Good models consistently perform well under multiple methods and datasets, showing robustness. Models with physical functional forms are found to perform better. The global IPE performed well for Italian data, implying insignificant regional differences for IPEs. This result encourages grouping intensity data collected from multiple geographic regions, both from the Internet and traditional surveys, into a larger dataset for the use of future model development and validation.

Online Material: Table of selected earthquakes, figure showing data for intensity prediction equation (IPE) evaluation, figures for additional residual analyses, and figures showing evaluations based on other performance metrics.

Introduction

The ground shaking generated by an earthquake can be predicted by empirical equations as a function of physical parameters of the earthquake and a target site. Ground motions described in physical units (e.g., acceleration and velocity) are predicted by ground-motion prediction equations (GMPEs) and in macroseismic intensity (hereafter referred to as intensity) by intensity prediction equations (IPEs). These predictive equations are formulas that take earthquake-specific parameters (e.g., magnitude) and site-specific parameters (e.g., hypocentral distance) as inputs and return the expected value of ground shaking and/or the uncertainty of the prediction as outputs. Their use has become general practice in engineering seismology and seismic-hazard assessments (SHAs).

Like with any predictive models, the scientific approach demands continuous testing of the predictive power of the models using prospective information (e.g., Diebold *et al.*, 1998, for financial forecasts; Jordan, 2006; Laio and Tamea, 2007, for hydrological forecasts; Schorlemmer and Gerstenberger, 2007, for earthquake forecasts). Testing is important for validating decisions made based on model predictions; test results form the basis of model improvement. Recently, the importance of model validation using independent observations has been recognized by the engineering seismology community: since the completion of the first phase of the Next Generation Attenuation project (Power *et al.*, 2008),

there have been a number of studies dedicated to validating and comparing the newly developed GMPEs (Stafford *et al.*, 2008; Delavaud *et al.*, 2009, 2012; Douglas and Mohais, 2009; Scasserra *et al.*, 2009; Kakkamanos and Baise, 2011; Arango *et al.*, 2012; Beauval *et al.*, 2012; Vilanova *et al.*, 2012). Through these studies, the improvements in predictive power of the new models over old models are documented and the applicability of individual models to specific regions are tested.

Macroseismic intensity (Grünthal, 2011) is a useful tool for purposes such as loss estimation (e.g., Eguchi *et al.*, 1997; Grünthal *et al.*, 2006; Wyss, 2008), communication of earthquake effects to the public (e.g., Wald, Quitoriano, Heaton, *et al.*, 1999), and SHAs in regions with abundant historical earthquake records and moderate seismicity but sparse seismometer coverage (e.g., Musson, 2000; Bindi *et al.*, 2011). Because the evaluation of building vulnerability is part of both macroseismic investigation and risk assessment, it is a natural advantage for intensity as the basis of risk assessment, leading to the continual popularity of the use of intensity for this purpose. Therefore, although physical intensities such as peak ground acceleration are often the unit of choice for recent SHAs, leading to an apparent decline in the use of IPEs, IPEs continue to be a crucial component for risk assessments.

Table 1
Selected Intensity Prediction Equations (IPEs) (Basic Information)

ID	Reference	Intensity Scale	Database	Region
1987GPT	Grandori <i>et al.</i> (1987)	Unspecified	Unspecified	Italy
1996P	Peruzza (1996)	MCS*	GNDT94 [†]	Italy
2000P	Peruzza (2000)	MCS*	DOM 4.1 [†]	Italy
2001G	Gasperini (2001)	MCS*	DOM 4.1	Italy
2004AD	Albarelo and D'Amico (2004)	MCS*	DOM 4.1	Italy
2006G	Gómez Capera (2006)	MCS*	DBMI04	Italy
2008PAG*	Pasolini <i>et al.</i> (2008)	MCS*	DBMI04	Italy
2010SSG	Sørensen <i>et al.</i> (2010)	MCS*	DBMI04	Campania, Italy
2012AWW	Allen <i>et al.</i> (2012)	Mixed	Atlas of ShakeMaps [§]	Global

*Mercalli–Cancani–Sieberg (MCS).

[†]See Data and Resources.

^{*}Model 2008PAG was separated into 2008PAG_I and 2008PAG_M in the current study.

[§]Allen *et al.* (2008).

Although new IPE models are being developed, the validation of IPEs has received less attention compared with that of GMPEs. Cua *et al.* (2010) evaluated a set of IPEs based on residual analysis, which provides a visual but less quantitative expression of the model performance. The issue of regional dependence of models was implicitly ignored when they evaluated region-specific models with a set of global data. Their study is seminal in this field, yet is preliminary.

The goal of our study is to extend the data-driven validation procedures from GMPEs to IPEs. We emphasize using prospective data (i.e., data not used for developing the models) to test the predictive power of the models. We use Italy as a test case for two reasons: first, there are a large number of IPEs developed specifically for Italy to allow meaningful comparison; second, the Italian intensity records, both retrospective and prospective, are among the best in the world in terms of quantity and quality. These two conditions enable the evaluation of region-specific models by region-specific data, and so avoid the issue of regional dependence of models.

We emphasize that data-driven validation is not equivalent to model selection. When selecting an attenuation model to be used in hazard modeling, for example, additional factors that involve expert judgment will need to be considered (e.g., Cotton *et al.*, 2006), because the available data are usually insufficient to cover all situations of interest, such as the performance of models when applied to earthquakes of unprecedentedly large magnitudes. On the other hand, within the parameter range where data are available, we consider the data-driven validation as the best estimation of model suitability.

Models

In this study, we evaluated Italian IPEs that match the following selection criteria:

1. The IPE is published in English in a journal or conference proceeding in complete article form, such that the model can be reasonably reproduced.

2. The IPE predicts an intensity value as a function of source–site distance. The predictions from various IPEs are therefore directly comparable.

We selected a total of eight Italian IPEs for comparison (Grandori *et al.*, 1987; Peruzza, 1996, 2000; Gasperini, 2001; Albarello and D'Amico, 2004; Gómez Capera, 2006; Pasolini *et al.*, 2008; and Sørensen *et al.*, 2010). Another four IPEs (Berardi *et al.*, 1994; Magri *et al.*, 1994; Rotondi and Zonno, 2004; and Faccioli and Cauzzi, 2006) were inspected but excluded from our study. Berardi *et al.* (1994) is not published in English, but this model was included as part of Peruzza (2000). Magri *et al.* (1994) and Rotondi and Zonno (2004) are two attempts to develop IPEs to address the probabilistic nature of the data and prediction; they both focus on methodological development but do not provide stand-alone models usable on prospective data. Faccioli and Cauzzi (2006) is not published in complete article form and does not provide sufficient details for the model to be reproduced.

In addition to IPEs specifically developed for Italy, we included one global IPE recently developed by Allen *et al.* (2012). Region-specific features are often considered to be significant in the development and application of GMPEs and IPEs. Including a global model could verify this general belief. The selected IPEs are summarized in Tables 1–3. Each of them is hereafter referred to by its model identification (see ID in Table 1).

