Review - MS No.: essd-2020-372

Title: INGe: Intensity-ground motion dataset for Italy
Author(s): Ilaria Oliveti et al.
MS type: Data description paper

General Comments

The publication of an intensity-ground motion dataset is in principle very welcome. It provides a good basis for further research. The dataset also enables better insights into previous studies on empirical relationships between instrumental ground motion data and intensities. The reprint, however, still contains a number of points that require clarification, addition or amendment. Some of these are of a fundamental nature.

Specific Comments

#1 line 7

It would be important to specify the type of magnitude used for the different earthquakes as part of the dataset. Even the given range can be different if you refer to Mw or ML.

#2 lines 21-26

Different DYFI approaches are discussed and the macroseismic scales they use. The following comments arise in this context:

#2-1 Provide the references for the three scales given in lines 25 and 26.

#2-2 Use the correct reference for the MM scale used for DYFI.

#2-3 Regarding the use of the MCS scale in Italy, it would be important for readers outside Italy or for readers who are not exactly insiders to know why the MSC scale, which is about 90 years old, is still used in Italy. If possible, also give the reference to the authorised translation into Italian, as presumably not all users of the scale in the country can use the original.

#2-4 With regard to the EMS, a distinction must be made between the 1992 and 1998 versions. The EMS-98 version is certainly meant; in this case, the designation must be made accordingly.

#3 line 54

What is meant by the slash in the expressions “(PGV)/intensity” and “(SA)/intensity”?

#4 line 57

It is stated which three spectral acceleration (SA) values are used. On the other hand, chapter 3 deals with practical concerns of the EC8. However, the three used spectral periods are only of limited

relevance with regard to the position of the plateau values of the design spectra of the EC8. Further SA values for periods lower than 0.3 s should be added.

#5 line 57
It says: ". . we . . .remove several earthquakes whose data were proven to be unreliable." It would be important for readers to know which earthquakes this concerns. The knowledge that the authors have gained in this respect should be shared (e.g. in a table with the related earthquakes).
What are the criteria that resulted in excluding certain earthquakes? Are these criteria or findings that have been expressed before or by other authors? Or is this fundamentally new knowledge?
A certain transparency would be expected.

#6 lines 14-65 (Meaning, this comment refers to the introduction in general.)
With regard to earlier datasets or works on empirical relations between ground motion data and intensities in Italy, reference is made to Faenza and Michelini (2010, 2011). There are, however, a number of other works specifically for Italy, but also works worth citing for other study areas. Shouldn't these be acknowledged in the introduction?
Further down in the text, Gomez-Capera et al. (2020) are mentioned. They refer to at least nine similar papers, five of them concerning Italy. All these papers are based on data sets obviously similar to the one described here by the authors of the reprint.

#7 lines 92-93
Here it is stated: ". . . the average macroseismic intensity value of a more or less large area with a point." In the following, individual questions on this:
#7-1 Does "more or less large area" mean the locality to which an MDP applies? Please specify.

#7-2 On the basis of the data set for the reprint https://doi.org/10.13127/inge.1, the localities are villages, small towns and up to medium-sized towns with their geographical extensions. With respect to Ancona it is the centro storico. Right?

#7-3 What might be meant by "the average macroseismic intensity value"? What are these average intensity values? And from which values was the average or mean calculated?
Intensities for a location are determined, as the authors know, from the totality of all macroseismic observations per location - never, as one might think from the wording in the reprint, as the average of point intensities for a location. Only from the totality of all macroseismic observations per location it can be determined in how many cases (or to what percentage) certain effects were observed or not observed. This is because the MCS scale is also based on the use of the frequency (or percentage) of occurrence of certain macroseismic effects per locality, including the types of building damage.
The wording in the lines mentioned would therefore have to be corrected or clearly described as to how it was actually done.

#8 line 94 ff.
#8-1 The intensity is always given in the text, i.e. not only in the line mentioned, with e.g. 6.0, i.e. a decimal number. However, intensities assigned from macroseismic observations are always positive integers from 1 to 12. Therefore, intensities are also indicated with Roman numerals in older literature to emphasise the integer nature of the intensity. However, Roman numerals are somewhat unsuitable for numerical applications. Only derived intensities, i.e. intensities calculated from other quantities by means of empirical relations, appear as decimal numbers.
Therefore, either change the intensities to integers or justify why decimals are used.

