

Comment on **essd-2022-103**

Mathilde Sørensen (Referee)

Referee comment on "The European Preinstrumental Earthquake Catalogue EPICA, the 1000–1899 catalogue for the European Seismic Hazard Model 2020" by Andrea Rovida et al., Earth Syst. Sci. Data Discuss., <https://doi.org/10.5194/essd-2022-103-RC3>, 2022

General comments:

This paper presents an updated pre-instrumental earthquake for Europe, EPICA, covering the time period 1000-1899. The dataset is developed through robust procedures and is well documented in the paper. Whereas the dataset is still expected to be highly incomplete, it represents the best possible dataset based on the data currently available. It furthermore opens the opportunity for future updates when new data become available.

The dataset is of high importance to the scientific community, documenting the seismic history of Europe. Besides being interesting in itself, information on past earthquakes is a fundamental requirement in seismic hazard studies where it is used to derive probabilities of future earthquakes.

The paper is mostly well-written and provides the necessary information about the dataset and how it has been developed. I suggest some minor revisions to improve the clarity and readability of the paper. Those are outlined in the following.

Specific comments:

- The language is generally good and easy to follow, but there are some cases of very long or unnecessarily complex sentences. I give some examples under "Technical

corrections”.

- Figure and table captions are generally very short and provide only the most basic information. To help “superficial readers”, I would suggest adding a bit more detail to the captions. The authors should also make sure that all figure content and abbreviations are explained in the captions.
- It is difficult to see the data in Fig. 2 with the symbology chosen. One options could be to use a less prominent symbol for events with 0 MDP, for example a small, red dot.
- The paragraph describing Fig. 3 (p.6-7) is difficult to follow. Adding a short introduction, framing the description of the time distribution of sources, may help. Also, the difference between Fig 3a and 3b should be clearly explained. There are several grammatical errors (is/are) in that paragraph.
- On page 8, line 214 it is stated that about 180 earthquakes from various catalogues have been identified as fakes. I suggest adding a list of those events (either by expanding the appendix or in an electronic supplement), as I think this list would be of interest to the wider community.
- The first paragraph on page 11 describes to what extent the three macroseismic location procedures have been used. It would be interesting to include some general information on what has driven the choice of method in different cases.
- Table 2 lists the relations between Mw and I0 used for each calibration region. I would suggest also including the attenuation models used for each region and method.
- The description of origin times on page 13 is very short. How are origin times derived from MDPs? And what is done if origin times are available both from MDP and catalog?
- On page 14, line 331, it is explained that the weights used in the magnitude determination are arbitrary. That is in principle ok, but it would be good to include some explanation of why those values were considered better than e.g. a 50/50 weighting. Also, please explain the term “reverse weights”.

Technical corrections:

- Page 1, line 18: Coming from “northernmost Europe” myself, I find this term very unspecific – to me, for example, that would also include the Svalbard area. Whereas this is certainly a very minor issue, I would encourage the authors to use a different term.
- Page 1, line 29: The last sentence in the paragraph seems to be incomplete.
- Page 4, line 94: “sparse” seems to be misplaced.
- Page 5, line 131: “innovations” does not seem to be the most appropriate word here. Use “new developments” instead?
- Page 5, lines 156-160: As I read these sentences, new datasets contribute 273000 MDP, whereas the final dataset contains 145500 MDP. How is that possible?
- On pages 5 and 6 there are several uses of the phrasing “the XX % of ...”. Here “the” should be removed.
- Page 8, lines 211-212: As I read this sentence, 3445 of the 5073 events in EPICA has both MDP and a catalogue listing, 2066 have only a catalogue listing and 177 only have MDP. These numbers do not sum up, please clarify.
- Page 11, line 279: This sentence gives the impression that Table 4 gives information on the regions. Should be rephrased.
- Page 18, line 415: The sentence starting “In addition to...” is unclear.

- Page 20, line 428: It is stated that "...the magnitude decrease is mostly due to the consistent amount of new macroseismic data today available...". Please explain why this leads to a magnitude decrease.
- Page 20, lines 452-456: This is a very long and complex sentence. Splitting it up would improve readability.