Cua *et al.* (2010) regarded IPEs that do not use moment magnitude and rupture distance as unfavorable because the moment magnitude is the current standard to represent earthquake size and the rupture distance avoids the unrealistic point-source assumption at the near field. Our study, however, included IPEs that use epicentral intensity (i.e., the intensity at the earthquake epicenter, denoted as I_0) as the measure of earthquake size. They are hereafter referred to as I_0 models, in contrast to M models that use magnitude. We also included models that use epicentral distance (R_e) as the measure of source–site distance. I_0 models are applicable in SHAs that

Table 2
Selected IPEs (Form and Range)

ID	Functional Form*	Input Parameters Range*	σ
1987GPT	$dI = \frac{1}{\ln w} \ln \{1 + (\frac{w-1}{\psi_0}) (\frac{R_e}{D_0} - 1)\}$	N/A [†]	N/A [†]
1996P	$dI = \frac{1}{\ln w} \ln \{1 + (\frac{w-1}{\psi_0}) (\frac{R_e}{D_0} - 1)\}$	VII $\leq I_0 \leq$ XI	
2000P	$dI = \frac{1}{\ln w} \ln \{1 + (\frac{w-1}{\psi_0}) (\frac{R_e}{D_0} - 1)\}$	VII $\leq I_0 \leq$ XI	
2001G	$dI = a_1 R_e + a_3$ if $R_e \leq 45$ $dI = a_1 R_e + a_3 + a_4 (R_e - 45)$ if $R_e > 45$	$10 \leq R \leq 180$	1.04 [‡]
2004AD	$I = c_3 I_0 - (a_1 \sqrt{R_e^2 + h^2} + a_2 \ln(\sqrt{R_e^2 + h^2}) + a_3)$	$15 \leq R \leq 300$	1.25
2006G	$dI = a_1 \sqrt[3]{R_e} + a_3$	VII $\leq I_0 \leq$ XI $R \leq 200$	0.94
2008PAG	$I = I_E - a_1 (\sqrt{R_e^2 + 3.91^2} - 3.91) - a_2 \ln(\frac{\sqrt{R_e^2 + 3.91^2}}{3.91})$ $I_E = c_1 M + c_2$ or $I_E = c_3 I_0 + c_4$	$4.4 \leq M_w \leq 7.4$ [§] $R \leq 200$	0.98 (for I_0) 0.87 (for M)
2010SSG	$I = (c_1 M + c_2) - \{a_1 (\sqrt{R_e^2 + h^2} - h) + a_2 \log_{10}(\frac{\sqrt{R_e^2 + h^2}}{h})\}$	IX – X $\leq I_0 \leq$ XI $6.3 \leq M_w \leq 7.0$ $R \leq 300$	0.971
2012AWW	if $R \leq 50$: $I = (c_1 M + c_2) - a_2 \ln(\sqrt{R_h^2 + R_m^2})$ else: $I = (c_1 M + c_2) - \{a_2 \ln(\sqrt{R_h^2 + R_m^2}) + a_4 \ln(\frac{R_h}{50})\}$	$4.9 \leq M_w \leq 7.9$ $6 \leq R \leq 300$	0.87

*Distance (R) and focal depth (h) in kilometers. Notations may deviate from those used in the original text to facilitate comparison.

Regression coefficients are denoted as a_i (for distance) and c_i (for magnitude).

[†]Unspecified in the reference.

[‡]Reported by Carletti and Gasperini (2003).

[§]Not specified in the original article but reported by Cua *et al.* (2010).

^{||}At $R_e = 60$; distance dependent.

use I_0 to describe earthquake size (Musson, 2000). This approach was adopted by a major SHA project in Italy (Slejko *et al.*, 1998). Epicentral distance is the only choice when an IPE is developed from historical earthquake data, in which information like focal depth and rupture plane are often impossible to accurately estimate. Because Italian IPEs are mostly developed using historical earthquake records, most of them use epicentral distance as the distance measure.

The 1987GPT (Grandori *et al.*, 1987), 1996P (Peruzza, 1996), and 2000P (Peruzza, 2000) models are the foundation of intensity attenuation of the SHA of Slejko *et al.* (1998). The 1987GPT model adopted an empirical functional form (see Table 2) that emphasizes the performance at near field and incorporated an I_0 -dependent attenuation rate. Peruzza (1996) divided Italy into a number of geographic regions and computed one set of coefficients for each region to create the 1996P model. It adopted the same functional form as 1987GPT but discarded the I_0 -dependent attenuation rate because the modelers found insufficient empirical support for I_0 -dependent attenuation. This IPE was then extended by dividing Italy into even more regions and became the 2000P model, which was finally adopted by Slejko *et al.* (1998). In this model, one set of coefficients was computed for each region using one representative (usually the strongest) earthquake from each respective region.

Although no further intensity-based SHAs had been developed for Italy of comparable scale since Slejko *et al.* (1998), Italian IPEs continued to evolve, resulting in the 2001G (Gasperini, 2001), 2004AD (Albarello and D’Amico, 2004), 2006G (Gómez Capera, 2006), and 2008PAG (Pasolini *et al.*, 2008) models. Many of these IPEs are based on the same or related

data sources as earlier models (Table 1), meaning that the data they used are highly overlapping. Various functional forms were investigated in Gasperini (2001), resulting in the bilinear 2001G model. Some function forms, such as the bilinear form for 2001G, the cubic form for 2006G, and the customized form for 1987GPT, 1996P, and 2000P, do not have direct physical meaning. On the other hand, the 2004AD and 2008PAG models are based on a classical functional form (von Kövesligethy, 1906; Gupta and Nuttli, 1976, hereafter referred to as “physical functional form”) with the physical terms of geometric spreading and anelastic attenuation, similar to the form used by modern GMPEs.

Unlike the above-mentioned models, which are designed for Italy in general, two of the IPEs analyzed in the present study are designed for other regions. The 2010SSG model (Sørensen *et al.*, 2010) is derived from large earthquakes in the volcanic region of Campania, and so is not expected to be comparable with models developed from both moderate and large earthquakes in an active shallow-crust region. We included this model for completeness only. The 2012AWW model (Allen *et al.*, 2012) is developed using 119 worldwide earthquakes occurring from 1960 to 2008, including only seven of which occurred in Italy; it is therefore not an indigenous Italian IPE. It uses the classical functional form similar to that used by 2004AD, 2008PAG, and 2010SSG but does not include the anelastic attenuation term. Furthermore, it is derived from a dataset of mixed intensity scales, assuming that they are equivalent to each other. In this study, the intensity values predicted by all IPEs are assumed to be in an equivalent scale, regardless of which intensity scale each IPE assumes.

Table 3
Selected IPEs (Data)

ID	Data Amount	Latest Event (yyyy/mm/dd)
1987GPT	19 earthquakes	Before 1987
1996P	54 earthquakes divided into 8 zones, each containing up to 13 earthquakes	Before 1994
2000P	59 zones, each containing 1 earthquake	Before 1997
2001G	~20,000 intensity reports	Before 1997
2004AD	369 earthquakes, 14,870 intensity reports	Before 1997
2006G	212 earthquakes, 20,873 intensity reports	2002/10/31
2008PAG	470 earthquakes, ~22,500 intensity reports	2002/10/31
2010SSG	9 earthquakes	2002/10/31
2012AWW	119 global earthquakes, 13,077 intensity reports	2002/10/31*

*For Italian earthquakes.

Several of the selected IPEs use various optimization methods or subsetting of the data when fitting the model, resulting in more than one set of coefficients. When the modelers did not specify which set of the resulting coefficients is preferred, we picked the one that makes the model the most comparable with other models. For models 2001G and 2006G, we took the coefficients computed using the entire set of data. For model 2010SSG, we took the coefficients computed using epicentral distance and Monte Carlo regression because only the epicentral distance is available in the majority of the test data. For model 2008PAG, both I_0 and moment magnitude (M_w) can be used as the parameter of earthquake size. We considered it as two separate models (2008PAG_I and 2008PAG_M).