#8-2 It is correct to use the notation 6-7 for cases of uncertain macroseismic findings where the intensity can be determined equally well, e.g., as 6 degrees or 7 degrees. But this notation is only used to express the uncertainty described. It is not a range between 6.0 and 7.0, as it says in the mentioned line. It is also by no means an intensity with the accuracy of half a degree. Otherwise the MCS scale would not have 12 degrees but 23.
The wording would therefore have to be changed in the reprint.

#9 line 96
#9-1 Explain the meaning of the abbreviations "(e.g. "HD", "D", or "F")"; i.e. what do "HD" etc. stand for?

#9-2 What means „class 4.0“? Is "class" used as a synonym for "intensity"? And if so, why?

#9-3 Aren't classes numbered as natural integers instead of decimals? Please, explain, why you prefer decimals.

#10 lines 114-118
#10-1 It would be extremely helpful for users if a table could be added with the 65 earthquakes that are dealt with here. Here all parameters, such as Mw and ML, can be specified (i.e. both magnitudes!), but also the number of available MDP per event (see lines 133-137). Many of these data are contained in the data set, but would have to be extracted first. A special table for the 65 earthquakes, as described, would be very user-friendly. Please add the number of available MDP per event to the table.

#10-2 The fact that the reprint does not distinguish whether a magnitude is Mw or ML is, as already mentioned, unacceptable. In addition, the Mw values should be available for all quakes considered in the data set. The CPTI15 file (Rovida et al. 2020; BEE vol. 18) lists all earthquakes in Italy with Mw >= 4.0 (either true or proxy). The CPTI15 file also contains the related uncertainty for each Mw value. The uncertainty in magnitudes should also be included in the required table.

#10-3 Why the geographical coordinates of, e.g., M6 earthquakes are given with an accuracy of up to 7 or 8 digits after the decimal point in view of a considerable extension of the fault plane?

#11 lines 119-120

#11-1 Three columns of v_{S30} values are given in the data set: vs30_m_sec, vs30_m_sec_WA, vs30_m_sec_shakemap. These designations would also have to be used in the two rows for clarity as to what is meant. Use a capital S, as it comes from S-wave.

#11-2 It would be important to add the uncertainties of the v_{S30} values. These differ in part significantly for the three types in the data set. In view of the differences in v_{S30} according to different types of determination at a measuring point of more than 500m/s² in the data set, this information is more or less useless without explicit error information.
If it is not possible to state the uncertainties for all measurement points and all types explicitly, at least summary estimates for certain data groups should be provided for the users of the data set. For future applications, such uncertainty assessments will play an increasing role.
In this context, the question arises why the v_{S30} in m/s are given in the three columns (vs30_WA, vs30_shakemap, vs30_MDP) with an accuracy of up to 6 or 7 digits after the decimal point (!). How does this supposed accuracy relate to the actual uncertainty of these quantities?

#12 lines 122

With regard to the distance between the location for which an intensity value is valid and the measuring point of the strong ground motion station (site_to_MDP_distance_km), there can only be rough estimates. An intensity value can only be representative for a locality (up to the size of a medium-sized town), which can have a considerable spatial extension. Is the edge of the town then used to determine the distance or its centre, or what? In any case, the uncertainty of the distance parameter also plays a decisive role here. It is essential to specify this uncertainty.
In the published data set, the distance in km is given with 9 digits after the decimal point (!) - and this with regard to a distance from a measuring point to an area of a town with an extension in the order of a few to several kilometres. Is there a reason for the remarkable accuracy of 9 decimal places?

#13 lines 125-126

Here it is stated: " . . . the maximum between the two horizontal components of the peak ground motion measures . . .".

#13-1 What do you mean with "the maximum between the two . . . components"?
Could it be "the maximum difference between the two components"? Or what?
Or perhaps "the maximum of the two components"?

#13-2 Measurements of PGA, PGV and calculations of derived vales like SA are connected with uncertainties. These have rarely been used so far, but will be of importance in the future. Since the already published data set covers a relatively large time span from 1972 to 2016, at least qualitative information on uncertainties would be unavoidable. It makes a big difference whether data from an early period of strong ground motion measurements are considered (PGA from low dynamic analogue instruments, later somehow digitised) or modern data (PGA measured with 24 bit digitizers). It is obvious that the uncertainty level has changed radically in such a time span.
This should be taken into account at least as a discussion.

