Earthquakes Parameters from Citizen Testimonies: A Retrospective Analysis of EMSC Database

Gianfranco Vannucci*[1], Paolo Gasperini[2], Laura Gulia[2], and Barbara Lolli[1]

Abstract

We aim to compute macroseismic parameters (location and magnitude) using the BOXER code for the first time on the citizen testimonies, that is, individual intensity data points (IDPs) at the global scale collected and made available by the LastQuake system of the European–Mediterranean Seismological Centre (EMSC). IDPs available for different earthquakes are selected to eliminate those that are geographically inconsistent with most data; then they are clustered spatially based on various methods. For each cluster with at least three IDPs, a macroseismic data point (MDP), corresponding to an intensity value assessed for given localities as in classical macroseismic studies, is computed by various central tendency estimators (average, median, and trimmed averages). Finally, macroseismic parameters are obtained by MDP distribution using two location methods of BOXER code. For each earthquake, we used raw and corrected intensities and 132 different combinations of grouping methods, estimators, and BOXER methods. We assigned a ranking to the combinations that best reproduce instrumental parameters and used such a ranking to select preferred combinations for each earthquake. We analyzed retrospectively the reliability of the parameters as a function of time and space. The results are essentially identical using original and corrected intensities and show higher reliability for BOXER's method 1 than for method 0; they are dependent on the geographical area, and generally improve over time and with the number of IDPs collected. These findings are useful for the future real-time analyses, and for evaluating the location and magnitude of earthquakes whenever a sufficient number of IDPs are available and with a distribution such that MDPs can be derived and the BOXER method applied.

Cite this article as Vannucci, G., P. Gasperini, L. Gulia, and B. Lolli (2023). Earthquakes Parameters from Citizen Testimonies: A Retrospective Analysis of EMSC Database, *Seismol. Res. Lett.* **XX**, 1–28, doi: 10.1785/0220230245.

Supplemental Material

Introduction

The macroseismic intensity, that is, the quantification of the severity of the ground motion, based on earthquake effects on humans, objects, natural environment, and buildings, is a tool for studying preinstrumental earthquakes used for seismic hazard assessment and seismic risk mitigation. The intensity assessed by macroseismic experts or other methods (e.g., Vannucci *et al.*, 2015) through macroseismic scales (e.g., Mercalli–Cancani–Sieberg [MCS]; Sieberg, 1912, 1932—European Macroseismic Scale [EMS98], Grünthal, 1998) is quantified using a damage scenario at the scale of localities, and their geographic distribution allows to assess reliable epicenter location and magnitude using various software codes (e.g., Bakun and Wentworth, 1997; Gasperini *et al.*, 1999, 2010; Pettenati and Sirovich, 2003; Musson and Jiménez, 2008). Gasperini *et al.* (2010) have shown how macroseismic intensities make it possible to calculate location, magnitude and, in the most favorable cases (e.g., earthquakes with magnitude ≥5.7), also the orientation of the source, with an accuracy comparable with the instrumental methods. Vannucci *et al.* (2019) also demonstrated that if the intensities are well distributed and quickly available after the occurrence of the earthquake, they can constrain well the macroseismic source, and provide useful information to civil protection and stakeholders even before reliable instrumental data be available. Therefore, the macroseismic intensities do not only provide information on preinstrumental earthquakes, but also on contemporary ones by taking advantage of the geographic abundance of information coming from different localities that are much denser than the instrumental stations. Such data also provide a direct check of the theoretical models of energy

1. Istituto Nazionale di Geofisica e Vulcanologia, Sezione di Bologna, Bologna, Italy, https://orcid.org/0000-0003-0918-0784 (GV); https://orcid.org/0000-0003-4186-9055 (BL); 2. Dipartimento di Fisica e Astronomia, Università di Bologna, Bologna, Italy, https://orcid.org/0000-0002-5314-0563 (PG); https://orcid.org/0000-0002-1474-230X (LG)

*Corresponding author: gianfranco.vannucci@ingv.it

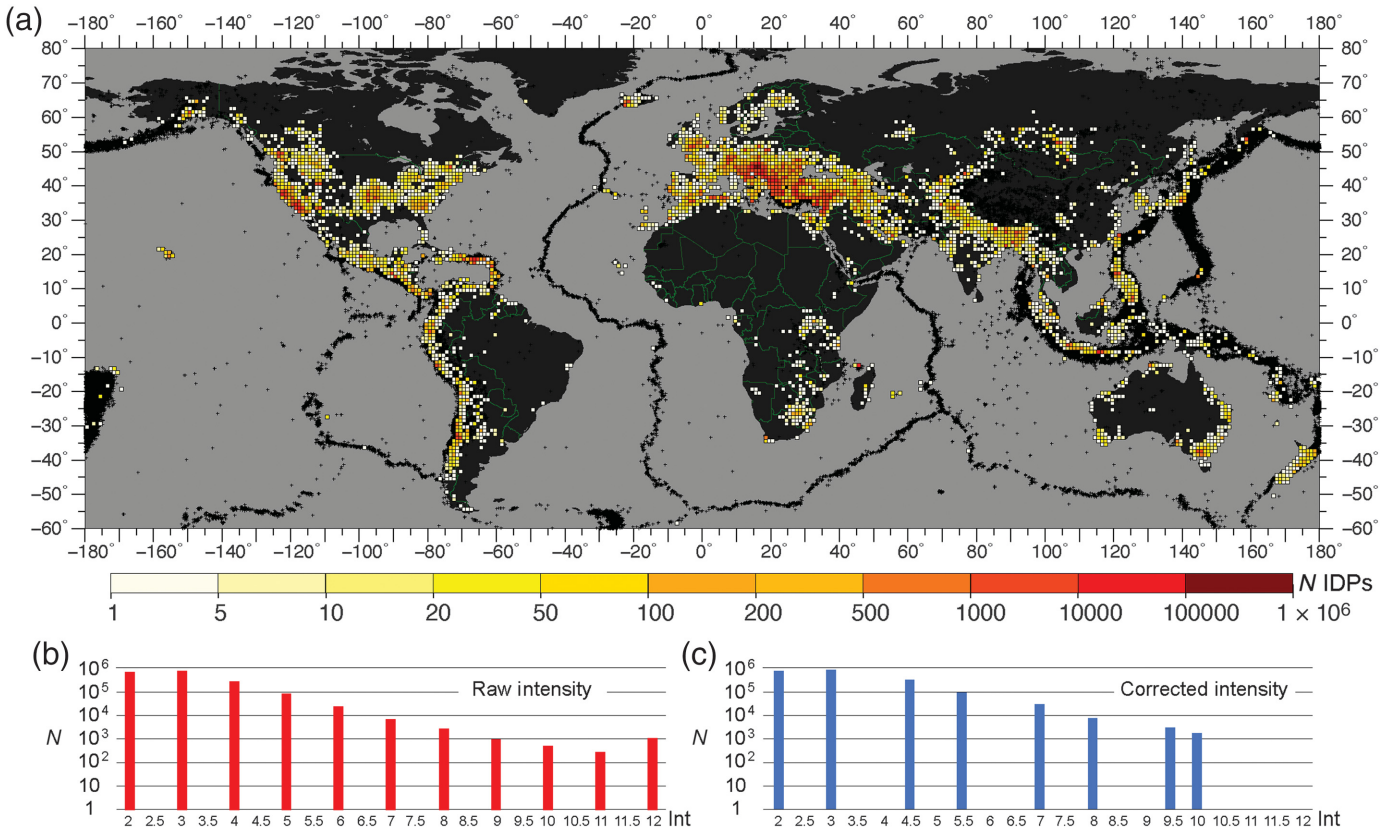


Figure 1. (a) In colors, numbers of individual intensity data points (IDPs) on a regular grid with mesh of 1° both in latitude and longitude. In black, seismicity from the revised catalogue of International Seismological Centre (ISC, 2022), with M > 3 in the time span 2013–2020. (b,c) Frequency distribution over intensity bin of 0.5° of IDPs of the European–Mediterranean Seismological Centre (EMSC) database. Raw intensities and corrected ones (Bossu *et al.*, 2017) in red and blue colors, respectively.

propagation (like SHAKEMAP, see Data and Resources section) for local calibration of expected effects.

Presently, the development of specific software applications allows to collect and elaborate testimonies of the shaking felt by individual citizens. Indeed, since several years, community intensities are collected by different agencies, for example, "Did you feel it?" (DYFI; Wald *et al.*, 1999, 2011, Dewey *et al.*, 2000), of the U.S. Geological Survey (USGS), "Hai sentito il terremoto?" (HSIT; Tosi *et al.*, 2015) of the Istituto Nazionale di Geofisica e Vulcanologia (INGV), New Zealand GeoNet questionnaires (GeoNet, Goded *et al.*, 2018), LastQuake system (Bossu *et al.*, 2015, 2018) of the European–Mediterranean Seismological Centre (EMSC). These data are collected at different spatial scales and with different methodologies. In particular, individual intensity data points (IDPs), that is, macroseismic intensity according to the EMS98 (Grünthal, 1998) and assessed by each eyewitness citizen, are collected and made available by LastQuake system. The IDP database is based on a worldwide community of people, number for which increases over time. Our aim is to use this basic information to develop methods to compute the location and the magnitude of the earthquake.

Through the LastQuake system, EMSC collected 1,874,376 IDPs (with intensity ≥ 2) of 51,359 global earthquakes (with magnitude ranging between 0.4 and 8.4) from 2012 to February 2023 (Fig. 1). Such data are freely available at EMSC website (see Data and Resources section). The number of collected IDPs generally increased over time (Fig. S1, available in the supplemental material to this article), because the popularity of the application increased, and the users became more and more involved in such activity (Bossu *et al.*, 2017). Each collected IDP provides latitude, longitude, raw (R), and corrected (i.e., revaluated) intensity (C). The raw intensity is assessed through the selection by each citizen (i.e., observer) of thumbnails that best represent the observed seismic effects, that is, by the correspondence between eyewitness observations and felt scenario representations, whereas the corrected intensity is computed, according to Bossu *et al.* (2017), to best reproduce DYFI intensities for a reference dataset of 17 earthquakes.

The number of collected intensities is a decreasing function of their value: the higher the intensity, the lower the number of reports, because higher intensities are generally limited to the areas close to the source (near field), whereas lower intensities occur at longer geographical distances (far field), with larger numbers of people and reports. This trend is generally valid for raw intensities, except for the extreme intensities 2 and 12 (Fig. 1).

Based on the geometric spreading of seismic energy, the effects of the earthquake should "ideally" propagate in any direction from the epicenter. However, cities and citizens are not evenly distributed throughout the territory, and the distribution of IDPs suffers sometimes from the lack of coverage in uninhabited areas. In general, the greater the earthquake magnitude the wider the area of effects and the higher the number of felt reports, but both the number and the distribution of IDPs are subject to a number of factors: geomorphological ones (presence of seas, lakes, mountains, and deserts), demographic ones (variable population density and presence or absence of cities), technological ones (internet coverage), and political ones (free or equitable access to internet, e.g., Hough and Martin, 2021).

The lack of IDPs in the epicentral area for earthquakes of strong magnitude and destructive effects could even be due just to the strength of such effects (e.g., destruction of buildings, infrastructures, and casualties) that might prevent the people to pay attention to the reports, so leaving the epicentral zone empty ("doughnut effect"; Bossu *et al.*, 2018). The IDPs may be absent where restrictive policies on the use of smartphone applications are in force (e.g., in China, North Korea etc., see Fig. 1). Hence, in some regions of the world where earthquakes are known to occur but where there are only a few IDPs (Fig. 1), the distribution of IDPs can be uneven: IDPs are not well distributed around the epicenter so that the maximum azimuthal gap of IDPs with respect to the epicenter is larger than 180°.

Another factor to consider is the presence of some IDPs that are inconsistent with the distribution of the most of other ones. These anomalous IDPs can be divided into two types: "intensity outliers" and "geographic outliers."

Intensity outliers are IDPs for which the assigned intensity values appear significantly inconsistent with respect to the other ones in the neighbor. They can be due to (1) the wrong judgment by the citizen who has emphasized the effects for emotional reasons or misjudgment, in most cases overestimating the macroseismic intensity; and (2) misreporting of intensity with the selection of the last of the available thumbnails in LastQuake system, which might also explain the unreasonably high frequency observed for the degree 12 of the raw intensity (Fig. 1).

Geographic outliers are IDPs located in areas far from the instrumental epicenter. They could be due to various reasons: (1) reports sent from a computer (not a smartphone) for which there is a wrong reporting of the geographical position due to the link to fixed network servers located up to tens of kilometers away from the observing site; (2) use of Virtual Private Network with geolocation up to thousands of kilometers away; (3) persons reporting an earthquake and its intensity on behalf of others, so associating the information with the georeferenced location of the reporting smartphone; (4) bad association between felt report and event; and (5) shocks due to other causes (e.g., quarry blasts). Geographic outliers, if present, are generally a small fraction of the total number of IDPs. Their presence in most cases enlarges the area covered by testimonies. During periods of intense seismicity (seismic sequences), the IDPs can be erroneously attributed to another shock that occurred almost simultaneously but located at long distance, generating intensity and geographical outliers. However, only very few earthquakes show a totally inconsistent association between instrumental epicenter and location of IDPs, that is, IDPs are located too far from the epicenter and cannot represent its effects, making these earthquakes unreliable and unusable for further analyses. The number of these unreliable epicenter–IDPs associations decreased in the course of time, probably owing to the increasing consciousness of people submitting their reports and to a more careful use of the application by the users.

Crowdsourcing projects such as DYFI (Wald *et al.*, 2011) and HSIT (Tosi *et al.*, 2015) collect intensities by citizens based on written questionnaires, and have already approached the problem of outliers by grouping single reports and deriving intensities at geographical localities as commonly done in standard macroseismic surveys. Geographical outliers can be detected on the basis of empirical magnitude–distance relationships, evidencing intensities at anomalous distances from the epicenter, whereas possible intensity outliers can be filtered out from the felt scenario by imposing an intensity threshold (e.g., <11, as in Bossu *et al.*, 2017). This can be justified by considering that the assessment of very heavy damage or destruction by citizens involved in them is unlikely, because they usually do not pay attention to sending smartphone reports while they are in danger of life. Following this approach, EMSC always consider degrees 11 and 12 as outliers, and provides corrected intensities by eliminating these values. However, the remaining high intensities (e.g., 8, 9, and 10), which define very severe and general damage, may also be unreliable, thus representing anomalous intensity values anyway.

Both DYFI and HSIT join individual IDPs using ZIP codes and municipal territories to obtain the MDPs, but, this is not possible for LastQuake data, because they are provided at a global scale in which such geographical subdivisions are not available.

Method: From IDP to Macroseismic Parameters

The distribution of IDPs provides a reasonable indication at a glance of the area of the effects and of the possible epicenter location. To compute the earthquake parameters such as location and magnitude, we use the BOXER code (Gasperini *et al.*, 1999, 2010)—a software widely used for macroseismic analysis for present (e.g., Vannucci *et al.*, 2019) and past earthquakes (e.g., Rovida *et al.*, 2020). However, the use of single IDP is too sensitive to the presence of outliers, and then it is preferable to use instead of macroseismic data points (MDPs), that is, intensities assigned to clusters of IDPs. We adopt the term MDP in analogy to an intensity value assessed for given localities as in classical macroseismic studies, although it has a different origin. Therefore, the quantitative computation of macroseismic

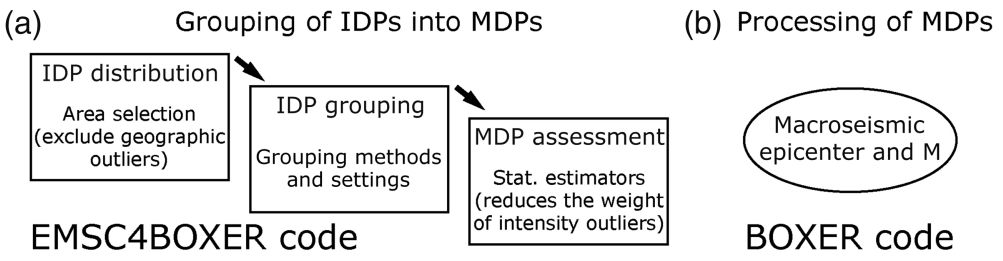


Figure 2. Procedure used from IDPs to assessment of macroseismic parameters. For details concerning the two steps (a and b) see text in the Method: From IDP to Macroseismic Parameters section.

parameters follows two main steps (Fig. 2): first, the grouping of IDPs and the assessment of intensity on MDPs and, second, the processing of MDPs to compute location and magnitude of the earthquake by BOXER.

Starting from the IDP distribution, we use an original code to constrain the area within which to select IDPs and outside which to eliminate geographic outliers. The IDPs are grouped using different grouping methods. If the number of IDPs in each cluster is larger than a given minimum threshold (e.g., 3 and 5), the MDP intensities are computed by various statistical estimators of central tendency as, for example, the average, the median, and the trimmed mean, so reducing the effects of intensity outliers. All MDP intensities are finally processed by the BOXER code to obtain macroseismic parameters and their uncertainties (Fig. 2).