Data

The Databases

We compared observed intensity values of a set of earthquakes with the predicted intensity values of each model. The observed intensities were taken from three intensity databases, namely “Hai sentito il terremoto?” (HSIT; Sbarra *et al.*, 2010), “Did You Feel It?” (DYFI; Wald, Quitoriano, Dengler, *et al.*, 1999), and Database Macrosismico Italiano 2011 (DBMI11). See Data and Resources for the online references for these databases. Both HSIT and DYFI collect questionnaires answered by Internet users and compile the responses from each locality (i.e., town, city) into an intensity value assigned to that locality. HSIT provides intensity reports in both Mercalli–Cancani–Sieberg (MCS; Sieberg, 1923) and EMS-98 (European Macroseismic Scale; Grünthal, 1998) scales. We took the former because most Italian IPEs were developed from data in the MCS scale; Musson *et al.* (2010) concluded equality between intensity assignments in using MCS and EMS-98. DBMI11 reports intensity in the traditional way, in which the intensity value is derived from expert reviews of the literature describing historical earthquakes and field reports or paper-based questionnaire surveys for modern earthquakes. It contains the largest amount of data among the three databases because it includes historical earthquakes dating back to 1000 years ago. It is also the only database that contains a signifi-

cant proportion of large earthquakes. The data selection process is described in the Data Selection section.

Input Parameters

Unless described in this section, all input parameters required by each IPE are provided directly by the earthquake parameters reported together with the intensity reports in each database. Some earthquakes are covered by multiple databases (see Table S1, available in the electronic supplement to this article), and their hypocentral parameters reported by different databases may not be identical. We tolerated this discrepancy because it did not affect the relative performance of the models.

Epicentral Intensity. The epicentral intensity, I_0 , is not reported in the HSIT and DYFI database but is a required input parameter for I_0 models. We followed the method described by Gasperini *et al.* (1999, their appendix 2) for I_0 estimation: I_0 is taken as the maximum reported intensity ($I_{\max}$) if the $I_{\max}$ value is assigned for at least two localities; otherwise, I_0 is taken as $I_{\max} - 1$. Formally, let I_i be the ordered observed intensity values such that $I_1 \geq I_2 \geq \dots \geq I_n$, in which n is the total number of observed intensity values in the test dataset. Then

$$I_0 = \begin{cases} I_1 & \text{if } I_1 = I_2, \\ I_1 - 1 & \text{otherwise.} \end{cases} \quad (1)$$

We found that the I_0 values reported in DBMI04 (an earlier version of DBMI11, see Data and Resources) are consistent with the above-mentioned method of I_0 estimation. In DBMI04, about two-thirds of earthquakes have $I_0 = I_{\max}$. For the remaining one-third, about 90% have $0 < I_0 - I_{\max} \leq 1$. Because all I_0 models evaluated in our study used (at least partly) the data from DBMI04 for model fitting (see Table 1), we consider our treatment of I_0 estimation appropriate.

Intensity reports from the DYFI and some reports from HSIT are not ordinal but with increments of 0.1. Therefore, it is less likely for the same $I_{\max}$ value to be assigned for multiple localities. To account for this, when estimating I_0 for these two databases, we modified the above method to

$$I_0 = \begin{cases} I_1 & \text{if } (I_1 - I_2) < 0.5, \\ I_1 - 1 & \text{otherwise.} \end{cases} \quad (2)$$

Magnitude. DYFI and DBMI11 report moment magnitudes, whereas HSIT reports only local magnitudes. For events in HSIT, we used the moment magnitude reported by Time Domain Moment Tensor (see [Data and Resources](#)) whenever possible, otherwise we used the local magnitude reported by HSIT.

Distance. All IPEs accept either the epicentral (R_e) or hypocentral distance (R_h) as input. These distances are either provided in the databases or can be easily calculated from the given information in the databases. The 2010SSG and 2012AWW models also accept the Joyner–Boore distance and rupture-plane distance, respectively, but we did not use these distance measures because for the moderate earthquakes that dominate our study, the point-source assumption is largely valid and so various definitions of magnitude and distance do not differ much. Besides, it is impossible in most cases to determine the rupture plane of the test events.

Data Selection

Although DBMI11 is a static database, the two Internet databases, HSIT and DYFI, are being updated continuously. We used data from HSIT that were recorded since the establishment of the database (in 2007) until the end of February 2013. We used data from DYFI since 2004, which is the time the database began to include non-U.S. earthquakes, until the end of February 2013. We used only earthquakes that occurred onshore in Italy to ensure the test data are suitable for the selected Italian IPEs. We used only widely felt (defined below) earthquakes because poorly reported earthquakes are unimportant for most hazard-related purposes; the data quality of poorly documented events is also questionable. For the DBMI11 database, widely felt earthquakes were defined as those reported by at least 20 localities. For the two Internet databases (HSIT and DYFI), only events with more than 100 returned questionnaires were used.

Some IPEs specify the range of the values of the data used in fitting the model (Table 2). To avoid extrapolation, the prescribed range of input values should be followed. It is impractical, however, to strictly follow the magnitude limit because there are too few earthquakes of significant size, especially in the modern databases. Releasing this limit is necessary for including more test data. In this study, the smallest magnitude used was 4.4, which is the lower limit for the 2008PAG model (Table 2). The minimum epicentral intensity used was V, which is lower than the limit of some IPEs but necessary to include more test data. Intensity reports with $I \geq \text{II}$ at $15 \leq R_e \leq 80$ km were used. The lower distance limit is the smallest number that is valid for all models. The upper limit was selected to restrict the data to the near field. We also tentatively used a larger upper bound of 180 km and a smaller lower bound of 0 km, but this did not significantly change the relative performance of the IPEs.

All selected IPEs are designed for shallow tectonic earthquakes so we excluded deep (focal depth > 30 km) and volcanic earthquakes near Mt. Etna (25 km within 37.73° E, 15.00° N) from our test data. The model 2012AWW requires hypocentral distance as input, so focal depth is required. For historical earthquakes in DBMI11 that have no estimation of focal depth, we assigned a hypothetical depth of 10 km. The traditional intensity values reported by DBMI11 occasionally come with an uncertainty. In such cases, we took the value as the average of the two bounds (e.g., VI–VII becomes 6.5).

It is important to differentiate between retrospective and prospective testing. Retrospective testing uses the same set (or a subset) of data used in fitting the model and shows only how well a model fits the input data. On the contrary, in prospective testing, model predictions are tested against data not seen by the model. This provides a better measure of a model's predictive power and hence its practical usefulness.

Data in HSIT and DYFI are prospective to all selected IPEs because all earthquakes in these databases occurred after the latest earthquakes that each IPE included (Table 3). On the other hand, all selected Italian IPEs used either DBMI04 directly or previous compilations of intensity data included in DBMI04 in their model developments (Table 1). Therefore, DBMI11 contains both prospective and retrospective test data. We isolated the portion of prospective data of DBMI11, denoted as DBMI11_p, which is the portion not included in DBMI04 (i.e., earthquakes occurred after October 2002). The retrospective portion of DBMI11 contains a large number of earthquakes and allows more detailed segregation. We divided it into two parts: DBMI11_{RM} for moderate earthquakes ($I_0 \leq \text{VIII}$) and DBMI11_{RL} for large earthquakes ($I_0 > \text{VIII}$). We focused on the IPE performance on prospective data but also show their performance under retrospective data as reference. Table 4 summarizes the datasets used in this study.