#14 lines 127-129 (macroseismic data)
#14-1 It is recommended that the already published data set be structured in such a way that, in addition to MCS intensities, those according to EMS-98 can also be included. On the one hand, this would ensure an opening with regard to other parts/countries of Europe where the EMS-98 is routinely used. On the other hand, such an extension could take into account the current developments in Italy to increasingly use the EMS-98 in macroseismology. This is expressed in the fact that more recent earthquakes are increasingly being processed using the EMS-98. Previous earthquakes, such as the 1976 Friuli earthquake of 6 May 1976, have been re-evaluated using the EMS-98 (Tertulliani et al. 2018; BGTA). Recent ground motion-to-intensity conversion equations (i.e., the application domain of the reprint reviewed here including dataset) also use intensity data in Italy in terms of EMS-98 (Zanini et al. 2019; Eng. Struct.) with data from 1983-2016.
It is true that, according to Musson et al. (2010), intensity assessments according to MSC and EMS-98 are comparable in principle, but in detail and regarding concrete MDPs, some differences become clearly visible (cf. Tertulliani et al. 2018). So, there would actually be no reason to ignore the described development of also using the EMS-98.

#14-2 v_{S30} should, according to the original definition, be a point information, i.e. the borehole location at which v_{S30} was determined. If, on the other hand, a single value is given for v_{S30} that is representative for the area of a medium-sized town, the error range of this information must not be neglected under any circumstances. This is because in many cases a considerable areal variation of this parameter is observed in the region of such a town. In the data set, v_{S30} is nevertheless also given here with an accuracy of 6 digits after the decimal point.

#15 line 133
Fig. 1 (p. 7) is mentioned here. In the figure, the assigned intensities are given for 3-4 to 10-11. These should therefore be intensity specifications, as it says in line 92, for the cases in which a clear intensity assignment for one degree of intensity in the form of an integer value is not possible. Why are only these uncertain indications shown graphically and not the intensities of the integer values from 4 to 10? These intensities are also more frequent in the data set than the uncertain ones with e.g. 6-7. For reasons of practicability, these can be shown with e.g. 6.5, but it should be clear what we are dealing with such a notation: an uncertain indication as a proxy for 6-7 but in no case an intensity determined exactly to half a degree.
Fig. 1 should be changed accordingly.

#16 line 135
Here is referred to Fig. 2. The same applies here as for Fig. 1 (cf. #15).

#17 line 177
It is said: „ . . . small intensity values (i.e. in the range $3 \leq MCS \leq 3.5$)“. Here again the at least implicit use of the integer values of intensity as a decimal quantity occurs. There exist in that case only the integer values of 3 and 4. If the observational data are so poor that it can be both 3 or 4 degrees, one writes 3-4. But what does the MCS scale provide for any values in between?

#18 line 186
A data set with magnitudes of 4.2-6.9 is mentioned here. In the abstract, the data set starts at 4.0. Which is correct?

#19 line 187-188
It is said: „ The dataset can be used as reference to benchmark studies seeking correlations between ground motion parameters and MCS macroseismic intensities". That is correct. However, the reader would have liked to see graphs showing exactly these data points. Preferably supplemented by the empirical adjustments based on earlier studies from Faenza & Michelini (2010) to Gomez-Capera et al. (2020); i.e. the graphical representations of the derived empirical relations. This would give a first, albeit only visual, impression of how the new data set behaves. Such a supplement would be very useful.

#20 line 198-199
The use of the data for the determination of Intensity Prediction Equations, as it is called here, should hardly be possible without precisely defined magnitudes in the data set, i.e. not knowing what type of magnitude we are dealing with in individual cases. Compare earlier comments on the specification of magnitude types.

Technical Corrections

#21 line 67

"ESM" needs also to be explained in the main body of the text, not only on the abstract.

#22 line 79 ff.
Mean shear wave velocity of the uppermost 30 m is given as "V_{S30}":
However, the derived physical quantity of velocity is abbreviated with a small v according to ISO (International Organization for Standardization). Unlike PGV, the small v has therefore be used for v_{S30}. A change according to the ISO standard is recommended.

#23 line 79 and other occurrences below
With regard to Eurocode 8, it is better not to quote "Code (2005)" but "Eurocode 8 (2005)". Further below, EC8 is used; however, without explanation of the abbreviation.

#24 line 158
Here it is given "Fig ??b". Insert the appropriate number.