

This article proposes a methodology to detect, through a probabilistic approach, the gaps in the information about the macroseismic effects in a number of selected sites (228 caused by historical and recent earthquakes (between 1000 and 2022, according to CPTI15-DBMI15). Therefore, its purpose is not a direct contribution to the understanding of the natural hazard rather the improvement of the data needed to estimate it. I think the article is well written, in a clear way. Its aim is coherent with the methodology and with the structure of the text.

I suggest a few changes. Some of them are small corrections, suggestions and mere style points: at the end, you can find a list of them. Moreover, I would suggest some points to add in the exposition.

- Line 110 and in the following: I would like to read a short seismological explanation about the exclusion of deep earthquakes. In the article, there is only a technical reason (“the calibration dataset of the used IPE does not consider these events.”), but I think that there would be an interesting reason from the macroseismic point of view. On the other hand, I’m not sure it is necessary to say how the exclusion with the CPTI15 code was carried out. Perhaps these lines can be spared.

- Line 132. At the end of Paragraph 3, would it be possible to insert an example that shows how the method works?

- Line 161. (Paragraph 5). Why 20km? Perhaps it would be useful to explain this distance. If not, could it be enough to indicate a reference: Antonucci 2021?

- Moreover, a question: is a single location with macroseismic observations within 20km sufficient to apply the method?

- Line 169. It says “we estimated the probability of having an undocumented intensity value at each considered site”: that is, of each of the other 224 sites? Will these data be published? Maybe in the supplements?

- Line 170. Is it possible to add a few lines to explain how the four sites were chosen, or what criteria were applied to do it?

- Line 221 and the following(Paragraph 5.2. “Geographical distribution...”). I wonder if it is possible to add time-windows. For example, Fig. 4XX. “Number of undocumented effects at each selected site with probabilities  $\geq 75\%$  of reaching or exceeding intensity (a) 9, (b) 8,(c) 7 and (d) 6 MCS, before 1500; before 1800; before 1900; during XXth and XXI century. Could it be possible to do the same thing with Fig. 5?

- Line 277 (Paragraph 6). I think it might be interesting to add the relationship between the number of documented effects of the sites examined and the number of “probable” undocumented effects. In other words, is it possible to say what is the relationship between the documented and the probable effects in each examined locality?
- Last question: has any test of the procedure been done?

### List of small corrections

line 55. “...on the Italian ~~area~~.”

“...on the Italian territory.”

line 71. “whereas the ~~latter~~ provided access to 346...”

“whereas the former provided access to 346...”

line 77. “...and most of ~~intensity data~~ (i.e...”

“...and most of IDPs (i.e...”

line 84-85. “The intensity data contained in DBMI15 are expressed ~~in terms of intensity degrees~~, mostly but not exclusively in the MCS macroseismic scale, as Arabic numbers...”

“The intensity data contained in DBMI15 are mostly expressed in the MCS macroseismic scale: see Locati et al. (2022) for details.”

line 112. “The ~~selection~~ is based on...”

“The exclusion is based on...”

line 248-249. “... was computed at ~~nearly the totality~~ of 228 considered sites, except for 75 sites mostly located in the same low seismicity areas.”

“... was computed at most of the 228 considered sites, except for 75 sites mostly located in the same low seismicity areas.”

line 276. “This ~~approach~~ provides...”

“This method provides...”

line 281. “...through a probabilistic ~~approach~~ that...”

“...through a probabilistic procedure that...”

line 288-289. “... greater than the 95% ~~at almost the totality~~ of the selected sites (i.e. 173 out of 228).”

. “... greater than the 95% at 173 out of 228 of the selected sites (i.e. almost the 76%).”

line 419. “Sieberg ..., ~~1932.~~”

“Sieberg ..., 1930.”