After the data selection, HSIT, DYFI, DBMI11_p, DBMI11_{RM}, and DBMI11_{RL} contain 23, 13, 11, 361, and 76 events, corresponding to 4525, 721, 1073, 17,678, and 7686 intensity reports, respectively. Eight of the 13 DYFI events are also included in HSIT, and one of the 13 DYFI events is included in DBMI11_p. The 2009 L'Aquila earthquake, which is the most significant Italian earthquake in the past decade, is included in HSIT, DYFI, and DBMI11_p. Because these databases collect data and estimate intensities in different ways, we consider them independent of each other despite the common events. The locations of the epicenters for the selected events are shown in Figure 1. [Ⓔ] Table S1 gives a list of events in the three prospective datasets. [Ⓔ] Figure S1 shows the distance distribution of the data.

Validation Methods

Cotton *et al.* (2006) provide criteria to be considered when selecting GMPEs. Although many of them are applicable for selecting IPEs, this study focuses on data-driven model validation, so expert judgments are not discussed here. Data-driven model validations can be categorized as two ap-

Table 4
Summary of the Five Datasets

Dataset*	Properties				$N_{eq}^{\dagger}$	$N^{\ddagger}$
HSIT	Internet	Pseudocontinuous	Prospective	§	23	4525
DYFI	Internet	Pseudocontinuous	Prospective	§	13	721
DBMI11 _P	Traditional	Ordinal	Prospective	§	11	1073
DBMI11 _{RM}	Traditional	Ordinal	Retrospective	$I_0 \leq VIII$	361	17,678
DBMI11 _{RL}	Traditional	Ordinal	Retrospective	$I_0 > VIII$	76	7686

*HSIT, “Hai sentito il terremoto?”; DYFI, “Did You Feel It?”; DBMI11, Database Macrosismico Italiano 2011.

[†]Number of selected events.

[‡]Number of intensity reports.

[§]See ⑤ Table S1 in the electronic supplement to this article for detailed information about earthquake size.

proaches, as described by [Stewart \(2010\)](#), his sections 3.2.3–3.2.4): the measure-oriented approach and the diagnostic approach. The measure-oriented approach uses a score to describe the overall goodness of fit between model predictions and actual observations. This unifying representation of model performance allows straightforward comparison between multiple models. The overall nature of the score, however, has the shortcoming of not indicating which part of a model performs well or badly. In contrast, the diagnostic approach focuses on investigating the performance of model components but is inconvenient for model comparison. In general, these two approaches complement one another to provide a more complete evaluation of model performance. The multiple models and datasets in this study highlight the importance of relative performance. Therefore, we used mainly the score to evaluate the IPEs. The diagnostic approach was used as a supplement to visually verify the results shown by the score and to ensure that the resulting score is representative instead of dominated by a subset of data.

Measure-Oriented Evaluation

We used the weighted mean absolute error (wMAE) as the measure of the goodness of fit between model predictions and actual observations. This is a widely used non-parametric score for evaluating point predictions. We tried another common nonparametric score, the mean squared error, and found the results to be similar. We did not use probabilistic scores as the major measure of model performance because three IPEs (1987GPT, 1996P, and 2000P) in this study are not probabilistic and predict only points. We acknowledge the potential advantage of probabilistic scores, and we discuss a tentative implementation in the [Appropriateness of the Score](#) section.

The wMAE for earthquake i is defined as

$$wMAE_i = \frac{\sum_{j=1}^{N_i} (w_{ij} |res_{ij}|)}{\sum_{j=1}^{N_i} w_{ij}}, \quad (3)$$

and

$$res_{ij} = I_{obs,ij} - I_{pre,ij}, \quad (4)$$

in which $I_{obs,ij}$ and $I_{pre,ij}$ are, respectively, the observed and predicted intensity values for record j of earthquake i . N_i is the number of records for earthquake i and w_{ij} is the weight for $I_{obs,ij}$. We weighted each intensity observation of the two Internet databases (HSIT and DYFI) by the number of questionnaires used to compile it, because we found the residuals, defined in equation (4), converge as the number of questionnaires increases (⑤ Figs. S2 and S3). For DBMI11, we did not weight the data (i.e., set $w_{ij} = 1$) because in the traditional intensity database there is no indicator for the accuracy of each intensity report. We tried two other weighting schemes, equal weighting and capped weighting (i.e., setting the maximum weight of an individual observation at 20), and found that the relative model performance was largely unchanged.

We took the average value of all event-based wMAE scores as the performance indicator for each IPE. Compared with simply averaging all absolute errors, our event-based approach ensures that the score is not dominated by a single earthquake that generated a large number of observations.

Although it is desirable for a score to have a unit with direct physical meaning, in practice most scores, even those as simple as the (weighted) mean absolute error used in this study or the root mean square error used in many other applications, do not have an absolute and quantitative physical meaning. Although the wMAE is a measure of the deviation of a model prediction from the observation, it is meaningful only in the relative sense that a model with a lower score is better than one with a higher score.

Diagnostic Evaluation

The diagnostic evaluation investigates the dependence of residuals (defined in equation 4) to input parameters. For each IPE, we plot the residuals against the two major input parameters, namely I_0 and R_c . The residuals for a good model should be completely random and show no trend.

Results

Figure 2 shows the development along events (ordered chronologically) of the average event-based wMAE scores for each IPE. The lines do not cross one another often, implying

Figure 1. Selected earthquakes for the five datasets (see Table 4). The size of a symbol is proportional to the magnitude of the earthquake. The color version of this figure is available only in the electronic edition.

that the overall performance for the good models remained stable over the events. Table 5 shows the ranking of the IPEs based on the wMAE averaged over all events. The oldest model (1987GPT) and the regional model (2010SSG) consistently

ranked last. For the three prospective datasets (HSIT, DYFI, and DBMI11_p) as well as the retrospective dataset for moderate earthquakes (DBMI11_{RM}), the rankings of the IPEs were highly consistent. The 2012AWW and 2004AD models

Figure 2. Development of the average value of event-based weighted mean absolute error ($wMAE_i$). The events are ordered chronologically. The score on each event is the average wMAE of all preceding events (inclusive). The score on the last event is the average of the wMAE of all events. The color version of this figure is available only in the electronic edition.

Table 5

Ranks by the Average Event-Based Weighted Mean Absolute Error (wMAE) Score

	HSIT	DYFI	DBMI11 _P	DBMI11 _{RM}	DBMI11 _{RL}
1987GPT	10	10	9	9	10
1996P	<i>4</i>	<i>4</i>	<i>3</i>	<i>4</i>	<i>6</i>
2000P	8	7	8	8	7
2001G	5	6	5	5	3
2004AD	2	<i>1</i>	2	2	<i>4</i>
2006G	6	5	6	6	2
2008PAG _I	7	8	7	7	5
2008PAG _M	3	3	4	3	<i>1</i>
2010SSG	9	9	10	10	9
2012AWW	<i>1</i>	2	<i>1</i>	<i>1</i>	8

Ranks 1–4 are italicized.

consistently ranked either first or second. The 2008PAG_M and 1996P models ranked either third or fourth. The 2001G and 2006G models ranked either fifth or sixth. The 2000P and 2008PAG_I models ranked either seventh or eighth. For the retrospective dataset for large earthquakes (DBMI11_{RL}), there was one major difference in the ranking from that of the other four datasets: the global model (2012AWW) ranked eighth, significantly worse than in the other four datasets. Its position was replaced by the 2006G model, which ranked second in DBMI11_{RL}. The ranks for other IPEs in DBMI11_{RL} were rearranged accordingly, differing from those in the other four datasets by one or two.