In this retrospective analysis of the EMSC database we clustered IDPs to derive MDPs both using raw and corrected intensities.

We discarded all intensities 2 and 12, thus reducing the range of the raw intensities to the interval from 3 to 11. This is because, even for true intensity estimates made by macroseismic experts, intensity 2 corresponds to a very weak perception of ground shaking (felt by very few people in particularly receptive conditions indoors) that it might remain unobserved in most cases, and it is also difficult to be distinguished from degree 3 (felt by few people indoors). For example, Bakun and Wentworth (1997), for their location and sizing method, choose to aggregate the degree 2 with degree 3, whereas Tosi *et al.* (2015) considered degree 2 equivalent to "not-felt" (degree 1) for HSIT data. Therefore, we preferred to simply discard degree 2, even considering the lower reliability of our intensities based on citizen testimonies. On the other hand, true intensities 12 were really never observed.

Classification of EMSC events

To select IDPs useful for computations and statistical retrospective analyses, we must first eliminate possible geographical outliers. In this retrospective analysis, we use the known instrumental epicenter and magnitude to constrain the geographic area of IDP coverage by the maximum distance prediction equation (MDPE; Fig. 3)—an empirical function aimed at discarding only the furthest geographical outliers, that links the magnitude of an event with the maximum distance of IDPs with respect to epicenter:

$$\mathrm{MDPE} = a + b \times M + \exp(c \times M), \quad (1)$$

in which M is the magnitude; and $a = 50$, $b = 70$, and $c = 0.9$ are fixed coefficients defined empirically by a trial-and-error procedure. The purpose is a quick preliminary selection of IDPs, deleting those located at distances longer than that predicted by the MDPE (Fig. 3) to significantly reduce the time required to assess the MDPs, considering the retrospective analysis of thousands of earthquakes of the EMSC dataset. The maxima and minima of the latitudes and longitudes of the IDPs within the MDPE radius defines a rectangular area for next elaborations (black solid lines in Fig. 3). For the sake of clarity, the application of the MDPE equation (1) and the filtering of geographical outliers is only done for this retrospective analysis, whereas the procedure requires no filter and knowledge of location and instrumental magnitude for event-by-event future near-real-time analyses (see Appendix A for more details).

We classify the earthquakes considering if:

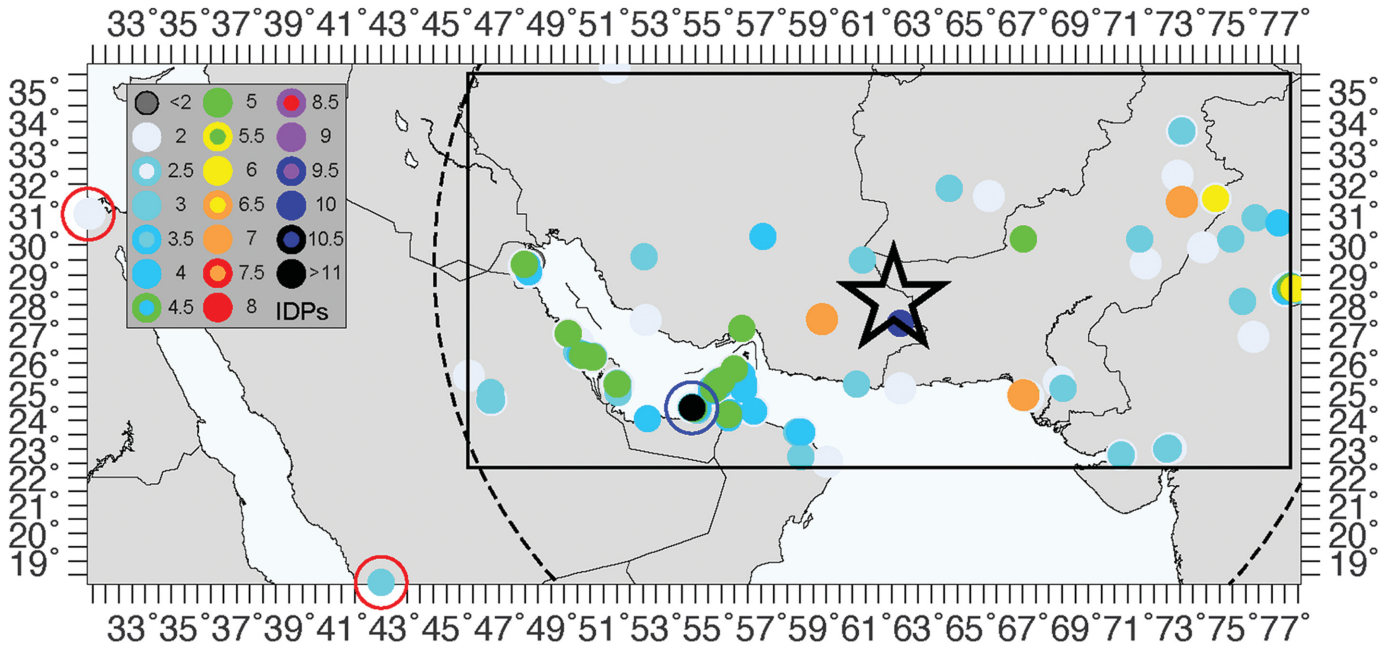
1. the epicenter is located inland or offshore;
2. the epicenter is located in or out the defined rectangular area; and
3. there are geographic outliers. Consequently, we assign a two-character code: the first one indicating whether the epicenter is inland or offshore (L or S, respectively), and the second one is:
   - if the epicenter is inside the area, without outliers;
   - if the epicenter is inside the area, with outliers;
   - if the epicenter is outside the area, without outliers; and
   - if the epicenter is outside the area, with outliers.

We provide a scheme of the eight main categories in Figure 4, identified by various codes (e.g., $L1$, $L4$, $S3$, …); some real examples of earthquakes classified following the previous scheme are plotted in Figure S2.

From IDPs to MDPs: data clustering

This procedure (see details in Appendix A) is structured in three steps (Fig. 5):

Step 1: definition of spatial areas or clusters where grouping IDPs;



Step 2: evaluation of the occurrence of IDPs in each spatial area or cluster; and
Step 3: assessment of MDPs.

Step 1 ("IDPs in/out" in Fig. 5): IDPs available for each earthquake can be clustered using different methods: within a given radius (RA), over a square grid (SQ), over a hexagonal grid (HE), within a radius, and over a square grid (RS, i.e., RA + SQ), within a radius and over a hexagonal grid (RH, i.e., RA + HE), or by DBSCAN (DB) method (see Appendix A for details of each of such methods). For the first five methods, fixed geometries are used to constrain the clustering areas, whereas the shape and the size of clustering areas can vary with the distribution of the data for DB method. We use a partitioning approach in which each IDP is assigned to only one cluster and cannot be shared by more clusters as in "hierarchical clustering" method (e.g., Amorese *et al.*, 2015).

Step 2 ("Occurrence" in Fig. 5): each cluster of IDPs collects intensities. The minimum number of IDPs to calculate a MDP intensity in an area/cluster could be taken in analogy with agencies that collect and provide "crowdsourced" intensities: 5, like HSIT (Tosi *et al.*, 2015) and 3, like DYFI (Wald *et al.*, 2011). Areas or clusters with a number of IDPs lower than the threshold are not evaluated, and the MDPs are not assessed. After several tests with different thresholds, we decided to use three IDPs (as done by DYFI).

Step 3 ("MDPs" in Fig. 5): On the IDPs in each area/cluster, we apply various statistical estimators of central tendency to derive both location (geographical coordinates) and the final MDP intensity of each cluster. We use the average (mnsa), the median (mdna), and the trimmed mean with four different intervals of the distribution of the sorted intensity values:

Figure 3. Example of geographical (circled in red) and intensity outliers (with raw intensity = 12, circled in blue), for the 16 April 2013 10:44 M 7.8 earthquake, number of IDPs: 408. The black dashed circle indicates the maximum distance prediction equation (MDPE) used to delete the farthest IDPs, and define the area (solid black line) of the minimum and the maximum latitude and longitude of selected IDPs. The black star indicates the instrumental epicenter.

10%–90% (mn10), 15%–85% (mn15), 20%–80% (mn20), and 25%–75% (mn25). Trimmed means are computed only if the tails of the distributions have at least one IDP, otherwise the simple average is used. The use of central tendency estimators reduces the effects of intensity outliers, because these are averaged with other IDPs in the clustering area and do not influence the final MDP intensity assessment too much. The approach followed is more conservative compared with HSIT and DYFI by preserving the intensities assessed by citizens as much as possible.

Available MDPs are therefore the results of the combination of grouping and central tendency methods using both raw (R) and corrected (C) intensities. Consequently, even the computed MDPs are hereinafter and analogously indicated as raw or corrected.

To calculate MDP, we used the minimum threshold of three IDPs, deleting geographical outliers and using the intensity in the range 3°–11° (3–10 for corrected intensities). Hence, the initial number of 51,359 earthquakes in the EMSC dataset reduces to 22,761 (Table S1). The selected earthquakes' instrumental epicenter which are located inland are about two-third of the total, covering a wide range of magnitudes. It is important to note that a threshold of five IDPs would immediately eliminate further 4291 earthquakes, and that only 2603 earthquakes have more

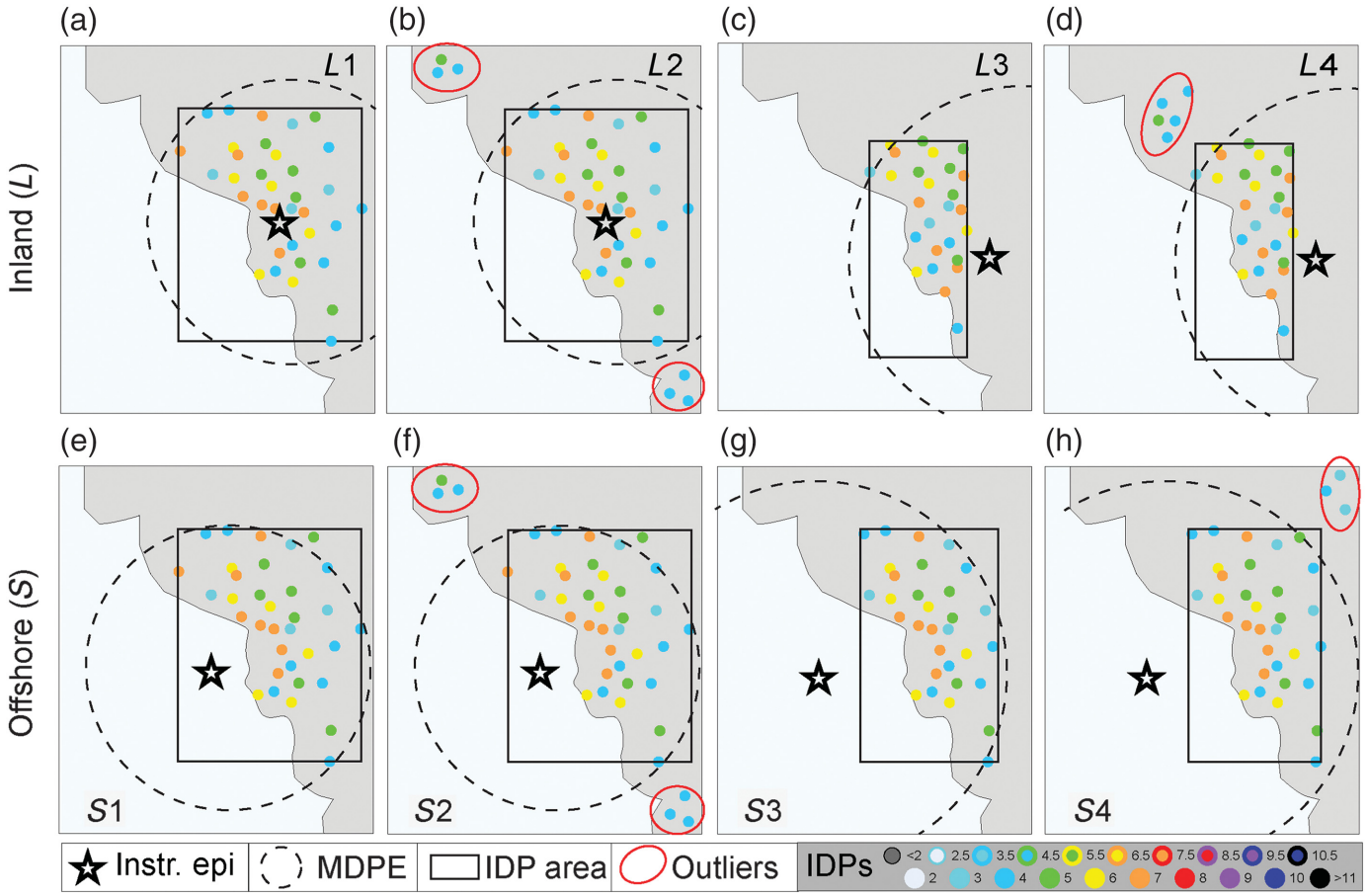


Figure 4. Scheme of classification of the distribution of IDPs. The star indicates the instrumental epicenter, the circular dashed black line is the MDPE radius, and the rectangular black line delimits the area of location of usable IDPs, that is, the minimum and the maximum latitude and longitude of usable IDPs, without any geographic outliers (circled in red color). Panels (a–h) refer to a combination of the classification in the same panel: instrumental epicentres located in land (L) or offshore (S) and presence and location of geographic outliers (from 1 to 4).

than 100 IDPs, whereas 20,159 earthquakes have IDPs ranging between 3 and 100 (panel B of Table S1).

From MDPs to macroseismic parameters

The BOXER code (Gasperini *et al.*, 1999, 2010) calculates macroseismic parameters, such as epicenter, magnitude, and their uncertainties using available MDPs. Among the different computation methods available in the code, we use only the n. 0 and n. 1, hereinafter indicated as BOXER-0 and BOXER-1 (or Bx-0 and Bx-1), respectively. Method 0 computes the epicenter as the barycenter of the sites with the most severe effects. Method 1 computes the center of the entire intensity distribution by a minimization of squared residuals of an attenuation function (Pasolini *et al.*, 2008). BOXER-0 can locate even the earthquakes with only one MDP, whereas BOXER-1 needs more than one MDP (we set the minimum of five MDPs) with the obvious consequence of reducing the total number of events for which macroseismic parameters can computed. However, the latter method allows to assessing the epicenter also for earthquakes located offshore or in uninhabited areas in most favorable cases. Macroseismic magnitude can also be estimated by different methods, depending on the number and the distribution of MDPs. The classical method described in Gasperini *et al.* (1999) uses both the epicentral intensity I_0 and the average distances R_I of various classes of intensities I. However, because the I_0 computed by questionnaire data is usually unreliable, we modified the original algorithm to only use the R_I. In any case, at least four MDPs are required (two intensity classes with two MDPs each) to compute a magnitude. The alternative methods described by Gasperini *et al.* (2010), based on a linear relation between I_0 and M, cannot be used in the present work for the poor reliability of I_0, as well as the new method described in Gasperini *et al.* (2010), because it was found to systematically underestimate the magnitudes.

