Comment on essd-2022-37
Ignacio Jara (Referee)

Referee comment on "LegacyPollen 1.0: A taxonomically harmonized global Late Quaternary pollen dataset of 2831 records with standardized chronologies" by Ulrike Herzschuh et al., Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2022-37-RC2, 2022

Herzschuh and collaborators present a pollen dataset of almost three thousand records from nearly all major terrestrial regions of the world. The dataset includes newly-developed chronologies for all sites, an harmonisation of taxa nomenclature, and open data and code availability. Hence, this compilation represents an immense work effort by the authors.

The LegacyPollen 1.0 dataset will undoubtedly be a great contribution to the palaeoecological and palaeoclimatological community, with enormous potential for paleo-environmental and climate change studies. In addition, the methodology used to harmonise the pollen taxonomy can be used as a guide for similar compilations in the future.

The manuscript is simple, informative, and well written; and therefore I strongly recommend its publication in Earth System Science Data.

Below are some minor corrections to the manuscript. These are very slight suggestions from my personal point of view, so no mandatory amendments are needed.

All the best,

Ignacio A. Jara
Minor suggestions

Line 34. Perhaps introducing the article with a more broader sentence regarding the importance of harmonized datasets. In my opinion the relevance of this type of global repositories goes beyond the validation of earth system models, including applications for climate change, biodiversity, and land use reconstructions to give a few examples

Line 84: “...nomenclature. We set...”. I suggest connecting both sentences with “...nomenclature. For doing so, we set..”

Line 90. How did you harmonize Gymnosperm taxa? Gymnosperm, especially conifers, are important elements of many records from mid and high latitude sites.

Line 96. I suggest to replace “in the following named” by “hereafter”

Lines 99-100: “Alternatively, the back-calculation of the pollen sum could be based on more elaborated methods...”. It is not clear to me if the authors are using these other methods, please clarify.

Line 107: “The LegacyPollen 1.0 metadata of 2831 records are provided for each pollen samples”. This sentence is hard to understand. Could you be more specific. Was the information of the 2831 sites assigned to each sample or site?

Line 138. I suggest adding two commas between “among them”, so that “...cover part of the Holocene, among them,...”

Line 173. I suggest replace “included” by “integrated” or other similar term, as “included” has already been used previously in the same sentence

Line 238. “Holocene” is not a timescale. I suggest changing this sentence, using “millennial timescales” or “Deforestation during the Holocene period is...”