Because of the highly consistent rankings among datasets, we here describe only the residual plots of the HSIT dataset (Figs. 3 and 4), whereas the plots for the other datasets are shown in (E) Figures S4–S11 for reference. For the residuals versus I_0 (Fig. 3), plots for models 2012AWW and 2004AD, which ranked the highest, do not show an obvious pattern. Plots for other models generally show an underprediction at small I_0 . Plots for models 1987GT and 2010SSG, which had the lowest rank, show a strong trend of underprediction and overprediction, respectively; we emphasize again that the model 2010SSG is designed for large earthquakes in a volcanic region so are not expected to perform well in the setting of the current study. Residual plots versus R_e (Fig. 4) show various degrees of a positive trend of residual against distance for all models, whereas the trends for the best-ranked models (2012AWW and 2004AD) are the weakest. Because a good model should show no trends on its residual plot, these patterns of the residual plots generally verify the ranking by the wMAE score.

We state again that the 2010SSG model is regional and not expected to perform well in the dataset we used. Results related to this model are shown only for reference. In general, it is practically impossible to empirically evaluate a highly region-specific model like 2010SSG using prospective observations because the restrictive region limits the amount of data that can be collected within the life span of human civilization.

Validity of Test

Data Completeness

Gasperini (2001) suspected that the intensity data at large distances might be incomplete in a sense that reports of small intensities are missing because weak ground shakings are less clearly felt and therefore not completely documented in historical records. In developing the 2001G model, they assumed the far-field data to be incomplete. In an attempt to avoid incomplete data, they excluded data beyond a distance $\tilde{R}_{IV}$, which is the distance where the intensity is expected to be IV. Their data selection scheme was followed by the modelers of the 2006G and 2008PAG models. On the other hand, Albarello and D'Amico (2004) inspected the attenuation values (i.e., $dI = I_0 - I_{obs}$) and found that the skewness of the data did not show a trend with distances, implying that reports of small intensities at large distances were not missing under the assumption of normally distributed dI . Therefore, the 2004AD model did not exclude distant intensity reports like the above three IPEs.

Although the data completeness issue is controversial, it is possible that unfiltered test data could reflect a distorted performance if the model has excluded potentially incomplete data in its data-fitting process. Therefore, we attempted to exclude data beyond $\tilde{R}_{IV}$ to check if this leads to any changes in model performance. Such a restriction, however, severely reduced the amount of data. After this data exclusion, many earthquakes contained fewer than five intensity reports for the prospective datasets (HSIT, DYFI, and DBMI11_P) for some models. The scores computed from such highly trimmed datasets are not representative. On the other hand, the retrospective dataset of DBMI11_{RM} contains a large number of earthquakes and retained a moderate amount of data even after this data exclusion. (E) The event-based wMAE scores for the trimmed DBMI11_{RM} dataset are shown in Figure S12. High ranking models such as 2012AWW and 2004AD retained their high ranks even after the data exclusion. This implies that the potentially incomplete data do not appear to have a predominant effect on the test results.

Appropriateness of the Score

The wMAE is a simple but widely used measure of goodness of fit in various fields. We used it because it is equally applicable to all selected IPEs. Some IPEs predict the uncertainty in addition to the predicted value. In that case, a score that captures the probabilistic nature of the prediction is desirable. The ignorance score (IGN; Roulston and Smith, 2002) has recently been popular in GMPE evaluations (e.g., Delavaud *et al.*, 2009, 2012; Beauval *et al.*, 2012; Edwards and Douglas, 2013) after Scherbaum *et al.* (2009), who named it the average sample log-likelihood difference, first applied it in the context of GMPE evaluation. We tentatively implemented the IGN to verify the result by wMAE. Some selected IPEs that report both a predicted value and uncertainty, either explicitly or implicitly, assume that the prediction

Figure 3. Box plots for residuals versus I_0 for “Hai sentito il terremoto?” (HSIT) data. Each box represents data of the same I_0 value. The box extends from the lower- to upper-quartile values of the data group, with a horizontal line at the median. ⑤ The corresponding plots for other datasets are given in Figures S4–S7, in the electronic supplement to this article. The color version of this figure is available only in the electronic edition.

follows a normal distribution with the predicted value being the mean and the uncertainty being the standard deviation. Models 1987GPT, 1996P, and 2000P do not include uncertainty, and we assigned the average of all other reported uncertainty (0.96) as their standard deviation. The IGN for earthquake i is defined as (see Scherbaum *et al.*, 2009, for mathematical details)

$$\text{IGN}_i = -\frac{1}{N_i} \sum_{j=1}^{N_i} \log_2[f(I_{\text{obs},ij})], \quad (5)$$

in which f is the likelihood of the prediction given the observation. ⑤ The result by IGN is shown in Figure S13. Models 2012AWW and 2004AD remain the highest ranks under both wMAE and IGN, implying robustness of their good performance.

Other scores proposed for GMPE validations (e.g., Scherbaum *et al.*, 2004; Kale and Akkar, 2013; Akkar and Kale, 2014; Mak *et al.*, 2014) are not implemented in this study because they are not as popular as the IGN.

Ordinal Intensities

Traditional macroseismic intensity is defined to be ordinal. The strict mathematical definition of an ordinal set describes only the order, but not the distance, between each member.

This may raise a concern of whether the measure of the mismatch between observed and predicted value, such as the residual and the absolute error, is meaningful because these measures implicitly assume an equal distance between successive intensity levels. We do not consider this issue critical in our validation because of three reasons. First, the model fitting processes of all the selected IPEs are designed for real numbers; and, because the resulting predictions are also real numbers, the intensity values have already been implicitly assumed to be equivalent to real numbers by the modelers. We were simply following this assumption. Second, the two major prospective databases, HSIT and DYFI, report intensity values as pseudocontinuous instead of truly ordinal numbers. The traditional ordinal nature of intensity is not strictly followed by modern web-based databases. Third, the results measured by the wMAE score and residual analyses are consistent among multiple datasets, implying robustness.

A model validation method was proposed by Albarello and D’Amico (2005) to address the issue of the ordinal nature of intensity values. We tentatively implemented this method to check if it provides any additional information on model performance. Readers are referred to the original text for the mathematical details of this method. The core concept of this method is to compare the predicted and observed numbers of occurrence of intensities exceeding a certain

Figure 4. Box plots for residual versus distance for HSIT data. Each box represents data of a 10 km interval. Adjacent data groups overlap by 5 km. The box extends from the lower- to upper-quartile values of the data group, with a horizontal line at the median. ⑤ The corresponding plots for other datasets are given in Figures S8–S11. The color version of this figure is available only in the electronic edition.

value. A significance test can be conducted to check if the two are statistically different. The concept of this method is similar to the N -test used in the Collaboratory for the Study of Earthquake Predictability (Schorlemmer *et al.*, 2007; Zechar *et al.*, 2010) for earthquake forecast evaluation, as well as the counting method proposed by Mucciarelli *et al.* (2008) for evaluating SHAs. Here, we call this method the N -test. Because it requires probabilistic predictions, we used the average uncertainty of other models as the uncertainty of models not reporting uncertainty (1987GPT, 1996P, and 2000P).