Results and Discussion

In Table 1, we show the distribution of the number of earthquakes as a function of the number of MDPs using both raw and corrected intensities. For ~7600 of the 22,761 initial earthquakes, we do not even have a single MDP. Consequently, the earthquakes with at least one MDP for which we can provide the location are ~15,000 (Table 1). Hence, only two-third of

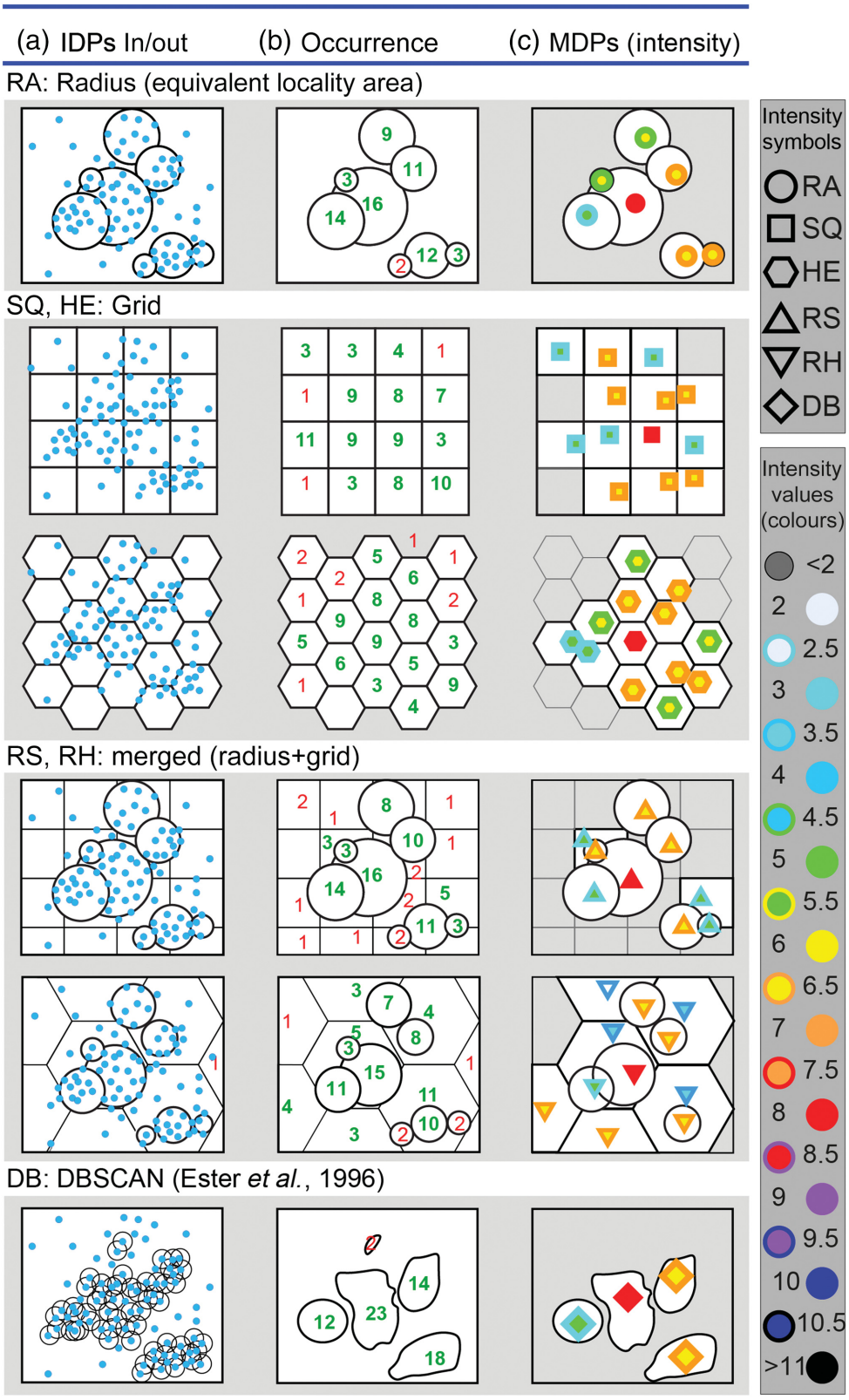


Figure 5. Methods of clustering of IDPs into MDPs through three steps: (a) IDPs available are grouped (or not) following the various methods; (b) for each area of grouping the occurrence of a sufficient number of IDPs is assessed (numbers in green) or not (numbers in red); and (c) IDPs are used to compute a combined intensity (MDPs), indicated with different colors and symbols, for selected area or clusters (in white colors) and using different central tendency estimators.

the events can be compared with instrumental locations to quantify the ability of BOXER code to provide reliable macroseismic parameters.

For each earthquake of the dataset with at least three IDPs (22,761 earthquakes), we combined 11 different grouping methods and six different central tendency estimators to assess MDPs (Appendix A and Table A1). Moreover, macroseismic locations and magnitudes are computed by two methods (BOXER-0 and BOXER-1). Hence, in total, we have 132 different alternative combinations of methods to test. The minimum threshold of three IDPs to locate an earthquake is the minimum but not a sufficient condition, because it is necessary that they all belong to the same clustering area to have a single MDP.

The comparison between macroseismic epicenters and magnitudes with instrumental data provides an estimate of the reliability of the computed parameters for different combinations both using raw and corrected intensities. Instrumental locations and magnitudes of each earthquake are taken from the EMSC webservice. In particular, magnitudes are homogenized to M_w using empirical formulas at the global scale of Lolli *et al.* (2014). For each earthquake, it is possible to evaluate the combinations of methods that separately minimize the distance between macroseismic and instrumental epicenter and the difference between the macroseismic and instrumental magnitude, but they are generally different for different earthquakes.

However, we can establish a ranking of combinations by

TABLE 1
Number of Earthquakes (n. Eqks) as a Function of the Number of Macroseismic Data Points (n. MDPs)

	n. Eqks	
n. MDPs	**Raw Intensities**	**Corrected Intensities**
0	7,658	7,661
1	4,940	4,940
2	2,550	2,551
3	1,473	1,477
4	1,003	998
5	742	742
6	516	515
7	424	426
8	369	367
9	279	280
10	256	255
11–15	763	762
16–20	430	430
21–30	439	440
31–50	380	378
51–75	207	207
76–100	104	104
101–150	101	101
151–200	46	47
201–500	66	65
501–1,000	12	12
1,001–2,000	3	3
All	22,758	22,758
≥ 1	**15,103**	**15,100**
≥ 3	**7,613**	**7,609**
≥ 5	**5,137**	**5,134**

The last three rows (in bold) show the cumulative number of earthquakes with MDP numbers ≥1, ≥3, and ≥5.

counting the number of earthquakes for which each combination best reproduces the instrumental parameters. To objectively compile such ranking, we consider datasets of earthquakes for which both the epicenter and the magnitude can be computed using all combinations. Such datasets include 1144 and 1082 earthquakes for raw and corrected intensities, respectively.

For each earthquake, we assign a score 3 to the combination of methods having, separately, the minimum epicentral distance and the minimum absolute magnitude difference, a score of 1 to all combinations with distances and differences within 5% of the minimum ones and no greater than 1 km and 0.2 m.u., and a score of 0 for all the other cases. We used such nonparametric approach (instead of, for example, the total root mean square error), because we are unsure that macroseismic locations and magnitudes are normally distributed, even considering the possible presence of intensity outliers in some earthquakes.

Such scores are reported in Table 2 for raw intensities (and in Table S2 for corrected intensities). Neither for localization distances nor for differences in magnitude, there is a combination that clearly overperforms all the other ones and which we can choose as the "preferred" one to use prospectively.

The best-performing combinations are different for epicentral location and magnitude, and for raw and corrected intensities. For epicentral location from raw intensities (Tables 2 and 3), the first 43 combinations in the ranking use BOXER method 1 and the first five the grouping method DB2. For corrected intensities (Tables S2 and S3), the first 27 use BOXER methods 1, and four of the first five use the grouping method DB2.

The results are less coherent for magnitude estimation. Using the raw intensities (Tables 2 and 4), in the highest rankings, we have an alternation of both BOXER methods and different grouping methods with a certain prevalence of BOXER-1 and grouping methods RA and DB. Using corrected intensities (Tables S2 and S4), the first 11 combinations use BOXER-1, whereas the preferred grouping methods vary from DB to RA and RH.

The better agreement of BOXER-1 with respect to BOXER-0, concerning the distance from the instrumental epicenter and the good performance of the grouping method DB2, can be immediately evidenced by plotting (Fig. 6) the values of Table 2 and Table S2 for the raw and corrected intensities, respectively: the greater the distance of each combination from the center of each Radar plot, the higher the score obtained by the combination. We also observe a prevalence of the median as central tendency estimator that minimizes the difference with the instrumental data for various grouping and BOXER combinations. About the difference in magnitude, the values are similar to each other, showing the lowest values for the DB2 grouping method; but there is not a clear prevalence of one BOXER or central tendency estimator method with respect to the others.

In general, not all earthquakes can be located and sized by the best performing combination; hence, even combinations other than the "top" ranking one must be used to determining the parameters for as many earthquakes as possible. To verify which combinations are mostly useful, we compute epicenters and magnitudes in our complete datasets of 22,761 earthquakes, using the combinations with higher ranking that are able, separately, to compute such parameters.

In the bottom sections of Tables 3 and 4 for raw intensities, we report the numbers of earthquakes located (15,103) and sized (5703) by each combination according to such procedure. The total number of located earthquakes is about ∼2/3 of the

TABLE 2

Scores (See From MDPs to macroseismic parameters Section) Obtained, Using Raw Intensities, by the 132 Combinations for Both the Distances between Macroseismic and Instrumental Locations (Di) and the Differences between Macroseismic and Instrumental Magnitudes (dM)

			Raw Intensities										
		den	2000	3500	5500				3500	3500			
Dset: A, n. eqks: 1144		MG	RA	RA	RA	SQ	SQ	HE	RS	RH	DB	DB	DB
		Size				1	2	2	3	3	0.5	1	2
Di	Bx-0	Mean	56	61	54	45	45	43	51	86	36	46	51
		mdna	70	55	52	64	56	42	62	54	51	49	60
		mn10	42	46	45	36	36	38	50	52	35	43	40
		mn15	42	42	47	30	37	47	52	48	36	53	41
		mn20	54	62	43	59	34	52	53	59	39	40	67
		mn25	53	63	48	58	38	45	54	67	44	48	54
	Bx-1	Mean	103	88	126	98	87	100	93	127	92	96	129
		mdna	106	91	119	112	93	97	100	109	79	78	144
		mn10	97	92	114	78	99	91	75	121	87	85	116
		mn15	92	97	90	87	81	75	72	90	74	85	145
		mn20	84	94	74	78	67	86	68	107	78	81	146
		mn25	82	101	92	79	75	88	90	101	86	77	130
dM	Bx-0	Mean	324	318	324	318	273	291	285	231	312	267	246
		mdna	330	342	309	300	258	258	258	234	300	252	279
		mn10	297	282	315	288	282	306	243	279	285	297	303
		mn15	285	297	309	303	294	267	246	252	330	237	315
		mn20	303	342	318	300	297	225	240	243	300	282	240
		mn25	288	258	303	285	309	252	246	261	291	270	258
	Bx-1	Mean	324	285	318	273	240	249	294	264	255	246	315
		mdna	294	258	282	285	249	222	258	324	237	285	324
		mn10	306	318	366	255	273	249	264	291	222	279	384
		mn15	273	315	288	246	276	276	270	297	243	276	357
		mn20	315	312	249	270	285	246	258	330	270	285	279
		mn25	261	297	285	255	255	234	258	321	264	252	291

The comparison refers to the dataset of common earthquakes (n. eqks) for which the parameters can be calculated by all the combinations of methods. Grouping methods are indicated (see Appendix A) as a function of population density (den), clustering method (MG), and grid or radius of the area (size), whereas the central tendency estimators are indicated by acronyms: average (mean), median (mdna), and trimmed averages with 10%, 15%, 20%, and 25% of tail trimming: (mn10, mn15, mn20, and mn25, respectively). Bx-0 and Bx-1 indicate BOXER methods 0 and 1, respectively. Results for corrected intensities are reported in Table S2. DB, density-based spatial clustering of applications with noise (DBSCAN) method; HE, hexagonal grid; RA, radius; RH, radius and hexagonal grid; RS, radius and square grid; and SQ, square grid.

22,761 earthquakes, and magnitudes can only be estimated for ~1/5 of the earthquakes. This is because, for the location, one MDP is sufficient, and at least four MDPs are needed for the magnitude. Hence, it is not possible to locate 7658 and to size 17,058 earthquakes. The results for corrected intensities are shown in Tables S3 and S4, with similar values for earthquakes located (15,100) and sized (5625), and not located (7661), and not sized (17,136).

For raw intensities (Table 2), combinations using BOXER-1 and BOXER-0 can locate ~1/3 and ~2/3 of the 15,103 earthquakes, respectively. In detail, 3321 (22%) earthquakes can be located using the "top"-scoring combination (DB2-20% trimmed average-BOXER-1), other 1816 (12%) earthquakes can be located by different combinations using BOXER-1. Overall, BOXER-1 locates at the best 5137 earthquakes, that is, all the events that have number of MDPs ≥ 5. BOXER-0 locates

TABLE 3

Ranking Order of the 132 Combinations of Methods Based on Distance Scores in Table 2 for Raw Intensities. Numbers of Events for Which Macroseismic Parameters Can Be Computed by Each Combination, Following the Order of the Ranking

			Raw Intensities										
		den	**2000**	**3500**	**5500**				**3500**	**3500**			
		MG	**RA**	**RA**	**RA**	**SQ**	**SQ**	**HE**	**RS**	**RH**	**DB**	**DB**	**DB**
Dset: A		**Size**				**1**	**2**	**2**	**3**	**3**	**0.5**	**1**	**2**
Di, n eqks: 1144	Bx-0	Mean	81	75	87	110	109	113	97	44	129	105	96
		mdna	66	82	94	71	80	115	74	85	95	99	76
		mn10	118	106	107	126	127	123	98	92	130	112	121
		mn15	116	117	103	132	125	104	93	100	128	89	119
		mn20	83	73	114	78	131	91	90	77	122	120	69
		mn25	88	72	102	79	124	108	84	70	111	101	86
	Bx-1	Mean	16	40	7	22	42	20	29	6	32	26	5
		mdna	15	34	9	12	28	23	19	13	54	57	3
		mn10	24	33	11	58	21	35	61	8	43	47	10
		mn15	30	25	37	41	52	60	65	38	64	48	2
		mn20	49	27	63	56	68	46	67	14	55	51	1
		mn25	50	18	31	53	62	39	36	17	45	59	4
Di, n eqks: 15,103, nd: 7,658	Bx-0	Mean	–	–	–	–	–	–	–	6409	–	–	–
		mdna	1864	–	–	–	3	–	20	–	–	–	–
		mn10	–	–	–	–	–	–	–	–	–	–	–
		mn15	–	–	–	–	–	–	–	–	–	–	–
		mn20	–	–	–	–	–	37	–	–	–	–	1633
		mn25	–	–	–	–	–	–	–	–	–	–	–
	Bx-1	Mean	–	–	49	–	–	127	–	1040	12	10	–
		mdna	116	–	–	232	–	–	148	–	–	–	–
		mn10	–	–	–	–	76	–	–	–	–	–	–
		mn15	–	–	–	–	–	–	–	–	–	–	–
		mn20	–	–	–	–	–	–	–	–	–	–	3321
		mn25	–	6	–	–	–	–	–	–	–	–	–

Acronyms as in Table 2; nd indicates the number of earthquakes that cannot be located or sized by any combination. Results for corrected intensities in Table S3.

9966 earthquakes, 6409 (42.4%) of which by the combination "3500 RH3-average," 1864 (12.3%), by the "2000 RA-median" 1633 (10.8%) by the "DB2-20% trimmed average." The latter three combinations correspond to the 44th, 66th, and 69th positions in the ranking, respectively (Table 3). Overall, 17 combinations are used to locating 15,103 earthquakes.

The situation is similar for corrected intensities (Table S2) in which 3319 (22%) earthquakes can be located by the same top-scoring combination for raw intensities, 1756 (11.6%) by other combinations using BOXER-1, and 10,025 (66.4%) by combinations using BOXER-0. Overall, BOXER-1 locates 5075 earthquakes of 5134 earthquakes with the number of MDPs ≥ 5 (Table 1). In addition, for corrected intensities, 17 combinations locate 15,100 earthquakes. Excluding the top-scoring combination (DB2-20% trimmed average-BOXER-1), the median and average are generally used for

TABLE 4
As in Table 3 for Magnitude Difference (dM) Scores and Raw Intensities

			Raw Intensities										
		den	2000	3500	5500				3500	3500			
		MG	RA	RA	RA	SQ	SQ	HE	RS	RH	DB	DB	DB
Dset: A		**Size**				**1**	**2**	**2**	**3**	**3**	**0.5**	**1**	**2**
dM, n eqks: 1144	Bx-0	Mean	10	15	11	17	77	49	56	129	25	86	116
		mdna	6	5	28	39	94	96	100	128	36	106	70
		mn10	41	66	23	54	69	31	120	72	57	40	33
		mn15	65	43	29	34	48	85	117	108	7	125	24
		mn20	32	4	18	37	42	130	123	121	38	68	122
		mn25	55	95	35	64	27	105	118	91	50	81	92
	Bx-1	Mean	12	60	16	80	124	109	47	87	101	114	21
		mdna	46	99	67	58	112	132	93	13	126	62	9
		mn10	30	19	2	102	78	110	89	51	131	71	1
		mn15	79	20	53	113	76	75	84	45	119	74	3
		mn20	22	26	111	83	59	115	98	8	82	61	73
		mn25	90	44	63	103	104	127	97	14	88	107	52
dM, n eqks: 5,703, nd: 17,058	Bx-0	Mean	36	3	5	157	7	15	19	2	7	2	1
		mdna	180	88	1	8	7	27	13	14	13	1	28
		mn10	3	–	1	–	1	134	1	38	1	16	13
		mn15	2	1	2	19	29	11	9	3	504	–	79
		mn20	4	196	1	11	7	1	–	–	–	13	–
		mn25	–	–	–	–	158	2	2	39	–	–	1
	Bx-1	Mean	–	–	–	–	–	–	25	–	–	–	2
		mdna	–	–	–	–	–	–	5	58	–	1	29
		mn10	2	–	156	–	–	–	–	2	–	–	3060
		mn15	–	–	–	–	–	3	2	10	–	–	44
		mn20	3	–	–	–	–	–	–	359	–	2	–
		mn25	–	–	–	–	–	–	–	4	–	–	–

Acronyms as in Table 2; nd indicates the number of earthquakes that cannot be located or sized by any combination. Results for corrected intensities in Table S4.

locating earthquakes (Table 3), in agreement with the highest-ranking values in Table 2 and Figure 6.

