

Response letter of 'LegacyPollen 1.0: A taxonomically harmonized global Late Quaternary pollen dataset of 2831 records with standardized chronologies'

Response to comments of Referee #2_Ignacio A. Jara

1. General comments

Reviewer comment: *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

Response: Thank you for your review, and the manuscript has been adjusted according to your suggestions.

2. Minor suggestions

Reviewer comment: (1) 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.

Response: Thank you for pointing us to this. We revised the text in the introduction.

New text (lines 31-33): “ Broad-scale palaeo-proxy databases provide important opportunities for making comparisons of palaeoenvironmental synthesis studies and for palaeodata-model validation, where harmonized data processing is the foundation (Gaillard et al., 2010; Cao et al., 2013; Trondman et al., 2015).”

Reviewer comment: (2) Line 84: “...nomenclature. We set...”. I suggest connecting both sentences with “...nomenclature. For doing so, we set..”

Response: We have adopted your suggestion.

New text (lines 77-78): “ First, we standardized the taxon nomenclature. To do so, we set up a master table containing all pollen taxa names from the 2831 records and made names consistent...”

Reviewer comment: (3) Line 90. How did you harmonize Gymnosperm taxa? Gymnosperm, especially conifers, are important elements of many records from mid and high latitude sites.

Response: We apologize for this and revised this part of the text.

New text (lines 83-85): “ Second, we harmonized the pollen taxa according to the classification of the Angiosperm Phylogeny Group IV system (APG IV; The Angiosperm Phylogeny Group et al., 2016) and the Gymnosperm Database (<https://www.conifers.org/>).”

Reviewer comment: (4) Line 96. I suggest to replace “in the following named” by “hereafter”

Response: We have adopted your suggestion.

New text (line 89): “ Although most pollen records contain the count data (‘raw’ data hereafter)...”

Reviewer comment: (5) 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.

Response: We deleted this part of the text.

Reviewer comment: (6) 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?

Response: Yes, each pollen sample of 2831 sites includes its site information. We also provide a new table, containing an overview of the 2831 site information, to PANGAEA.

New text (lines 99-104): “ The metadata for each site in the LegacyPollen 1.0 dataset includes the following: Event (PANGAEA dataset identifier), Data Source, Data Type (raw or digitized), Site ID (in the source datasets), Dataset ID (in the LegacyPollen 1.0 dataset), Site Name, Location (longitude, latitude, elevation, and continent), Archive Type (e.g., peat, lake sediment core), Site Description (from original publication/Neotoma), and Reference. All site-specific metadata are available at PANGAEA (<https://doi.pangaea.de/10.1594/PANGAEA.929773>; Herzschuh et al., 2021a) in the “Further details” section (“Site metadata of LegacyPollen 1.0 dataset.csv”).”

Reviewer comment: (7) Line 138. I suggest adding two commas between “among them”, so that “...cover part of the Holocene, among them, ...”

Response: We have adopted your suggestion.

New text (lines 127-129): “Almost all records (97.8%) cover part of the Holocene, among them, 65.2, 79.5, and 89.5% cover the early Holocene (11.7–8.2 cal. ka BP), middle Holocene (8.2–4.2 cal. ka BP), and late Holocene (4.2–0 cal. ka BP), respectively.”

Reviewer comment: (8) Line 173. I suggest replace “included” by “integrated” or other similar term, as “included” has already been used previously in the same sentence

Response: We have adopted your suggestion.

New text (lines 158-159): “To our knowledge, LegacyPollen 1.0 is the largest harmonized fossil pollen dataset including more than twice the number of records integrated in previously published datasets...”

Reviewer comment: (9) Line 238. “Holocene” is not a timescale. I suggest changing this sentence, using “millennial timescales” or “Deforestation during the Holocene period is...”

Response: Thanks for your suggestion, we have revised the text.

New text (lines 218-220): “Deforestation during the Holocene period is of particular relevance which, with the help of the LegacyPollen 1.0 dataset, can now be investigated at the hemispheric scale.”