⑤ The N -test results for each dataset are shown in Figure S14. In general, all models failed the N -test at intensity values exceeded by a sufficient number of observations. This is not surprising, because Albarello and D'Amico (2005) showed in their test case that the 2004AD model failed even when using exactly the same set of data used in model fitting. Because most models performed poorly under this test, we consider that this test is not informative under the current status of IPE development.

Discussions

Data Quantity

Plentiful data do not necessarily lead to a good model. For example, the 1996P model is derived from significantly

fewer data than both the 2001G and 2006G models but did not perform worse in the present study. Of course, this principle cannot be pushed to the extreme. For example, the 2000P model is updated from the 1996P model (see The Databases section), so they are based on highly overlapping data. The finer zonation of 2000P has led to using only one earthquake in each zone, which does not appear to be conducive to good performance.

Although the data sources for all Italian IPEs after 2000 are highly overlapping, their predictive powers were found to be quite different in the present study. In addition to the total amount of available data, decisions made by the modeler, such as data selection, functional form selection, and optimization method, jointly determine the predictive power of a model. We discuss the issue of functional form in the Functional Form section. In general, however, it may not be possible to isolate the effect of each component and identify the best decision. This again points to the importance of empirical evaluation using prospective data because the apparent superiority of a particular decision does not always lead to a predictive model.

Models 2008PAG_M and 2008PAG_I are based on the same functional form, whereas the relative performances are quite different (see Table 5). It is possible that this is due to the uncertainty in the estimation of I_0 . The presumably more-

accurate estimation of magnitude than of epicentral intensity could be the reason for the better performance of 2008PAG_M than 2008PAG_I. The current study used the reported parameters from each database as it is and assumed they are sufficiently accurate.

Functional Form

The three models that ranked the highest (2004AD, 2012AWW, and 2008PAG_M) are based on physical functional forms (see the [Models](#) section and Table 2), although they differ in details such as how (or whether) to apply a hypothetical depth or an anelastic attenuation term. There has been intensive debate on whether IPEs should adopt forms with physical meanings. [Berardi et al. \(1994\)](#) concluded that the nonphysical cubic root form was the best to describe Italian intensity attenuation. [Gasperini \(2001\)](#) claimed that a bilinear form was the most appropriate based on their analysis. The current study does not support these early findings. The simple empirical functional forms were designed to use the least number of coefficients to avoid overfitting. It remains unknown whether this has led to a failure to capture the general attenuation characteristics, thus limiting each model's predictive power. Although a simpler functional form is generally preferred when various forms fit the data equally well, the current study seems to indicate that a physical functional form provides better predictive power.

Regional Dependence

The importance of the regional difference in ground-motion attenuation among regions sharing the same tectonic feature has long been controversial. Case studies (e.g., [Atkinson and Morrison, 2009](#); [Chiou et al., 2010](#)) suggest that regional differences in ground motions are significant for small earthquakes but diminish for moderate to large earthquakes. In practice, regional data are often combined to enhance data quantity (e.g., [Power et al., 2008](#)). The mechanism of regional differences in ground motion may not apply directly to IPEs. For example, regional differences due to source characteristics presumably do not affect I_0 models. Because of the good performance of the global model 2012AWW, the current study does not support regional differences in IPEs. The use of global data on model development and evaluation is therefore justified. This principle, again, cannot be pushed to the extreme. IPEs developed for a highly specific region are not expected to be applied to other regions, as shown by the poor performance of the regional model of 2010SSG in the current study.

Newer $\neq$ Better

The current study shows that models 2012AWW, 2004AD, and 2008PAG_M consistently score better than other models on the datasets HSIT, DYFI, DBMI11_P, and DBMI11_{RM}. These four datasets contain mainly moderate earthquakes. Models 1996P and 2004AD are not the most recent Italian models, but they outperformed newer models like 2000P, 2001G (for

1996P), and 2006G, 2008PAG (for 2004AD). It is common in the engineering seismology community to develop new models once new techniques or data become available. The presumption that newer models should perform better implicitly follows a straightforward (and deductive) reasoning that improved methods and data should lead to improved results. Our test results show that this reasoning may not be always correct. Empirical justification remains critically important. The counterexamples of models 1996P and 2004AD reiterate the need of investing resources in continuous and prospective model evaluations, in addition to continuously developing new models.

Internet-Based Intensity

The model performances under modern Internet-based datasets (HSIT and DYFI) and traditional datasets (DBMI11_P and DBMI11_{RM}) are consistent. This indirectly implies that data from the two kinds of collection methods are compatible and justifies the use of Internet-based data, which will likely dominate future data, on model developments and evaluations. As their properties are becoming better understood (e.g., [Mak and Schorlemmer, 2015](#)), Internet-based intensity data will likely play a more important role in hazard mitigation in the future.

Model Performance on Large Earthquakes

The relative performance pattern was different in the dataset DBMI11_{RL}, which contains mainly strong earthquakes. We consider the IPE performance on strong earthquakes as inconclusive for two reasons. First, the dataset DBMI_{RL} is retrospective. Evaluations based on retrospective data do not reflect the true predictive power of a model. Second, there is a different use of large intensity values between historical and modern earthquakes. Intensity values $\geq X$ were often assigned for historical earthquakes but rarely used in modern earthquakes. This may explain the poor performance of the model 2012AWW, which is developed using solely strong modern earthquakes. In fact, intensity values predicted by the 2012AWW are capped at about 8.5. In the foreseeable future, the IPE performance for large earthquakes cannot be objectively and empirically evaluated using only Italian earthquake data. This is not a unique shortcoming of the current study or IPE comparisons, but a common problem for all model comparisons for ground-motion models, including GMPEs. Currently, expert judgments seem the only means to evaluate model performance on large earthquakes.

Concluding Remarks

We evaluated the performance of a collection of IPEs developed over the past two decades using both prospective and retrospective intensity observations in Italy. The results show the following:

1. The simple score of mean absolute error is sufficient to represent the general performance of a model. Other

methods verified the major results. More complicated methods may have theoretical advantages and could be more applicable to future generations of models.

2. The 2012AWW, 2004AD, and 2008PAG_M models consistently showed better performance than other IPEs on four independent datasets that consist of mainly moderate earthquakes. Because of the retrospective nature of strong earthquake data and the different use of high-intensity values between historical and modern earthquakes, the model performance on strong earthquakes remains inconclusive.
3. IPEs that performed well in this study use physical functional forms.
4. Both good I_0 (2004AD) and M (2012AWW and 2008PAG_M) models exist. These models are suitable for intensity predictions based on different representations of earthquake source.
5. Both indigenous (2004AD and 2008PAG_M) and global (2012AWW) IPEs performed well. More importantly, the global IPE (2012AWW) is applicable to Italy. This supports (a) grouping data from multiple geographic regions of the same tectonic regime to enlarge the pool of data for model development and evaluation and (b) the application of foreign IPEs when indigenous models are not available.