For raw intensities (Table 4), combinations using BOXER-1 and BOXER-0 assign the magnitude at best to 3767 and 1936 events, respectively (i.e., ∼2/3 and ∼1/3 of the total of 5703 earthquakes). This preference for BOXER-1 is even more pronounced with corrected intensities (Table S4) with 4959 (88%) of the total of 5625 events, whereas combinations with BOXER-0 assess the magnitude at best for only 666 events (12%). Using raw intensities (Table 4), 3060 (53.7%) magnitudes can be determined by the top-scoring combination (DB2-10% trimmed mean-BOXER-1), other 707 (12.4%) by combinations using BOXER-1, 1936 (34.4%) by combinations using BOXER-0. In all, 71 combinations are used to compute

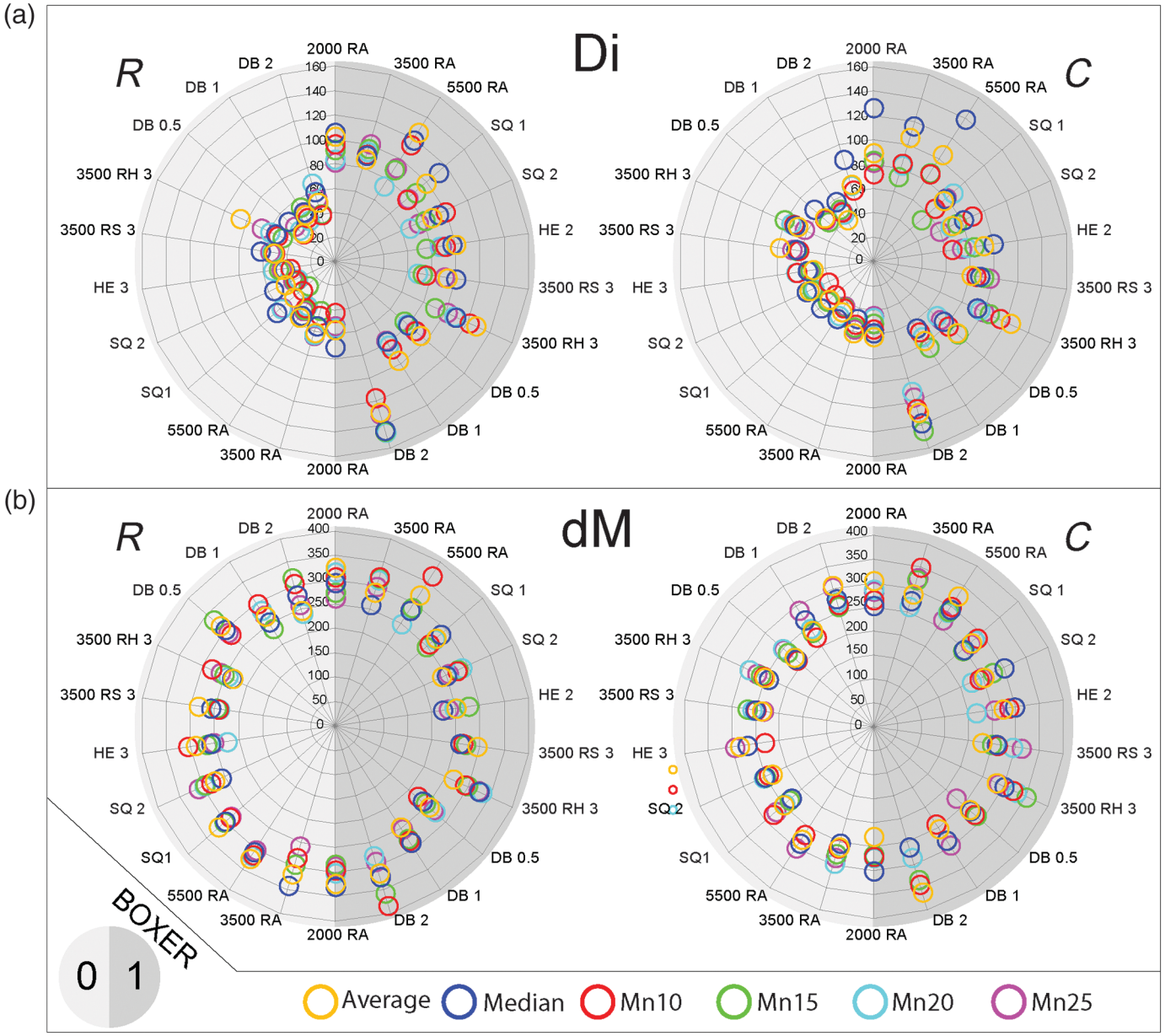


Figure 6. Radar diagrams of the data represented in Table 2 and Table S2 for (a) distance (Di) and (b) difference of magnitude (dM) for BOXER-0 (Bx-0) and BOXER-1 (Bx-1). The light gray areas refer to BOXER-0, and the dark gray ones refer to BOXER-1. Colored symbols (small circles) refer to central tendency estimators used to compute MDPs, plotted as a function of methods used to cluster raw IDPs (codified as in Table 2). The number of earthquakes (and then the agreement with instrumental data) increases from the center of each circle outward. Results for raw intensities (R), and corrected ones (C).

the 5703 magnitudes. Using corrected intensities, 2984 (53%) magnitudes can be determined by the top-scoring combination (DB2-mean-BOXER-1), other 1965 (35.1%) by combinations using BOXER-1, and 666 (11.8%) by combinations using BOXER-0. In all, 83 combinations are used to compute 5625 magnitudes. All the grouping, central tendency, and BOXER methods are necessary to compute epicenters of magnitudes for all earthquakes.

From a first analysis of the correspondence between macroseismic and instrumental parameters in Figure 7, a geographical heterogeneity is quite evident: a fairly good agreement is observed in Europe and North America, and some greater discrepancy in other areas of the World. For this reason, we will analyze the results not only at a global scale but also for the five macroareas indicated in Figure 7: Europe (EU), Asia and Oceania (AO), North America (US), South America (SA), and Africa (AF). It is obvious to relate the agreement and disagreement between macroseismic and instrumental parameters with the number of IDPs available in the different areas. In fact, the larger number of IDPs in the EU and US with greater density and continuity (Fig. 1) corresponds to

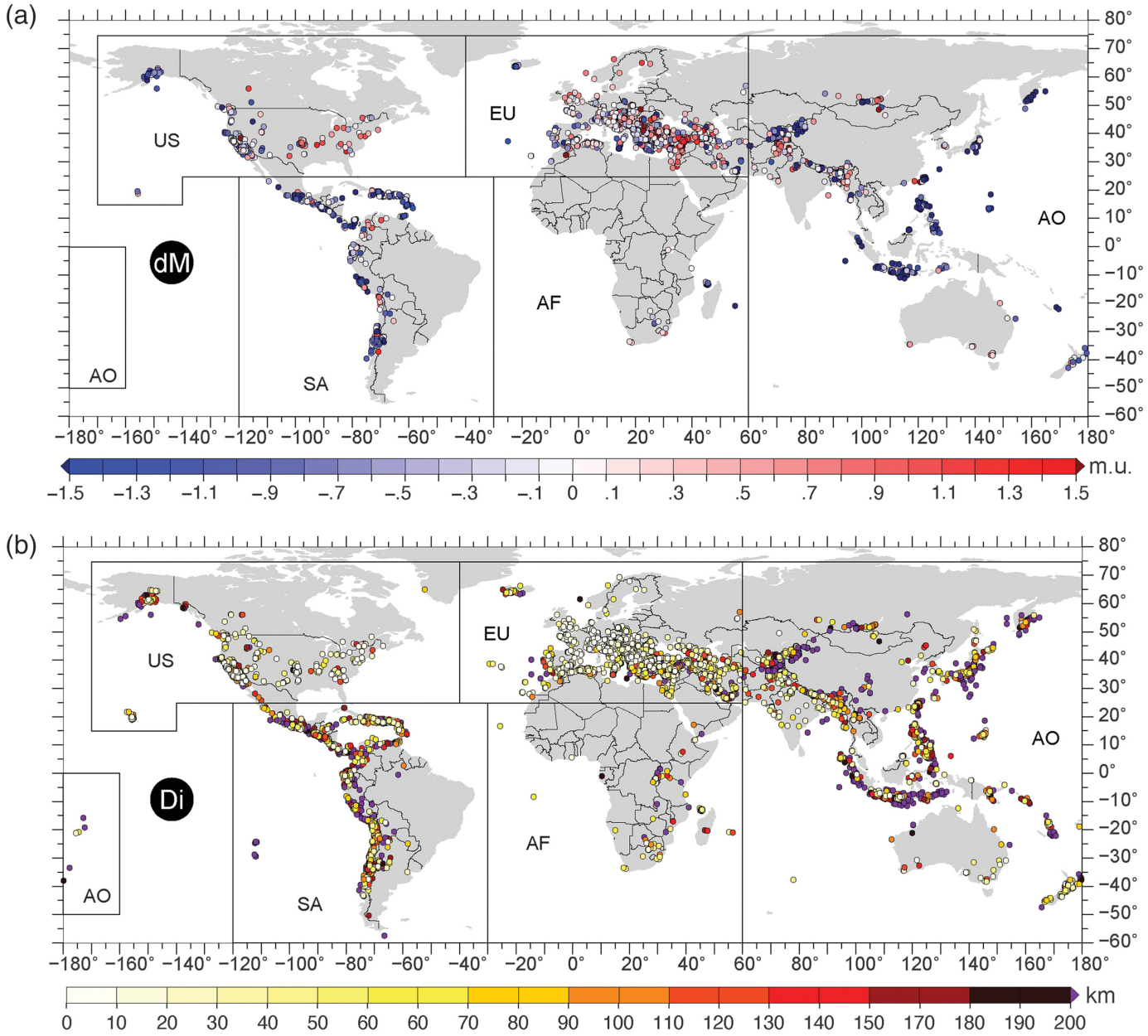


Figure 7. Plot of (a) difference of magnitude (dM, 5703 earthquakes) and (b) distance (Di, 15,103 earthquakes) between "preferred" macroseismic parameters and instrumental data for raw intensity. Five zones (AF, Africa; AO, Asia and Oceania; EU, Europe; SA, South America; and US, North America) are shown.

a higher average number of MDPs in the same areas for each analyzed earthquake (Table 5 for raw and corrected intensities). Therefore, such larger number of MDPs per earthquake manages to better constrain location and macroseismic magnitude, improving the agreement with the instrumental data at the global scale (see Fig. S3 and Table S5).

We calculated the frequency histograms in various ranges of distances and magnitude differences both at the global scale and for different macroareas (Fig. 8 with numerical values in Tables S6 and S8 for raw intensities and Fig. S4, Tables S8 and S9 for corrected ones). The lower the values, the better the fit of macroseismic to instrumental values. All earthquakes (a) have also been divided into categories or subsets, depending on whether they are located inland (*L*) or offshore (*S*), have the maximum gap between available MDPs and epicenter less than 180°(*g*), and, for epicentral distance only, have at least three MDPs (*n*). We do not consider the latter subdivision for magnitudes, because the minimum number of MDPs for computing them is 4. As well, three MDPs are at least required for a gap <180°. This comparison between macroseismic and instrumental parameters is displayed in Figure 8, both in terms of number of events and of percentage of the total number.

Both for the distance and for the difference in magnitude, at the global scale, the earthquakes located on land (*L*) are ~2/3 of the total (*a*) and ~1/3 are located offshore (*S*). The agreement is generally better for the former ones than for the latter ones

TABLE 5
Average Number of Macroseismic Data Points (MDPs) per Earthquake (n. MDPs/Eqk) at Global Scale and for Macroareas (as in Fig. 7) Using Raw and Corrected Intensities

n. MDPs/Eqk		Global (W)	Europe (EU)	Asia and Oceania (AO)	North America (US)	South America (SA)	Africa (AF)
Raw intensity	Di	5.1	5.6	3.8	6.5	3.2	4.4
	dM	11.4	11.7	9.4	13.9	8.9	7.5
Corrected intensity	Di	5.1	5.5	3.8	6.5	3.2	4.4
	dM	11.5	11.9	9.6	14.2	9.1	7.8

Values for distance (Di) and difference of magnitude (dM) are shown.

and improves by a few percentage points by only considering earthquakes with at least three MDPs (*n*). The agreement further improves for earthquakes with maximum azimuthal gap lower than 180°, which number, however, is about one-fourth of the total for the location and to about one-half for the magnitude. For about 30% of earthquakes, the distance exceeds 50 km; and it exceeds 100 km for about 15% of them. Only 40% of the earthquakes have magnitude differences less than 0.6 m.u. This indicates a certain difficulty of the macroseismic magnitudes in reproducing the instrumental ones.

Analyzing the results by macroareas, the correspondence between macroseismic and instrumental data shows significant variations: ∼2/3 of the earthquakes are concentrated in Europe, whereas the other macroareas have about 1200–2200 earthquakes with location and 400–700 with magnitude except for the African area having about 200 events (Fig. 8). Compared with the data at a global scale, a clearly better agreement between macroseismic and instrumental parameters is observed for the EU and the US areas, and a worse agreement for AO, SA, and AF (Fig. 8).

It is also clear that events in the sea (*S*) have worse agreement with the instrumental data than all the other datasets (*a*, *L*, *n*, and *g*). Compared with the whole dataset (*a*), the trend of improvement of the agreement is evident for the subsets *L*, *n*, and *g*. It follows that the number of MDPs (*n*), possibly well distributed around the epicenter (*g*), are factors that improve the quality of the final macroseismic data, making the calculated parameters more reliable. Increasing more and more the number of IDPs and then of MDPs is a goal and a mean to obtain realistic estimates of macroseismic parameters. The use of corrected intensities leads to results substantially similar to those calculated with raw intensities, with some slight improvements at the shortest distances and smaller magnitude difference (Fig. S5).

To show the evolution over time of the agreement between macroseismic and instrumental data, we subdivided the results by year, from 2012 to 2022 (excluding 2023 which has only two months of data). Figure 9 shows the overall results of the distance and magnitude difference at a global scale, both in terms of number of earthquakes and of percentage. For each year, the earthquakes are divided into subsets (SaLng for distance and SaLg for magnitude difference) analogously to Figure 8. It is possible to observe how the number of events macroseismic parameters for which are estimated increases over time, except for year 2022 in which it decreases. The agreement between macroseismic and instrumental parameters remains similar to each other, even within the subsets of earthquakes.

Over time, the distances and the differences in magnitude decrease: in 2020–2022 for the subsets "L," "n," and "g," the percentage of earthquakes located within 10 km from the instrumental epicenter is about 20%–30%, about 30%–50% within 20 km, about 50%–70% within 30 km, and about 80% within 50 km. For the differences in magnitude, a slight percentage improvement over time is observed with about 25%–40% of the earthquakes of the subsets "L" and "g" within about 0.3 m.u. and about 65% of events within 0.6 m.u. The trend of improvement over time is even more visible, considering events beyond certain values (e.g., 100 km away and 1 degree of magnitude), which halves their percentages compared to the first few years. It should also be noted that some years like 2016 and 2017 have percentages in line or even better in terms of agreement than the most recent years.

