Interactive comment on "INGe: Intensity-ground motion dataset for Italy" *by* Ilaria Oliveti et al.

Antoine Schlupp (Referee)

antoine.schlupp@unistra.fr

A) General comment.

The benefit or interest doing such a dataset is clear and well explained. Personally, I am convinced as we studied macroseismic and instrumental correlation in the past (Lesueur et al. 2013, ref that could be add as very close to the paper approach, see full reference here under). In that work we faced the difficulty to have numerous seismic stations and macroseismic information separated by short distance to be able to compare them. In the submitted paper and dataset, the very near common observations, less than 3 km and many times less, is great for further studies. Therefore, I support the publication and the built dataset that will save huge time for future users. The paper is well written and explains well the aim of the dataset. I have nevertheless few questions to clarify the text and quite a list of comments or questions about the dataset

itself. They should be checked or clarified before publication, despite the format of the publication is with a short length.

Lesueur, C., M. Cara, O. Scotti, A. Schlupp, and C. Sira. Linking ground motion measurements and macroseismic observations in France: a case study based on accelerometric and macroseismic databases. Journal of Seismology, 17 (2) pp. 313-333, doi 10.1007/s10950-012-9319-2, 2013.

B) Specific comments:

1) There is one main point that must be clarified for a good use of the dataset. The station is placed at a very precise point. But the MDP location corresponds to a "a pair of geographical coordinates matching the average macroseismic intensity value of a more or less large area with a point". How the user will know the extent of the area, is it the whole urban area of the city? If the locality is large, this point could be at more than 3 km despite the station is inside the city. Did you exclude the MDP in that case? This method probably aggregated sometimes areas with a part at rock and a part at "site effect" but the whole zone will be affected to one point. How the user will know if the "location point" calculated is consistent with the whole area? If the calculated point will be at rock when the most area is at sediment, how can we correlate the measurements? Does the VS30 at the MDP corresponds to that specific point or to the average VS30 of the whole area used to determine the intensity? You said at 105-106 that you "check of geological and topographical condition match". Is it for the whole area or for the "reduce" location at a point? As user, I would like really to know that before using the dataset and making conclusions on the results.

2) Line 100: Please clarify in the following sentence what was not reliable, it is not clear for me. " To generate our dataset, we have extracted all the MDPs from DBMI15 corresponding to earthquakes listed in the ESM flatfile and that have not been listed in the Macroseismic Bulletin since this latter data source has been proved to be largely unreliable Âż. You mention also "unreliable" in line 58, but with no explanation. Could

you precise it shortly?

3) The criteria of 3 km is very interesting. But the variation in ground motion with distance (site to MDP distance) is higher at short epicentral distance then long epicentral distance. The criteria could be too large for location very near epicenter (you have very short epicentral distances) and too restrictive for large distance (for example 100 km or more). I understand the need to have one distance value but is seem more important for me to know if the local "site effect" are similar or different between station and MDP. See first specific comment. I advise to include a comment on that point in the text.

4) line 20 to 24: The citations (DYFI of USGS, INGV, EMSC) seems to be related to Italian territory. It should be specified (they are other numerous and high quality "DYFI" procedure in other European countries)

5) line 29: several observatories are already producing shakemaps that include instrumental and macroseismic data (USGS, BCSF at www.franceseisme.fr and many others). They are directly in the topic of the paper bringing together these two types of data. Include at least the information would support the paper, with some citations.

6) line 95 to 99: I did not understand if you are using these data with high uncertainty at the end. Could you clarify it ?

7) line 108: your selection is on the closest one, not on sites conditions. It is your choice well explained here but why not including all of them within the 3 km distance? It would give an idea of variability within this 3km distance.

Comments or question about the dataset:

- the number of decimals in most of the dataset is not homogeneous and sometimes they are 10 and more decimals which has no meaning (for example event latitude at a precision of less than one millimeter? distance between epicenter and MDP much lower than one millimeter? ...). It should be homogenized and with a number of decimal consistent with the uncertainties on the value.

- the dataset does not include any uncertainty on the value. This can decrease the use of the dataset because the user will need to correlate this dataset and the original catalog (with the event ID) for instrumental data which is in the opposite of the aim of this work (at least for ev_lat, ev_long, ev_depth, Magnitude)

- the stations are called "site" or "station" or "st" in the header line. It would be better to homogenized to avoid misunderstanding.

- there is a lack of explanation of each column or header meaning. I did not find what is vs30_m_sec_WA for example. It should be available with the dataset.

- the type of distance to epicenter is not explained. Or in other words, are you able to specify when it is Repi and when it is RJB ?

- What does mean the starts behind EC8 code. Not reliable? In that case why?

- a column with "MDP epicentral distance" would have been appreciated even they are not far from instrumental data.

- What is the magnitude type ?

- you have "distance_km" with values of 0 which are not consistent with epicenter location and station location. Is "distance_km" the epicentral distance of the station?

- you have many type of Vs30 values, it would help by specifying which one is measured on the field (I speculate from the value that it is this one "vs30_m_sec")

Some of the requests in this list can be found from original seismic catalog but it is really pity that it is not available directly in this dataset.

C) Technical corrections - line 77 you use "s" and line 78 "Hz" for the "range", I would advise using the same scale.

- Line 94 you do not use Roman number for intensity, and you put one decimal. For Intensity we should always prefer roman number. Same comment for lines 96 and 97.

- line 99: need a ")" after 2007

- line 158, specify the figure number

- line 250: where can we find the 2011 version of DATABASE MACROSISMICO ITAL-IANO, could you add a link ?

- line 271: Could you precise where do we find this version? Is it this version found on web that should be refered as following in that case ? De Rubeis, V., Sbarra, P., Tosi, P., and Sorrentino, D. (2019). Hai Sentito Il Terremoto (HSIT) - Macroseismic intensity database 2007-2018, version 1, https://doi.org/10.13127/HSIT/I.1

- figure 3: variation in magnitude legend not progressive (M4 very small / others)

- Figure 4: I think that "No Observation" means "observation number". You want to say "number of observations" that should be written "Nb. Observations".

- figure 5: for clarification, I would add in the vertical axis MDP => "Vs30 (ms-1) at MDP" or like in the dataset "Vs30_MDP"

- figure 7. The boundaries between proportion is not easy to see with the color used. Add for example a black line in-between. And correlation between color and legend could be impossible for some readers. Either put a label for each "piece of cake" or change color progression.

- figure 10: same as previous comment about epicentral distance. Do you really have 0km distance? With the dataset, it seems not (using event coordinate and station coordinate).

- figure 11: "ShakeMap grid" is from Michelini et al. 2020 ? If yes precise it at least in the figure legend text.

Interactive comment on Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2020-372, 2020.