Data and Resources

Macroseismic intensity data of HSIT, DYFI, and DBMI11 were downloaded from www.haisentitoilteremoto.it, earthquake.usgs.gov/earthquakes/dyfi, and emidius.mi.ingv.it/DBMI11, respectively. Moment tensors for earthquakes in the HSIT database were downloaded from the Time Domain Moment Tensor catalog (cnt.rm.ingv.it/tfmt.html). I_0 values of earthquakes in DBMI04 were downloaded from emidius.mi.ingv.it/DBMI04. Figure 1 was prepared using Generic Mapping Tools (GMT) (gmt.soest.hawaii.edu). All websites were last accessed in March 2013. The details of databases GNDT94 and DOM 4.1 presented in Table 1 are available from emidius.mi.ingv.it/GNDT and emidius.mi.ingv.it/DOM, respectively.

Acknowledgments

The study is developed under the financial support of the Global Earthquake Model Foundation. Comments from Gottfried Grünthal, Paolo Gasperini, and two anonymous reviewers have helped to improve the manuscript. The first author thanks Álvaro González for his help in compiling Italian earthquake data for an early version of this study.

References

- Akkar, S., and Ö. Kale (2014). Reply to "Comment on 'A new procedure for selecting and ranking ground-motion prediction equations (GMPEs): The Euclidean distance-based ranking (EDR) method' by Özkan Kale and Sinan Akkar" by Sum Mak, Robert Alan Clements and Danijel Scholarenmer, *Bull. Seismol. Soc. Am.* **104**, no. 6, 3141–3144.
- Albarelo, D., and V. D'Amico (2004). Attenuation relationship of macroseismic intensity in Italy for probabilistic seismic hazard assessment, *Bollettino di Geofisica Teorica ed Applicata* **45**, no. 4, 271–284.
- Albarelo, D., and V. D'Amico (2005). Validation of intensity attenuation relationships, *Bull. Seismol. Soc. Am.* **95**, no. 2, 719–724.
- Allen, T. I., D. J. Wald, A. J. Hotovec, K. Lin, P. S. Earle, and K. D. Marano (2008). An atlas of ShakeMaps for selected global earthquakes, *U.S. Geol. Surv. Open-File Rept. 2008-1236*, 47 pp.
- Allen, T. I., D. J. Wald, and C. B. Worden (2012). Intensity attenuation for active crustal regions, *J. Seismol.* **16**, 409–433.
- Arango, M. C., F. O. Strasser, J. J. Bommer, J. M. Cepeda, R. Boroschek, D. A. Hernandez, and H. Tavera (2012). An evaluation of the applicability of current ground-motion models to the South and Central American subduction zones, *Bull. Seismol. Soc. Am.* **102**, no. 1, 143–168.
- Atkinson, G. M., and M. Morrison (2009). Observations on regional variability in ground-motion amplitudes for small-to-moderate earthquakes in North America, *Bull. Seismol. Soc. Am.* **99**, no. 4, 2393–2409.
- Beauval, C., H. Tasan, A. Laurendeau, E. Delavaud, F. Cotton, P. Guguén, and N. Kuehn (2012). On the testing of ground-motion prediction equations against small-magnitude data, *Bull. Seismol. Soc. Am.* **102**, no. 5, 1994–2007.
- Berardi, R., C. Petrungaro, L. Zonetti, L. Magri, and M. Mucciarelli (1994). *Mappe di sismicità per l'area Italiana*, ENEL, Rome, Italy, 60 pp.
- Bindi, D., S. Parolai, A. Oth, K. Abdrakhmatov, A. Muraliev, and J. Zschau (2011). Intensity prediction equations for central Asia, *Geophys. J. Int.* **187**, no. 1, 327–337.
- Carletti, F., and P. Gasperini (2003). Lateral variations of seismic intensity attenuation in Italy, *Geophys. J. Int.* **155**, no. 3, 839–856.
- Chiou, B., R. Youngs, N. Abrahamson, and K. Addo (2010). Ground-motion attenuation model for small-to-moderate shallow crustal earthquakes in California and its implications on regionalization of ground-motion prediction models, *Earthq. Spectra* **26**, no. 4, 907–926.
- Cotton, F., F. Scherbaum, J. J. Bommer, and 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**, no. 2, 137–156.
- Cua, G., D. J. Wald, T. I. Allen, D. Garcia, C. B. Worden, M. Gerstenberger, K. Lin, and K. Marano (2010). "Best practices" for using macroseismic intensity and ground motion intensity conversion equations for hazard and loss models in GEM1, *Technical Rept.*, GEM Foundation, Pavia, Italy.
- Delavaud, E., F. Scherbaum, N. Kuehn, and T. Allen (2012). Testing the global applicability of ground-motion prediction equations for active shallow crustal regions, *Bull. Seismol. Soc. Am.* **102**, no. 2, 707–721.
- Delavaud, E., F. Scherbaum, N. Kuehn, and C. Riggelsen (2009). Information-theoretic selection of ground-motion prediction equations for seismic hazard analysis: An applicability study using Californian data, *Bull. Seismol. Soc. Am.* **99**, no. 6, 3248–3263.
- Diebold, F. X., T. A. Gunther, and A. S. Tay (1998). Evaluating density forecasts with applications to financial risk management, *Int. Economical Rev.* **39**, 863–883.
- Douglas, J., and R. Mohais (2009). Comparing predicted and observed ground motions from subduction earthquakes in the Lesser Antilles, *J. Seismol.* **13**, 577–587.
- Edwards, B., and J. Douglas (2013). Selecting ground-motion models developed for induced seismicity in geothermal areas, *Geophys. J. Int.* **195**, no. 2, 1314–1322.
- Eguchi, R. T., J. D. Goltz, H. A. Seligson, P. J. Flores, N. C. Blais, T. H. Heaton, and E. Bortugno (1997). Real-time loss estimation as an emergency response decision support system: The early post-earthquake damage assessment tool (EPEDAT), *Earthq. Spectra* **13**, no. 4, 815–832.
- Faccioli, E., and C. Cauzzi (2006). Macroseismic intensities for seismic scenarios, estimated from instrumentally based correlations, *Proc. of the First European Conference on Earthquake Engineering and Seismology*, Geneva, Switzerland, 3–8 September 2006, 569.
- Gasperini, P. (2001). The attenuation of seismic intensity in Italy: A bilinear shape indicates the dominance of deep phases at epicentral distances longer than 45 km, *Bull. Seismol. Soc. Am.* **91**, no. 4, 826–841.