The temporal behavior in the different macroareas compared with the global scale (Fig. 9) shows different results both in terms of percentage and of the number of earthquakes. For Europe (Fig. 10), it can be observed that the number of earthquakes slightly decreases in 2018 and in 2022, but increases the percentage of earthquakes that have relatively shorter distances and smaller magnitude differences. Furthermore, over the years, we can note a marked decrease in the percentages of earthquakes with distances longer than 100 km and magnitude differences greater than 1° (Fig. 10). In the other macroareas (AO, US, SA, and AF), we have about one-third of the total number of earthquakes analyzed. For certain years and/or certain subsets of earthquakes, the small number of available events makes the statistics scarcely significant. The North America (US) area has similar or even slightly better agreement than that of Europe, with the exception of years 2012 and 2013 when the statistics are insignificant due to the low number of events (Figs. S6 and S7

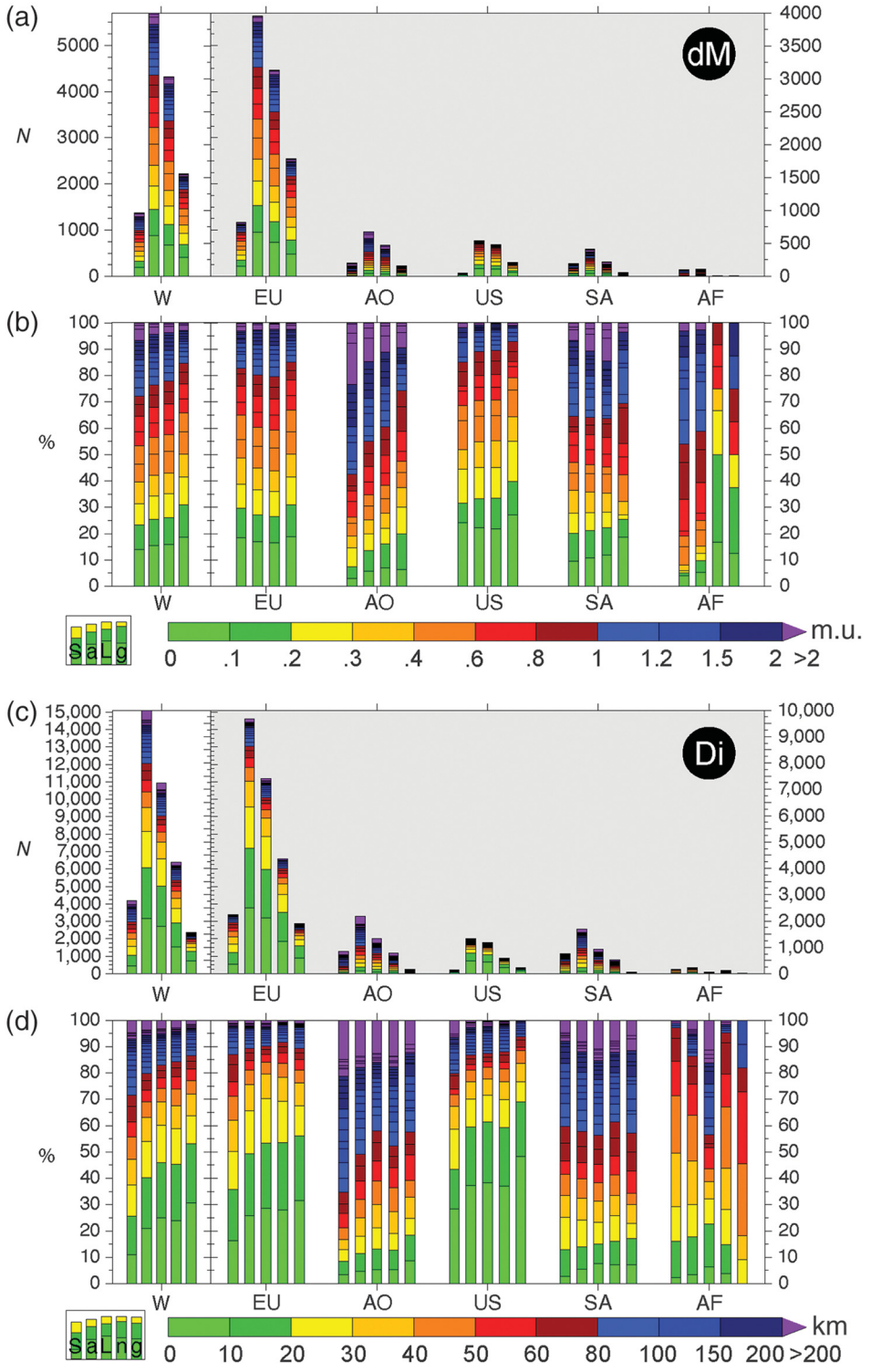


Figure 8. Statistical results of the comparison between macroseismic and instrumental parameters (represented in Fig. 7). Plots display numbers of event (*N*) and percentages (%) for (a,b) magnitude differences (dM) and (c,d) distances (Di). Columns refer to global scale (*W*) and different macroareas (AF, Africa; AO, Asia and Oceania; EU, Europe; SA, South America; and US, North America). The columns of each zone (see the legend in lowest left corner) indicate, from left to right, the earthquakes located offshore (*S*), all the earthquakes (*a*), earthquakes located inland (*L*), earthquakes with the number of MDPs ≥ 3 (*n*), and earthquakes with azimuthal gap less than to 180° (*g*). The area in gray highlights the macroareas with respect to the global area (*W*). The scales of the numbers of earthquakes (*N*) are different for the global area (left) and the macroareas (right).

for raw and corrected intensities, respectively). In the other macroareas (Figs. S8 and S9), the percentage of well-localized events and well-assigned magnitudes also drops significantly due to the reduced number of MDPs per earthquake (Table 5). Using the corrected intensity gives similar results (Figs. S10–S15). For the sake of clarity, we also provide an example of the entire procedure from IDPs to macroseismic parameters for the 19 September 2020 California earthquake (06:38 UTC, M 4.5) in Appendix B.

Conclusions

We analyzed the database of individual intensities provided by citizens (1,874,376 IDPs), collected and made available online by the EMSC for 51,359 earthquakes. The database provides two intensity values: raw and corrected (i.e., eliminating intensities >10 and applying an empirical formula to the raw data, according to Bossu *et al.*, 2017). We applied various methods for grouping the IDPs on both the raw and corrected datasets. We tested the combinations of 11 clustering methods and six central tendency estimators (mean, median, trimmed means with various trimming intervals) to derive an MDP intensity for each cluster with at least three IDPs. The MDPs thus available were processed with methods 0 and 1 of the BOXER code (Gasperini *et al.*, 2010). Therefore, there are 132 possible combinations of methods for each event, for each type of intensity, which allow to compute epicenter and macroseismic magnitude. The threshold of at least three IDPs for deriving an MDP significantly

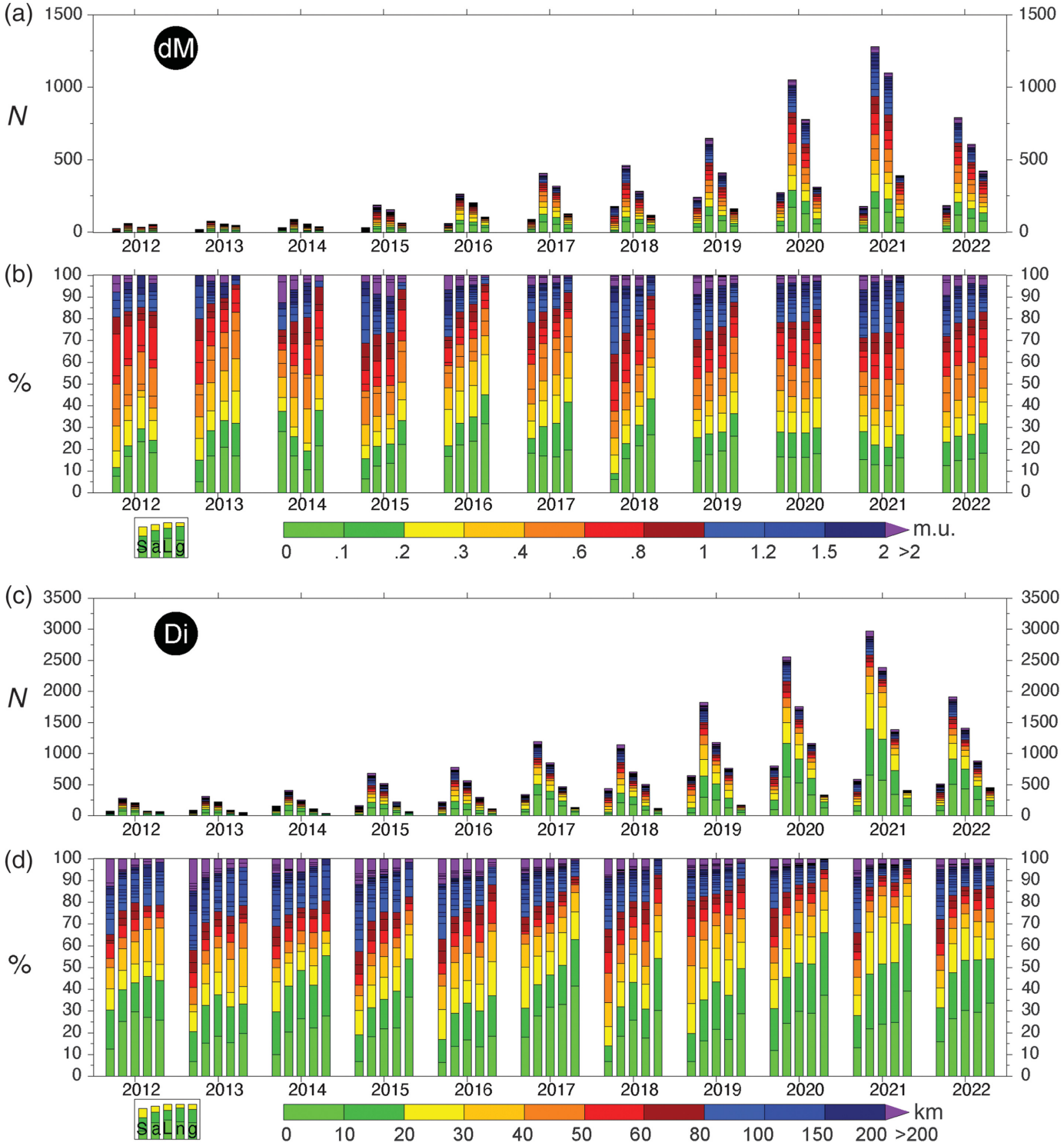
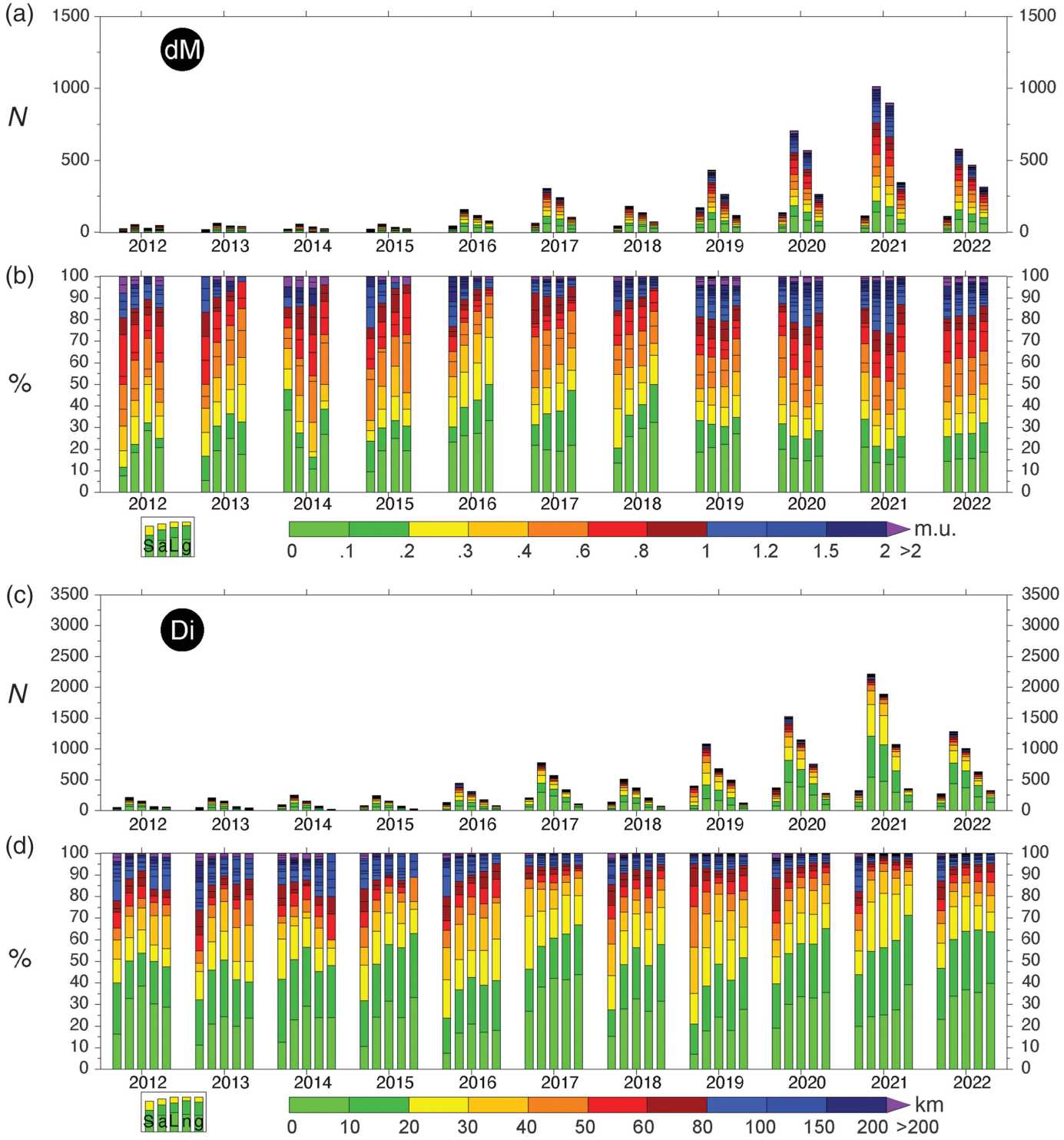


Figure 9. Same as in Figure 8, at the global scale and for different years.

lowers the number of earthquakes for which macroseismic parameters can actually be calculated. Furthermore, at least four MDPs are required for the calculation of magnitude. Therefore, it is possible to compute an epicenter and a magnitude for ~15,000 and ~5,700 earthquakes, respectively.

The calculated macroseismic parameters can be compared with the instrumental ones to evaluate the reliability of the entire methodology. To identify the combination that minimizes the difference with the instrumental data, separately for distance and magnitude, we selected about a thousand earthquakes for which the parameters could be calculated for all the possible combinations. A score was assigned to each



combination, based on its ability to well reproduce the instrumental parameters of each earthquake. This systematic approach shows similar score values for several combinations of methods, especially concerning the difference in magnitude. Considering the distance alone, however, the better overall results are obtained by BOXER-1 compared with BOXER-0. The score assigned to the different combinations for all earthquakes defines a ranking that can be used to select the most preferable ones in a prospective view.

Figure 10. Same as in Figure 9 for Europe.

Because not all earthquakes can be located and sized by the best performing combination, other combinations must also be used to determine the parameters for as many earthquakes as possible. In particular, most earthquakes can only be located using BOXER-0, because it requires less MDPs than BOXER-1 to be applied.

In addition to the complete dataset of available earthquakes (1), we also considered subsets of events with epicenter located on land (*L*), offshore (*S*), with number of MDPs ≥ 3 (*n*), and with azimuthal gap between MDPs and instrumental epicenters <180° (*g*).

The analyses we brought, not only at a global scale but also for five macroareas (EU, AO, US, SA, and AF), show substantially similar results between raw and corrected intensities. The distribution of available earthquakes shows a clear concentration in Europe with ∼2/3 of the total data. In general, the fit between macroseismic and instrumental parameters shows an increasing trend from the dataset of earthquakes located offshore (*S*) up to the dataset of earthquakes with a gap (*g*) of less than 180°, with intermediate results for other datasets (*a*, *L*, and *n*). Moreover, compared with the global scale, some macroareas (EU and US) have a better fit than others (AO, SA, and AF). We can argue that the larger numbers of MDPs per earthquake that we have in EU and US has a role in improving the agreement with instrumental parameters. In the practice, future near-real-time analyses will take advantage of knowing the macroarea where an event occurs to give a preliminary assessment of the likely reliability of calculated parameters.

Analyzing the results as a function of time and macroareas, we can observe increasing trends for subsets as well as for the complete dataset, except for certain areas or certain years for which the low number of events makes the statistics poorly significant. With the increase over time of the number of MDP available per earthquakes, an improvement of the fit between macroseismic and instrumental parameters is generally observed. In certain areas such as Europe and North America, 60%–70% of the events are localized within about 30 km from the instrumental epicenter with a magnitude difference of <0.6 m.u.; and, above all, there is a strong reduction over time of extreme differences (more than 100 km of distance or >1 of magnitude). However, in other areas the agreement is still not so good probably due to the still low number of MDPs. Therefore, it is desirable to continue to increase the number of IDPs, and to overcome the economic and political barriers, which today exclude large areas of the Earth from the possibility of providing such information.

Finally, the reporting of IDPs could also be influenced by the thumbnails representing the different scenarios associated with various degrees used in the LastQuake system. In particular, the types of houses and furniture depicted in them are more similar to European and North American environments than to those of other macroareas, and this makes it more difficult to apply the EMS98 scale to the damage scenario.

The processing performed by applying the BOXER code to the IDPs data in an original way is essential and preparatory for future applications in near-real-time. When, for an event, EMSC starts collecting IDPs from citizens, an automatic procedure can be run. If the number of IDPs is enough to allow their grouping into MDPs, it will be possible to assess location and magnitude with BOXER following a preferential ranking order. The greater the number of MDPs, the greater the reliability of the result. In particular, we believe that the threshold of five MDPs allowing the application of the BOXER-1 method is a discriminating element to give greater reliability to the results. Obviously, further comparative tests of the results at time intervals will have to be conducted, exploiting the delay-time information with respect to the time T_0 origin of the event; but all this will be the subject of further specific work and is beyond the scope of the present purposes.

Data and Resources

Boxer code freely available at https://emidius.mi.ingv.it/boxer/. Cities500.txt database, available at https://www.geonames.org. "Did You Feel It?" (DYFI) available at http://earthquake.usgs.gov/earthquakes/dyfi/. Individual intensity data points (IDPs) are downloaded by European–Mediterranean Seismological Centre (EMSC) available at www.seismicportal.eu/testimonies-ws/; for example, http://www.seismicportal.eu/testimonies-ws/api/search?unids=20210629_0000012&includeTestimonies=true. EMSC available at https://www.emsc-csem.org. GHSL database available at https://ghsl.jrc.ec.europa.eu/index.php. GeoNames database available at https://www.geonames.org. GeoNet New Zealand questionnaires available at https://www.geonet.org.nz. "Hai sentito il terremoto?" (HSIT) available at http://www.haisentitoilterremoto.it/. SHAKEMAP—a tool for Earthquake Response available at https://earthquake.usgs.gov/data/shakemap/. The supplemental materi al includes figures and tables that provide further information and details of the main text. Moreover, similar elaborations, plots, and figures are given for the "corrected" intensities in as for the "raw" intensities in the main text. All websites were last accessed in July 2023.

Declaration of Competing Interests

The authors acknowledge that there are no conflicts of interest recorded.

Acknowledgments

This article benefitted from funding provided by the H2020 EU project RISE (Real-time earthquake rIsk reduction for a reSilient Europe) n. 821115 and Istituto Nazionale di Geofisica e Vulcanologia-Sezione di Bologna. The authors thank Rémy Bossu and Matthieu Landès for data, discussion, and suggestions in the preparation and drafting of the article.

References

Amorese, D., R. Bossu, and G. Mazet-Roux (2015). Automatic clustering of macroseismic intensity data points from internet questionnaires: Efficiency of the partitioning around medoids (PAM), *Seismol. Res. Lett.* **86,** no. 4, 1171–1177.

Bakun, W. H., and C. M. Wentworth (1997). Estimating earthquake location and magnitude from seismic intensity data, *Bull. Seismol. Soc. Am.* **87,** 1502–1521.

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

Bossu, R., M. Laurin, G. Mazet-Roux, F. Roussel, and R. Steed (2015). The importance of smartphones as public earthquake-information tools and tools for the rapid engagement with eyewitnesses: A case

study of the 2015 Nepal earthquake sequence, *Seismol. Res. Lett.* **86,** no. 6, 1587–1592.
Bossu, R., F. Roussel, L. Fallou, M. Landès, R. Steed, G. Mazet-Roux, A. Dupont, L. Frobert, and L. Petersen (2018). LastQuake: From rapid information to global seismic risk reduction, *Int. J. Disast. Risk Reduction* **28,** 32–42.
Dewey, J., D. Wald, and L. Dengler (2000). Relating conventional USGS modified Mercalli intensities to intensities assigned with data collected via the Internet, *Seismol. Res. Lett.* **71,** 264.
Ester, M., H. P. Kriegel, J. Sander, and X. Xiaowei (1996). A density-based algorithm for discovering clusters in large spatial databases with noise, *KDD'96: Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, 2–4 August 1996, Portland, Oregon, U.S.A., 226–231.
Florczyk, A., C. Corbane, M. Schiavina, M. Pesaresi, L. Maffenini, M. Melchiorri, P. Politis, F. Sabo, S. Freire, D. Ehrlich, *et al.* (2019). GHS Urban Centre Database 2015, multitemporal and multidimensional attributes, R2019A, European Commission, Joint Research Centre (JRC), available at https://ghsl.jrc.ec.europa.eu/ghs_stat_ucdb2015mt_r2019a.php.
Gasperini, P., F. Bernardini, G. Valensise, and E. Boschi (1999). Defining seismogenic sources from historical earthquake felt reports, *Bull. Seismol. Soc. Am.* **89,** 94–110.
Gasperini, P., G. Vannucci, D. Tripone, and E. Boschi (2010). The location and sizing of historical earthquakes using the attenuation of macroseismic intensity with distance, *Bull. Seismol. Soc. Am.* **100,** 2035–2066, doi: 10.1785/0120090330.
Goded, T., N. Horspool, S. Canessa, A. Lewis, K. Geraghty, A. Jeffrey, and M. Gerstenberger (2018). New macroseismic intensity assessment method for New Zealand web questionnaires, *Seismol. Res. Lett.* **89,** no. 2A, 640–652, doi: 10.1785/0220170163.
Grünthal, G. (Editor) (1998). *European Macroseismic Scale 1998, Conseil de l'Europe*, Cahiers du Centre Européen de Géodynamique et de Séismologie, Vol. 13, Luxembourg, 99 (in French).
Hough, S. E., and S. S. Martin (2021). Which earthquake accounts matter? *Seismol. Res. Lett.* **92,** no. 2A, 1069–1084, doi: 10.1785/0220200366.
International Seismological Centre (2022). On-line bulletin, doi: 10.31905/D808B830.
Lolli, B., P. Gasperini, and G. Vannucci (2014). Empirical conversion between teleseismic magnitudes (m_b and M_s) and moment magnitude (M_w) at the Global, Euro-Mediterranean and Italian scale, *Geophys. J. Int.* **199,** 805–828, doi: 10.1093/gji/ggu264.
Musson, R. M. W., and M. J. Jiménez (2008). Macroseismic estimation of earthquake parameters, NA4 Deliverable D3, NERIES Project, available at https://emidius.mi.ingv.it/neries_NA4/docs/NA4_D5.pdf (last accessed July 2023).
Pasolini, C., D. Albarello, P. Gasperini, 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, doi: 10.1785/0120070021.
Pettenati, F., and L. Sirovich (2003). Tests of source-parameter inversion of the U.S. Geological Survey intensities of the Whittier Narrows, 1987 earthquake, *Bull. Seismol. Soc. Am.* **93,** 47–60.
Rovida, A., M. Locati, R. Camassi, B. Lolli, and P. Gasperini (2020). The Italian earthquake catalogue CPTI15, *Bull. Earthq. Eng.* **18,** 2953–2984, doi: 10.1007/s10518-020-00818-y.
Sieberg, A. (1912). Über die makroseismische Bestimmung der Erdbebenstärke, in *Ein Beitrag zur seismologische Praxis*, G. Gerlands (Editor), Vol. 11**,** Beiträge zur Geophysik, 227–239 (in German).
Sieberg, A. (1932). Erdebeben, in *Handbuch der Geophysik*, B. Gutenberg (Editor), Vol. 4, 552–554 (in German).
Tosi, P., P. Sbarra, V. De Rubeis, and C. Ferrari (2015). Macroseismic intensity assessment method for web-questionnaires, *Seismol. Res. Lett.* **86,** 985–990, doi: 10.1785/022014022.
Vannucci, G., P. Gasperini, B. Lolli, and L. Gulia (2019). Fast characterization of sources of recent Italian earthquakes from macroseismic intensities, *Tectonophysics* **750,** 70–92, doi: 10.1016/j.tecto.2018.11.002.
Vannucci, G., D. Tripone, P. Gasperini, G. Ferrari, and B. Lolli (2015). Automated assessment of macroseismic intensity from written sources using the Fuzzy sets, *Bull. Earthq. Eng.* **13,** 2769–2803, doi: 10.1007/s10518-015-9759-5.
Wald, D. J., V. Quitoriano, L. Dengler, and J. W. Dewey (1999). Utilization of the Internet for rapid community intensity maps, *Seismol. Res. Lett.* **70,** 87–102.
Wald, D. J., V. Quitoriano, B. Worden, M. Hopper, and J. W. Dewey (2011). USGS "Did You Feel It?" Internet-based macroseismic intensity maps, *Ann. Geophys.* **54,** 688–707, doi: 10.4401/ag-5354.

Appendix A

Method for grouping IDPs and compute MDPs

The transformation from individual intensity data points (IDPs) to macroseismic data points (MDPs) implies a delimitation of the IDPs in a felt area and in some cases a selection of IDPs discarding the out-of-area data.

For retrospective statistical analyses or to derive relationships from available IDPs, the area is limited to the threshold distance defined by MDPE (equation 1). The use of such a filter does not change or modify the number of MDPs that are actually calculated, because it only eliminates isolated IDPs (i.e., geographic outliers), but it does significantly reduce the calculation time required to create the subsequent geographic grids (for more than 15,000 earthquakes) on which to check cell by cell the relative occurrence of IDPs. IDPs available for each earthquake can be clustered or not in areas by six different methods (Fig. 5):

1. radius (RA);
2. square grid (SQ);
3. hexagonal grid (HE);
4. radius and square grid (RS, i.e., RA + SQ);
5. radius and hexagonal grid (RH, i.e., RA + HE); and
6. DBSCAN method (DB).

In detail:

- RA method uses georeferenced localities (from a database) as cluster centers. Starting from the identification location, the radius constrains a representative surface of the location. IDPs can be in or out of the "city-equivalent" area. We use a database of global localities (i.e., the open source cities 500.txt, see Data and Resources section) that also provide the number of inhabitants for each locality, however, without

population density information. Some databases, such as GHS Urban Centre Database (Florczyk *et al.*, 2019), from GHSL (see Data and Resources), also collect open source information on surface, but only of 13,000 cities worldwide. By setting a population density (e.g., 2000, 3500, and 5500 inhabitants/km^2), it is possible derive a "city-equivalent" area, i.e., a spatial area roughly proportional to the number of inhabitants. The radius is computed as $\sqrt{(\text{area}/\pi)}$. IDPs located within the radius from the locality belong (In) to the locality or not (Out). Within each area, the clustering of IDPs starts with the localities with the smallest number of inhabitants and continues by grouping the remaining IDPs following the localities with increasing numbers of inhabitants.

- SQ and HE methods use a regular equal-areal grid with squared and hexagonal mesh, respectively. The center of development of the grid is fixed to the average of coordinates of all IDPs inside the area.
- RS (RA + SQ) and RH (RA + HE) methods combine the method of clustering 1 with 2 and 3, respectively: first the RA method is applied, then remaining IDPs are grouped by the SQ or HE method. This approach overcomes, in certain cases, the simplification of equating the locality area to a circle based on a fixed population density and allows to retrieve information about IDPs outside of RAs, but in a sufficient number so that to compute residual MDPs over grid.
- DB (Density-based spatial clustering of applications with noise [DBSCAN], Ester *et al.*, 1996) is a method based on the grouping of IDPs located less than an arbitrary distance that successively can aggregate neighbour clusters and IDPs. If an IDP is close to another one (that is it is located at a distance, or "EPS" [Epsilon-neighborhood] radius smaller than a given value), the two IDPs are grouped together in the same aggregation area. However, if one of aggregated IDPs is close to another IDP at a distance smaller than the EPS radius, then the latter IDP is joined together with the aggregation area to which the former IDP belongs. This technique proceeds in a chain by joining IDPs to the cluster and is able to discover clusters of arbitrary shape. IDPs at a distance greater than the EPS radius from all the IDPs of the cluster are external or belong to other, distinct, aggregation areas.

For each grouping method (1–6), setting parameter (e.g., population density, grid side, or EPS distance) constrains the areas for IDP grouping. We derived MDPs using various combinations of grouping methods and central tendency estimators, and by varying the reference settings. In particular (Table A1), we used:

1. three population densities of 2000, 3500, and 5500 inhabitants/km^2 using the RA method;
2. regular equidimensional grids with side of the mesh of 1 and 2 km for squared cells (SQ) and 2 km for hexagonal cells (HE);
3. a population density of 3500 inhabitants/km^2 and side of the grid of 3 km, both for RS and RH methods; and
4. three eps radii (0.5, 1, and 2 km) for DB methods

TABLE A1
Summary of Grouping Methods and Settings Used to Derive Macroseismic Data Points (MDPs) from Individual Intensity Data Points (IDPs)

Intensities: Raw (*R*), Corrected (*C*)

	Central Tendency Estimators (Mean, mdna, mn10, mn15, mn20, and mn25)			
Grouping Methods	**Population Density (den) (Inhabitants/km^2)**	**Side of Cells (gr) (in km)**	**EPS Radius (eps) (in km)**	**Speed Test**
RA	2000, 3500, and 5500			Fast
SQ		1 and 2		Slow+
HE		2		Slow–
RS	3500	3		Medium+
RH	3500	3		Medium–
DB			0.5, 1, and 2	Fast

The speed test is a relative indication of the processing time of MDPs from the slowest to the fastest, with further intermediate levels (+ and –). In total, 11 grouping methods, six central tendency estimators, and two type of intensities (*R* and *C*) are used for comparative analyses on the EMSC earthquakes. DB, density-based spatial clustering of applications with noise (DBSCAN) method; HE, hexagonal grid; RA, radius; RH, radius and hexagonal grid; RS, radius and square grid; and SQ, square grid.

The methods for assessing MDPs are not equivalent to each other in terms of computing time. In Table A1, the last column gives a raw evaluation of the computational speed of the method (high, average, low speed, and relative comparison with "+" and "–" symbols). Grouping methods based on SQ and HE grids require more computer time than RA or DB (Table A1) methods. The RS and RH methods are intermediate between the previous approaches. The construction of grids requires a complete coverage of the whole area; and the smaller the size of the grid side, the longer the time to construct the grid and therefore to search for IDPs within each cell. The tessellation with hexagonal cells, due to a higher complexity, is more time consuming than that with square cells. RS and RH methods use grids with sides slightly wider than SQ and HE ones, so they are faster than SQ and HE methods. In any case, the higher the level of detail one wants to achieve as spatial coverage, the more the time needed to perform computations. The DB method (Fig. 5) is independent of external data, like locality databases, or of grid tessellation, and is based only on the available information (location and intensity) of the IDPs.

We tested the combinations of different grouping methods and settings for the all the earthquakes of the EMSC dataset.

To simplify the discussion of analyses and statistics, we then selected some settings only indicated in Table A1. Combinations can be represented by combining acronyms: "3500RH3-mean" uses both the radius (R) of "city-equivalent" area, based on a population density of 3500 inhabitants/km^2 and hexagonal cells (H) of side 3 km as grouping method, and the median to derive the MDP intensity, while "2000RA-mdna" uses a radius (R) with population density of 2000 inhabitants/km^2 and the median.

Other methods of data clustering (e.g., based on polylines of the limit of urban areas at different sites) are not available on a global scale with the same quality: some countries may have these data even for small locations, while others do not. Determining whether or not an IDP falls within a polyline is a time-consuming calculation.

For future near-real-time analyses, the instrumental location and magnitude of events are unknown when IDPs are made available since the event time (T_0). The IDPs collected at the subsequent time steps (T_1, T_2, T_n…) directly define the maximum and the minimum latitude and longitude of the survey area, because the analysis of only one event at a time does not create problems of excessive calculation time. All the IDPs (even the geographical outliers) will be tested to verify their occurrence in the grouping areas for the assessment of the MDPs. In any

TABLE B1

Numerical Values of the Data in Figure B3 for BOXER Method (Bx-mth) 0 and 1, 11 Grouping Methods (MGs), and Used Settings (den, gr or eps as in Table A1) e Six Central Tendency Estimators (CTEs)

BX-mth	Lon (° + km)	Lat (° + km)	M	Di (km)	dM (m.u.)	n. MDPs	den	MGs	gr or eps (km)	CTEs
0	−118.1107 ± 3.6	34.0106 ± 2.4	4.49 ± 0.22	3.02	−0.01	171	2000	RA	–	mnsa
	−118.1012 ± 2.7	34.0185 ± 2.4	4.65 ± 0.24	1.96	0.15					mdna
	−118.1027 ± 3.4	34.0200 ± 2.6	4.49 ± 0.22	2.09	−0.01					mn10
	−118.1057 ± 3.0	34.0224 ± 2.2	4.5 ± 0.22	2.38	0					mn15
	−118.0968 ± 3.5	34.0214 ± 2.7	4.49 ± 0.19	1.55	−0.01					mn20
	−118.1075 ± 3.4	34.0197 ± 2.4	4.49 ± 0.19	2.54	−0.01					mn25
	−118.1308 ± 3.8	34.0290 ± 2.8	4.52 ± 0.23	4.79	0.02	146	3500			mnsa
	−118.1101 ± 3.6	34.0097 ± 2.8	4.32 ± 0.26	3	−0.18					mdna
	−118.1212 ± 3.4	34.0162 ± 4.4	4.39 ± 0.24	3.82	−0.11					mn10
	−118.1247 ± 3.6	34.0241 ± 3.7	4.4 ± 0.24	4.14	−0.1					mn15
	−118.1268 ± 3.7	34.0051 ± 3.5	4.38 ± 0.25	4.62	−0.12					mn20
	−118.1318 ± 3.8	34.0057 ± 2.7	4.37 ± 0.25	5.03	−0.13					mn25
	−118.162 ± 3.9	34.0311 ± 4.1	4.49 ± 0.24	7.65	−0.01	126	5500			mnsa
	−118.1302 ± 3.2	33.9933 ± 2.7	4.46 ± 0.50	5.5	−0.04					mdna
	−118.1334 ± 3.6	33.9983 ± 4.3	4.49 ± 0.22	5.48	−0.01					mn10
	−118.1328 ± 3.6	33.9983 ± 4.3	4.49 ± 0.22	5.44	−0.01					mn15
	−118.1011 ± 3.4	33.9812 ± 3.7	4.41 ± 0.28	4.73	−0.09					mn20
	−118.1108 ± 3.9	33.9877 ± 3.5	4.44 ± 0.36	4.58	−0.06					mn25
	−118.2493 ± 2.9	34.0989 ± 1.0	4.42 ± 0.19	17.9	−0.08	184	-	SQ	1	mnsa
	−118.1518 ± nd	34.1158 ± nd	4.3 ± 0.24	12.54	−0.2					mdna
	−118.2493 ± 2.9	34.0989 ± 1.0	4.43 ± 0.19	17.89	−0.07					mn10
	−118.2647 ± 2.3	34.0972 ± 1.1	4.38 ± 0.19	19.06	−0.12					mn15

Macroseismic latitudes, longitudes, and magnitudes also report the uncertainties values computed by BOXER. Distance (Di) and difference of magnitude (dM) with respect to instrumental values are indicated. The preferred data (BOXER-1, DBSCAN with eps 2 km, trimmed mean 20 for location, and mean 10 for magnitude) in bold characters. We use the average (mnsa), the median (mdna), and the trimmed mean with four different intervals of the distribution of the sorted intensity values: 10%–90% (mn10), 15%–85% (mn15), 20%–80% (mn20), and 25%–75% (mn25). DB, density-based spatial clustering of applications with noise (DBSCAN) method; den, population density; HE, hexagonal grid; MDP, macroseismic data points; MG, clustering method; RA, radius; RH, radius and hexagonal grid; RS, radius and square grid; and SQ, square grid.