- Gasparini, P., F. Bernardini, G. Valensise, and E. Boschi (1999). Defining seismogenic sources from historical earthquake felt reports, *Bull. Seismol. Soc. Am.* **89**, no. 1, 94–110.
- Gómez Capera, A. A. (2006). Seismic hazard map for the Italian territory using macroseismic data, *Earth Sci. Res. J.* **10**, no. 2, 6790.
- Grandori, G., F. Perotti, and A. Tagliani (1987). On the attenuation of macroseismic intensity with epicentral distance, A. S. Cakmak (Editor), *Proc. of the Third International Conference Soil Dynamics and Earthquakes Engineering, Ground Motion and Engineering Seismology*, Princeton University, Princeton, New Jersey, 22–24 June 1987, 581–594.
- Grünthal, G. (1998). *European Macroseismic Scale 1998 (EMS-98)*, Vol. 15, Cahiers du Centre Européen de Géodynamique et de Séismologie, Conseil de l'Europe, Luxembourg, 1–99.
- Grünthal, G. (2011). Earthquake, intensity, in *Encyclopedia of Solid Earth Geophysics, Encyclopedia of Earth Sciences*, H. K. Gupta (Editor), Springer, Dordrecht, The Netherlands, 237–242.
- Grünthal, G., A. H. Thieken, J. Schwarz, K. S. Radtke, A. Smolka, and B. Merz (2006). Comparative risk assessments for the city of Cologne—Storms, floods, earthquakes, *Nat. Hazards* **38**, 2144.
- Gupta, I. N., and O. W. Nuttli (1976). Spatial attenuation of intensities for central U.S. earthquakes, *Bull. Seismol. Soc. Am.* **66**, 743–751.
- Jordan, T. (2006). Earthquake predictability, brick by brick, *Seismol. Res. Lett.* **77**, no. 1, 3–6.
- Kaklamanos, J., and L. G. Baise (2011). Model validations and comparisons of the next generation attenuation of ground motions (NGA-west) project, *Bull. Seismol. Soc. Am.* **101**, no. 1, 160–175.
- Kale, Ö., and S. Akkar (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**, no. 2A, 1069–1084.
- Laio, F., and S. Tamea (2007). Verification tools for probabilistic forecasts of continuous hydrological variables, *Hydrol. Earth Syst. Sci.* **11**, 1267–1277.
- Magri, L., M. Mucciarelli, and D. Albarello (1994). Estimates of site seismicity rates using ill-defined macroseismic data, *Pure Appl. Geophys.* **143**, no. 4, 617–632.
- Mak, S., and D. Schorlemmer (2015). What makes people respond to “Did You Feel It?” *Seismol. Res. Lett.* **87**, no. 1, doi: [10.1785/0220150056](https://doi.org/10.1785/0220150056).
- Mak, S., R. A. Clements, and D. Schorlemmer (2014). Comment on “A new procedure for selecting and ranking ground-motion prediction equations (GMPEs): The Euclidean distance-based ranking (EDR) method” by Özkan Kale and Sinan Akkar, *Bull. Seismol. Soc. Am.* **104**, no. 6, 3139–3140.
- Mucciarelli, M., D. Albarello, and V. D’Amico (2008). Comparison of probabilistic seismic hazard estimates in Italy, *Bull. Seismol. Soc. Am.* **98**, no. 6, 2652–2664.
- Musson, R. M. W. (2000). Intensity-based seismic risk assessment, *Soil Dynam. Earthq. Eng.* **20**, 353–360.
- Musson, R. M. W., G. Grünthal, and M. Stucchi (2010). The comparison of macroseismic intensity scales, *J. Seismol.* **14**, 413–428.
- Pasolini, C., D. Albarello, P. Gasparini, V. D’Amico, and B. Lolli (2008). The attenuation of seismic intensity in Italy, part II: Modeling and validation, *Bull. Seismol. Soc. Am.* **98**, 692–708.
- Peruzza, L. (1996). Attenuating intensities, *Ann. Geofisc.* **39**, no. 5, 1079–1093.
- Peruzza, L. (2000). Macroseismic attenuation relationships of Italian earthquakes for seismic hazard assessment purposes, *Bollettino di Geofisica Teorica ed Applicata* **41**, no. 1, 31–48.
- Power, M., B. Chiou, N. Abrahamson, Y. Bozorgnia, T. Shantz, and C. Roblee (2008). An overview of the NGA project, *Earthq. Spectra* **24**, no. 1, 3–21.
- Rotondi, R., and G. Zonno (2004). Bayesian analysis of a probability distribution for local intensity attenuation, *Ann. Geophys.* **47**, no. 5, 1521–1540.
- Roulston, M. S., and L. A. Smith (2002). Evaluating probabilistic forecasts using information theory, *Mon. Weather Rev.* **130**, 1653–1660.
- Sbarra, P., P. Tosi, and V. De Rubeis (2010). Web-based macroseismic survey in Italy: Method validation and results, *Nat. Hazards* **54**, 563–581.
- Scasserra, G., J. P. Stewart, P. Bazzurro, G. Lanzo, and F. Mollaioli (2009). A comparison of NGA ground-motion prediction equations to Italian data, *Bull. Seismol. Soc. Am.* **99**, no. 5, 2961–2978.
- Scherbaum, F., F. Cotton, and P. Smit (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**, no. 6, 2164–2185.
- Scherbaum, F., E. Delavaud, and C. Riggelsen (2009). Model selection in seismic hazard analysis: An information-theoretic perspective, *Bull. Seismol. Soc. Am.* **99**, no. 6, 3234–3247.
- Schorlemmer, D., and M. Gerstenberger (2007). RELM testing center, *Seismol. Res. Lett.* **78**, no. 1, 30–36.
- Schorlemmer, D., M. Gerstenberger, S. Wiemer, D. D. Jackson, and D. A. Rhoades (2007). Earthquake likelihood model testing, *Seismol. Res. Lett.* **78**, no. 1, 17–29.
- Sieberg, A. (1923). *Geologische, physikalische und angewandte Erdbebenkunde*, G. Fischer, Jena, Germany.
- Slejko, D., L. Peruzza, and A. Rebez (1998). Seismic hazard maps of Italy, *Ann. Geofisc.* **41**, no. 2, 183–214.
- Sørensen, M. B., D. Stromeyer, and Grünthal (2010). Intensity attenuation in the Campania region, southern Italy, *J. Seismol.* **14**, 209–223.
- Stafford, P. J., F. O. Strasser, and J. J. Bommer (2008). An evaluation of the applicability of the NGA models to ground-motion prediction in the Euro-Mediterranean region, *Bull. Earthq. Eng.* **6**, 149–177.
- Stewart, J. P. (2010). Analysis of regional ground motion variations for engineering application, Chapter 3, in *Earthquake Engineering in Europe*, M. Garevski and A. Ansal (Editors), *Geotechnical, Geological and Earthquake Engineering*, Vol. 17, Springer, The Netherlands, 67–75.
- Vilanova, S. P., J. F. B. D. Fonseca, and C. S. Oliveira (2012). Ground-motion models for seismic-hazard assessment in western Iberia: Constraints from instrumental data and intensity observations, *Bull. Seismol. Soc. Am.* **102**, no. 1, 169–184.
- von Kövesligethy, R. (1906). Seismonomia, *Bollettino della Società Sismologica Italiana* **XI**, 113–250.
- Wald, D. J., V. Quitoriano, L. Dengler, and J. Dewey (1999). Utilization of the Internet for rapid community intensity maps, *Seismol. Res. Lett.* **70**, 680–697.
- Wald, D. J., V. Quitoriano, T. H. Heaton, H. Kanamori, C. W. Scrivner, and B. C. Worden (1999). Trinet “ShakeMaps”: Rapid generation of peak ground-motion and intensity maps for earthquakes in southern California, *Earthq. Spectra* **15**, no. 3, 537–556.
- Wyss, M. (2008). Estimated human losses in future earthquakes in central Myanmar, *Seismol. Res. Lett.* **79**, no. 4, 520–525.
- Zechar, J. D., M. C. Gerstenberger, and D. A. Rhoades (2010). Likelihood-based tests for evaluating space-rate-magnitude earthquake forecasts, *Bull. Seismol. Soc. Am.* **100**, no. 3, 1184–1195.

Helmholtz-Zentrum Potsdam Deutsches GeoForschungsZentrum GFZ
Helmholtzstraße 6
14467 Potsdam, Germany
smak@gfz-potsdam.de