(Continued next page.)

TABLE B1 (continued)

Numerical Values of the Data in Figure B3 for BOXER Method (Bx-mth) 0 and 1, 11 Grouping Methods (MGs), and Used Settings (den, gr or eps as in Table A1) e Six Central Tendency Estimators (CTEs)

BX-mth	Lon (° + km)	Lat (° + km)	M	Di (km)	dM (m.u.)	n. MDPs	den	MGs	gr or eps (km)	CTEs
	−118.1518 ± nd	34.1158 ±	4.34 ± 0.23	12.54	−0.16					mn20
	−118.1518 ± nd	34.1158 ± nd	4.33 ± 0.23	12.54	−0.17					mn25
	−118.1938 ± 3.1	34.0475 ± 1.6	4.58 ± 0.13	10.93	0.08	347			2	mnsa
	−118.1567 ± 3.4	34.0089 ± 1.5	4.57 ± 0.12	7.17	0.07					mdna
	−118.2 ± 2.8	34.0354 ± 1.5	4.57 ± 0.14	11.2	0.07					mn10
	−118.2043 ± 2.8	34.0289 ± 1.6	4.55 ± 0.14	11.5	0.05					mn15
	−118.1676 ± 3.7	34.0091 ± 1.6	4.59 ± 0.12	8.16	0.09					mn20
	−118.1631 ± 3.7	34.0111 ± 1.6	4.58 ± 0.12	7.73	0.08					mn25
	−118.1771 ± 2.0	34.0511 ± 1.5	4.58 ± 0.14	9.6	0.08	356	–	HE		mnsa
	−118.1386 ± 3.1	34.0058 ± 1.3	4.55 ± 0.12	5.63	0.05					mdna
	−118.1544 ± 2.4	34.0388 ± 1.6	4.58 ± 0.14	7.17	0.08					mn10
	−118.1554 ± 2.7	34.0448 ± 1.8	4.54 ± 0.13	7.47	0.04					mn15
	−118.1274 ± 3.3	33.9897 ± 1.5	4.54 ± 0.13	5.52	0.04					mn20
	−118.1371 ± 3.0	33.9966 ± 1.5	4.54 ± 0.13	5.87	0.04					mn25
	−118.1647 ± 3.0	34.0423 ± 1.8	4.54 ± 0.14	8.19	0.04	358	3500	RS	3	mnsa
	−118.1364 ± 4.3	33.9753 ± 8.0	4.58 ± 00.12	7.19	0.08					mdna
	−118.1776 ± 2.7	34.0368 ± 2.0	4.52 ± 0.14	9.18	0.02					mn10
	−118.1722 ± 2.4	34.0366 ± 2.1	4.5 ± 0.14	8.7	0					mn15
	−118.1364 ± 4.3	33.9753 ± 8.0	4.55 ± 0.13	7.19	0.05					mn20
	−118.1364 ± 4.3	33.9753 ± 8.0	4.55 ± 0.13	7.19	0.05					mn25
	−118.1652 ± 3.2	34.0490 ± 2.5	4.55 ± 0.17	8.49	0.05	268	3500	RH	3	mnsa
	−118.1441 ± 3.4	34.0276 ± 1.8	4.47 ± 0.19	5.97	−0.03					mdna
	−118.1475 ± 3.9	34.0330 ± 2.4	4.55 ± 0.17	6.38	0.05					mn10
	−118.1458 ± 3.9	34.0370 ± 2.7	4.56 ± 0.17	6.35	0.06					mn15
	−118.1423 ± 5.3	34.0055 ± 2.4	4.49 ± 0.19	5.96	−0.01					mn20
	−118.1446 ± 5.0	34.0076 ± 2.3	4.5 ± 0.19	6.12	0					mn25
	−118.2151 ± 4.1	34.1075 ± 3.9	4.48 ± 0.19	15.8	−0.02	179	–	DB	0.5	mnsa
	−118.1953 ± 4.0	34.0920 ± 2.6	4.34 ± 0.24	13.3	−0.16					mdna
	−118.2152 ± 4.1	34.1075 ± 3.9	4.49 ± 0.19	15.81	−0.01					mn10
	−118.1897 ± 4.7	34.1010 ± 4.4	4.49 ± 0.20	13.54	−0.01					mn15
	−118.1953 ± 4.0	34.0920 ± 2.6	4.39 ± 0.23	13.3	−0.11					mn20
	−118.1953 ± 4.0	34.0920 ± 2.6	4.39 ± 0.24	13.3	−0.11					mn25
	−118.191 ± 2.8	34.0133 ± 3.8	4.6 ± 0.17	10.26	0.1	241			1	mnsa

Macroseismic latitudes, longitudes, and magnitudes also report the uncertainties values computed by BOXER. Distance (Di) and difference of magnitude (dM) with respect to instrumental values are indicated. The preferred data (BOXER-1, DBSCAN with eps 2 km, trimmed mean 20 for location, and mean 10 for magnitude) in bold characters. We use the average (mnsa), the median (mdna), and the trimmed mean with four different intervals of the distribution of the sorted intensity values: 10%–90% (mn10), 15%–85% (mn15), 20%–80% (mn20), and 25%–75% (mn25). DB, density-based spatial clustering of applications with noise (DBSCAN) method; den, population density; HE, hexagonal grid; MDP, macroseismic data points; MG, clustering method; RA, radius; RH, radius and hexagonal grid; RS, radius and square grid; and SQ, square grid.

(Continued next page.)

TABLE B1 (continued)

Numerical Values of the Data in Figure B3 for BOXER Method (Bx-mth) 0 and 1, 11 Grouping Methods (MGs), and Used Settings (den, gr or eps as in Table A1) e Six Central Tendency Estimators (CTEs)

BX-mth	Lon (° + km)	Lat (° + km)	M	Di (km)	dM (m.u.)	n. MDPs	den	MGs	gr or eps (km)	CTEs
	–118.1405 ± 4.6	34.0022 ± 3.4	4.41 ± 0.21	5.91	–0.09					mdna
	–118.1668 ± 3.6	34.0107 ± 4.0	4.65 ± 0.17	8.07	0.15					mn10
	–118.1669 ± 3.6	34.0107 ± 4.0	4.63 ± 0.17	8.08	0.13					mn15
	–118.1355 ± 5.7	33.9773 ± 4.0	4.62 ± 0.15	6.98	0.12					mn20
	–118.1354 ± 5.4	33.9726 ± 3.8	4.62 ± 0.15	7.34	0.12					mn25
	–118.1115 ± 6.5	33.9731 ± 1.3	4.61 ± 0.32	5.96	0.11	128			2	mnsa
	–118.061 ± 8.2	34.0041 ± 11.8	4.65 ± 0.23	2.48	0.15					mdna
	–118.1119 ± 6.5	33.9731 ± 1.3	4.62 ± 0.32	5.99	0.12					mn10
	–118.112 ± 6.4	33.9725 ± 1.3	4.61 ± 0.31	6.05	0.11					mn15
	–118.0366 ± 13.9	33.9725 ± 20.3	4.61 ± 0.29	6.63	0.11					mn20
	–118.0368 ± 13.9	33.9728 ± 20.3	4.6 ± 0.29	6.59	0.1					mn25
1	–118.0934 ± 2.3	34.0103 ± 2.7	4.49 ± 0.23	1.64	–0.01	171	2000	RA	–	mnsa
	–118.0959 ± 2.1	34.0307 ± 2.6	4.66 ± 0.29	1.89	0.16					mdna
	–118.08 ± 2.2	34.0264 ± 4.1	4.5 ± 0.22	0.72	0					mn10
	–118.0786 ± 2.3	34.0249 ± 3.5	4.51 ± 0.23	0.56	0.01					mn15
	–118.0801 ± 2.4	34.0314 ± 3.2	4.5 ± 0.25	1.27	0					mn20
	–118.1013 ± 2.3	34.0367 ± 2.7	4.5 ± 0.26	2.7	0					mn25
	–118.1202 ± 3.2	34.0522 ± 3.4	4.55 ± 0.24	5.16	0.05	146	3500			mnsa
	–118.1267 ± 3.1	34.0573 ± 2.9	4.34 ± 0.26	5.98	–0.16					mdna
	–118.1176 ± 3.2	34.0552 ± 3.9	4.42 ± 0.25	5.23	–0.08					mn10
	–118.1344 ± 4.1	34.0659 ± 3.2	4.43 ± 0.26	7.15	–0.07					mn15
	–118.1441 ± 4.3	34.0597 ± 3.2	4.39 ± 0.26	7.38	–0.11					mn20
	–118.1683 ± 3.7	33.9997 ± 3.0	4.36 ± 0.26	8.45	–0.14					mn25
	–118.1394 ± 3.8	34.0625 ± 3.2	4.51 ± 0.30	7.23	0.01	126	5500			mnsa
	–118.1114 ± 2.5	34.0323 ± 3.4	4.47 ± 0.50	3.2	–0.03					mdna
	–118.1567 ± 3.0	34.0008 ± 2.8	4.49 ± 0.30	7.39	–0.01					mn10
	–118.1608 ± 3.0	34.0008 ± 2.7	4.49 ± 0.30	7.75	–0.01					mn15
	–118.1143 ± 3.2	34.0493 ± 3.4	4.4 ± 0.24	4.54	–0.1					mn20
	–118.1164 ± 3.4	34.0479 ± 3.7	4.39 ± 0.27	4.57	–0.11					mn25
	–118.1637 ± 2.1	34.1082 ± 2.9	4.55 ± 0.21	12.48	0.05	184	–	SQ	1	mnsa
	–118.1542 ± 1.6	34.1147 ± 1.5	4.3 ± 0.24	12.55	–0.2					mdna
	–118.1628 ± 2.0	34.1088 ± 2.6	4.55 ± 0.21	12.48	0.05					mn10
	–118.176 ± 2.7	34.1026 ± 6.6	4.5 ± 0.21	12.75	0					mn15

Macroseismic latitudes, longitudes, and magnitudes also report the uncertainties values computed by BOXER. Distance (Di) and difference of magnitude (dM) with respect to instrumental values are indicated. The preferred data (BOXER-1, DBSCAN with eps 2 km, trimmed mean 20 for location, and mean 10 for magnitude) in bold characters. We use the average (mnsa), the median (mdna), and the trimmed mean with four different intervals of the distribution of the sorted intensity values: 10%–90% (mn10), 15%–85% (mn15), 20%–80% (mn20), and 25%–75% (mn25). DB, density-based spatial clustering of applications with noise (DBSCAN) method; den, population density; HE, hexagonal grid; MDP, macroseismic data points; MG, clustering method; RA, radius; RH, radius and hexagonal grid; RS, radius and square grid; and SQ, square grid.

(Continued next page.)

TABLE B1 (continued)

Numerical Values of the Data in Figure B3 for BOXER Method (Bx-mth) 0 and 1, 11 Grouping Methods (MGs), and Used Settings (den, gr or eps as in Table A1) e Six Central Tendency Estimators (CTEs)

BX-mth	Lon (° + km)	Lat (° + km)	M	Di (km)	dM (m.u.)	n. MDPs	den	MGs	gr or eps (km)	CTEs
	−118.1533 ± 1.7	34.1142 ± 1.6	4.33 ± 0.23	12.46	−0.17					mn20
	−118.1534 ± 1.7	34.1144 ± 1.5	4.32 ± 0.23	12.49	−0.18					mn25
	−118.099 ± 2.0	34.033 ± 2.8	4.63 ± 0.15	2.27	0.13	347			2	mnsa
	−118.1007 ± 2.1	34.0123 ± 2.5	4.6 ± 0.13	2.09	0.1					mdna
	−118.102 ± 2.1	34.0239 ± 2.7	4.62 ± 0.15	2.07	0.12					mn10
	−118.1023 ± 2.0	34.0278 ± 2.8	4.61 ± 0.16	2.23	0.11					mn15
	−118.1121 ± 2.3	34.02772 ± 2.9	4.62 ± 0.13	3.06	0.12					mn20
	−118.1142 ± 2.3	34.0316 ± 2.9	4.6 ± 0.14	3.4	0.1					mn25
	−118.1206 ± 2.1	34.052 ± 2.4	4.6 ± 0.16	5.17	0.1	356	–	HE	2	mnsa
	−118.1109 ± 2.1	34.0143 ± 2.2	4.56 ± 0.16	2.92	0.06					mdna
	−118.0914 ± 1.9	34.0193 ± 2.5	4.57 ± 0.15	1.06	0.07					mn10
	−118.1056 ± 2.0	34.0341 ± 2.9	4.55 ± 0.15	2.84	0.05					mn15
	−118.1114 ± 2.1	34.0096 ± 2.0	4.54 ± 0.14	3.12	0.04					mn20
	−118.1269 ± 2.2	34.0191 ± 2.3	4.54 ± 0.14	4.33	0.04					mn25
	−118.1092 ± 2.1	34.0056 ± 1.8	4.54 ± 0.17	3.13	0.04	358	3500	RS	3	mnsa
	−118.1342 ± 2.5	34.051 ± 3.0	4.56 ± 0.13	6.07	0.06					mdna
	−118.1014 ± 2.0	34.0014 ± 1.7	4.53 ± 0.17	2.86	0.03					mn10
	−118.1186 ± 2.4	34.0058 ± 1.7	4.51 ± 0.18	3.89	0.01					mn15
	−118.1226 ± 2.4	34.0526 ± 2.7	4.54 ± 0.14	5.34	0.04					mn20
	−118.1348 ± 2.3	34.0542 ± 2.6	4.53 ± 0.14	6.32	0.03					mn25
	−118.1092 ± 3.5	34.0141 ± 3.0	4.53 ± 0.20	2.77	0.03	268	3500	RH	3	mnsa
	−118.1257 ± 2.5	34.0375 ± 2.1	4.47 ± 0.19	4.63	−0.03					mdna
	−118.1139 ± 2.8	34.0213 ± 3.9	4.54 ± 0.19	3.13	0.04					mn10
	−118.1209 ± 3.3	34.0185 ± 4.0	4.55 ± 0.20	3.78	0.05					mn15
	−118.1385 ± 2.7	34.0154 ± 2.5	4.5 ± 0.20	5.42	0					mn20
	−118.1407 ± 2.8	34.0219 ± 2.6	4.51 ± 0.19	5.6	0.01					mn25
	−118.154 ± 3.0	34.1121 ± 4.4	4.56 ± 0.23	12.3	0.06	179	**–**	**DB**	0.5	mnsa
	−118.1536 ± 2.7	34.1112 ± 3.2	4.38 ± 0.24	12.2	−0.12					mdna
	−118.151 ± 2.8	34.1116 ± 4.2	4.57 ± 0.22	12.1	0.07					mn10
	−118.151 ± 2.9	34.1124 ± 4.0	4.54 ± 0.22	12.18	0.04					mn15
	−118.1516 ± 2.8	34.1098 ± 3.5	4.43 ± 0.23	11.97	−0.07					mn20
	−118.1521 ± 2.8	34.1103 ± 3.4	4.43 ± 0.24	12.04	−0.07					mn25
	−118.0917 ± 2.5	34.0065 ± 3.2	4.61 ± 0.18	1.85	0.11	241			1	mnsa

Macroseismic latitudes, longitudes, and magnitudes also report the uncertainties values computed by BOXER. Distance (Di) and difference of magnitude (dM) with respect to instrumental values are indicated. The preferred data (BOXER-1, DBSCAN with eps 2 km, trimmed mean 20 for location, and mean 10 for magnitude) in bold characters. We use the average (mnsa), the median (mdna), and the trimmed mean with four different intervals of the distribution of the sorted intensity values: 10%–90% (mn10), 15%–85% (mn15), 20%–80% (mn20), and 25%–75% (mn25). DB, density-based spatial clustering of applications with noise (DBSCAN) method; den, population density; HE, hexagonal grid; MDP, macroseismic data points; MG, clustering method; RA, radius; RH, radius and hexagonal grid; RS, radius and square grid; and SQ, square grid.

(Continued next page.)

TABLE B1 (continued)

Numerical Values of the Data in Figure B3 for BOXER Method (Bx-mth) 0 and 1, 11 Grouping Methods (MGs), and Used Settings (den, gr or eps as in Table A1) e Six Central Tendency Estimators (CTEs)

BX-mth	Lon (° + km)	Lat (° + km)	M	Di (km)	dM (m.u.)	n. MDPs	den	MGs	gr or eps (km)	CTEs
	–118.1157 ± 2.7	34.0076 ± 3.0	4.41 ± 0.20	3.56	–0.09					mdna
	–118.0806 ± 2.6	34.0043 ± 2.8	4.66 ± 0.19	1.74	0.16					mn10
	–118.0917 ± 2.5	34.0117 ± 3.0	4.65 ± 0.20	1.42	0.15					mn15
	–118.1091 ± 2.8	34.0011 ± 2.8	4.62 ± 0.19	3.41	0.12					mn20
	–118.1164 ± 2.9	33.9924 ± 2.6	4.62 ± 0.20	4.55	0.12					mn25
	–118.0483 ± 2.9	33.9885 ± 2.9	4.59 ± 0.30	4.56	0.09	**128**			**2**	mnsa
	–118.0378 ± 3.7	33.9972 ± 3.4	4.65 ± 0.46	4.64	0.15					mdna
	–118.0481 ± 2.9	33.9853 ± 2.7	**4.62 ± 0.27**	4.86	**0.12**					**mn10**
	–118.0469 ± 2.8	33.9844 ± 2.6	4.6 ± 0.28	5	0.1					mn15
	–118.0443 ± 2.7	**33.9829 ± 2.5**	4.61 ± 0.27	**5.28**	0.11					**mn20**
	–118.0407 ± 4.7	33.9799 ± 4.3	4.6 ± 0.29	5.74	0.1					mn25

Macroseismic latitudes, longitudes, and magnitudes also report the uncertainties values computed by BOXER. Distance (Di) and difference of magnitude (dM) with respect to instrumental values are indicated. The preferred data (BOXER-1, DBSCAN with eps 2 km, trimmed mean 20 for location, and mean 10 for magnitude) in bold characters. We use the average (mnsa), the median (mdna), and the trimmed mean with four different intervals of the distribution of the sorted intensity values: 10%–90% (mn10), 15%–85% (mn15), 20%–80% (mn20), and 25%–75% (mn25). DB, density-based spatial clustering of applications with noise (DBSCAN) method; den, population density; HE, hexagonal grid; MDP, macroseismic data points; MG, clustering method; RA, radius; RH, radius and hexagonal grid; RS, radius and square grid; and SQ, square grid.

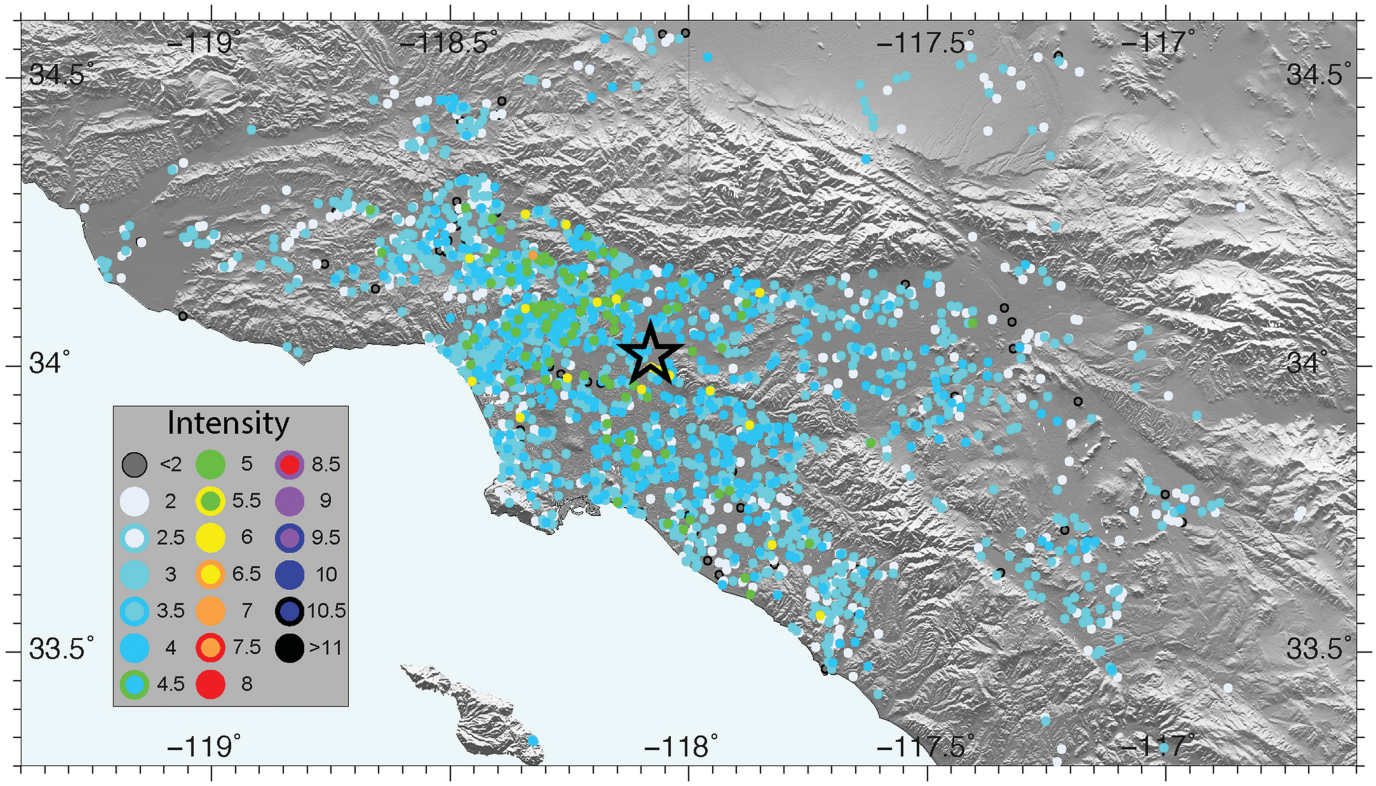


Figure B1. Plot of 2692 individual intensity data points (IDPs; raw intensities) of the event 19 September 2020. The star represents the instrumental epicenter.

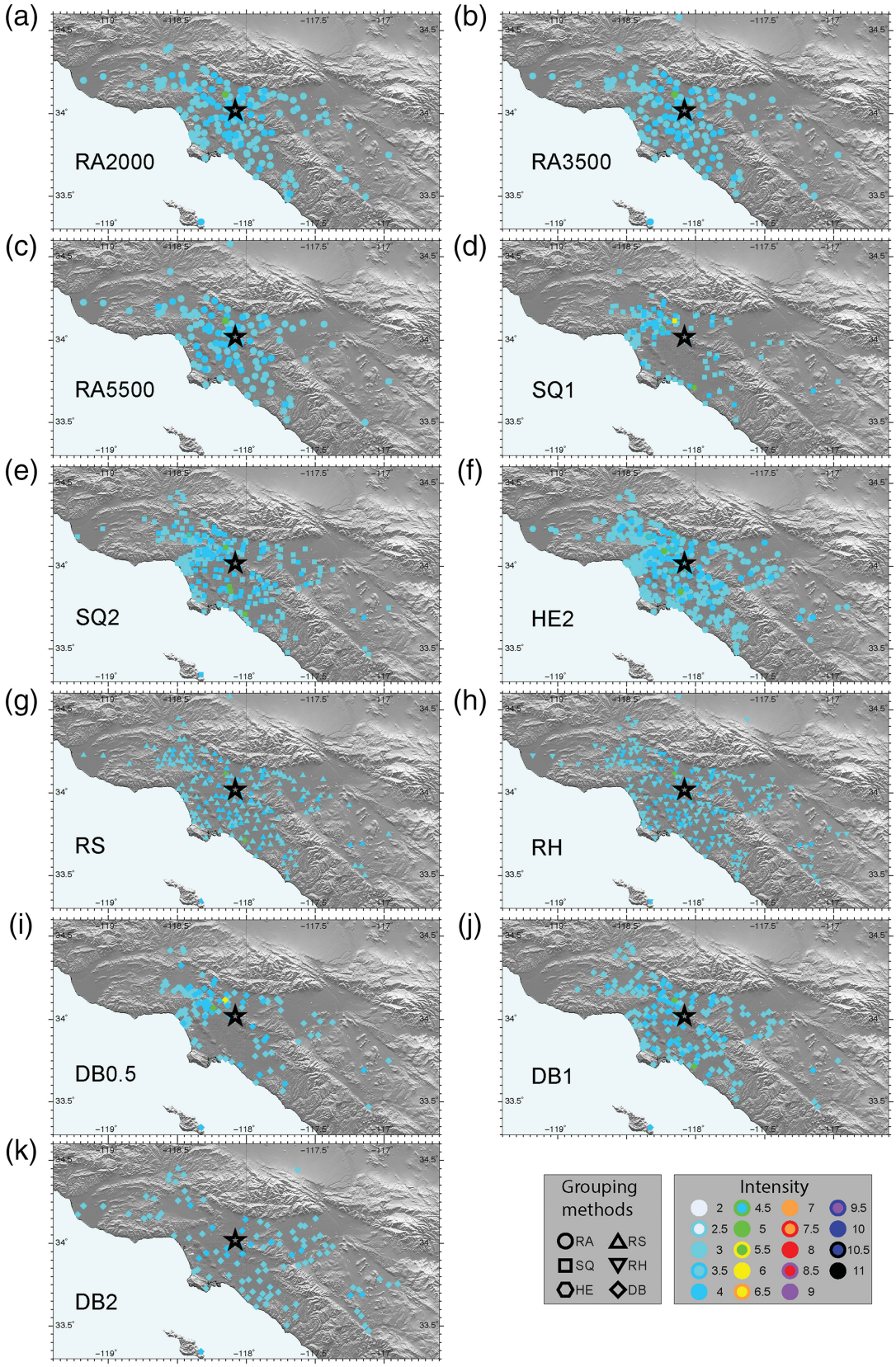


Figure B2. Macroseismic data points (MDPs) for 11 different grouping methods: radius (RA, panels a,b,c); square grid (SQ panels d,e); hexagonal grid (HE, panel f); radius and square grid (RS, panel g); radius and hexagonal grid (RH, panel h); density-based spatial clustering of applications with noise [DBSCAN] method (DB, panels i,j,k), and median as central tendency estimator. See Appendix A and Table A1 for details. The star represents the instrumental epicenter.

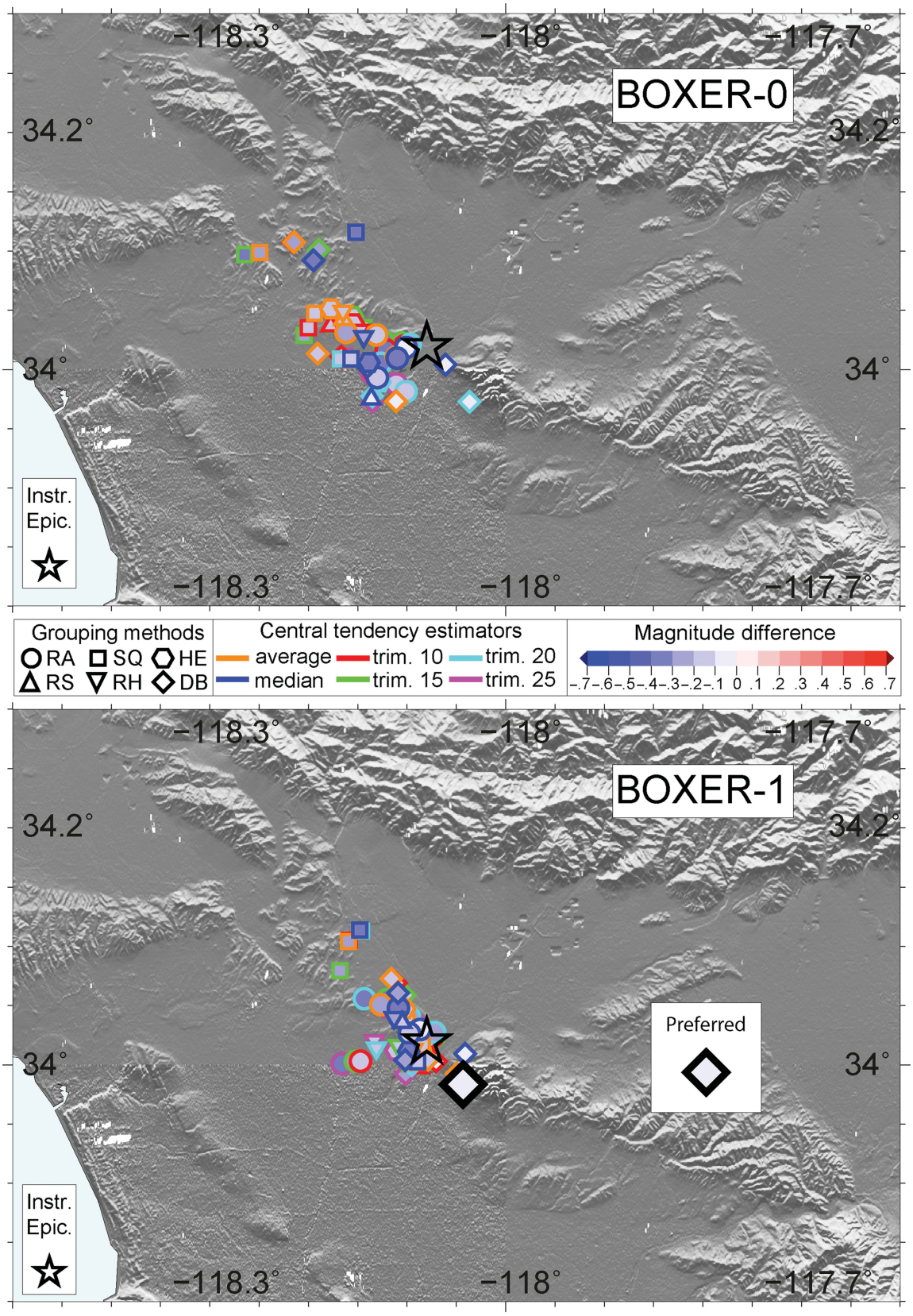


Figure B3. Macroseismic parameters (location and difference of magnitude with respect to instrumental one) with BOXER-0 and BOXER-1 for a total of 132 different MDPs distributions. The preferred solution following the ranking order is indicated (numerical values in Table B1).

case, geographic outliers are generally isolated (i.e., below the expected threshold (3) of the minimum IDP occurrence to assign an MDP) and do not contribute to the creation of MDPs.

Appendix B

Example of the procedure from IDPs to macroseismic parameters

To better represent the procedure from the IDPS to the choice of preferred macroseismic parameters of an earthquake, we show, as an example, the earthquake of 19 September 2020 (06:38 UTC, latitude: 34.02, longitude: −118.08, M_w 4.5) referred to the raw intensities only.

For this event, EMSC provides 2192 IDPs (Fig. B1). The grouping of IDPs into MDPs involves selecting IDPs by discarding geographical outliers (not present in the example, however) and grouping them into MDPs using clustering methods and central tendency estimators for a total of 66 possible MDP distributions (see also Appendix A and Table A1 for details).

Figure B2 shows the MDPs obtained by applying 11 grouping methods and the median as the central tendency distribution. For each group of 66 combinations, the BOXER provides location and magnitude with methods 0 and 1, giving a total of 132 locations and magnitudes. Figure B3 (with numerical values in Table B1) shows the epicenters and the differences in magnitude with respect to the instrumental values. Most of the macroseismic epicenters are very close to the true instrumental one (the maximum distance is about 19 km), generally with small differences in magnitude (the overall range is between −0.3 and 1.5 m.u.).

To choose a preferred location and magnitude, we applied the ranking order (Tables 3 and 4). For the example earthquake, macroseismic parameters are available for all 132 possible combinations (Table B1), so the first ranked combination was chosen for both distance (DBSCAN with eps-radius 2 km, trimmed mean 20 and BOXER-1, Table 3) and magnitude (DBSCAN with eps 2 km, trimmed mean 10, and BOXER-1, Table 4). The macroseismic preferred solution is located at l atitude: 33.9829, longitude: −118.0443, with magnitude: 4.62 (Fig. B3). The distance with respect to instrumental epicenter is 5.28 km, and the difference of magnitude is 0.1 m.u.

Manuscript received 25 July 2023
Published online 30 October 2023