Response to comments of Anonymous Referee #1

Summary

Cao et al. describe a new database of modern lake sediment pollen samples from the central and eastern Tibetan Plateau. Additionally, the authors utilize two statistical analytical metrics, weighted-averaging partial squares and random forest, in order to determine the climatic controls to pollen distribution. Additional attention is needed in small but key areas including, minor clarifications within the methodology and overall MS structure. Beyond these minor changes, this paper and accompanying dataset will compliment existing pollen datasets on the Tibetan Plateau as well as provide useful data for paleoclimate studies moving forward.

Main Comments (Manuscript)

Some of the methodology is insufficient. While I can understand the necessity of road access, I would be interested to see whether a GoogleEarth or similar initial scouting of the area would produce a more even distribution of sample sites. Additionally, there is mention of long-distance pollen transport within the data description section (lines 173-176). However, no explanation how this is determined is given. You mentioned that bodies of water directly connected to rivers were avoided but did not go further. Please clarify this assertion. Within the methodology section, there is no discussion regarding sources of error (sampling, etc.) or how they are treated. Please clarify.

Response: The Tibetan Plateau is a quite remote area and road access is essential for collecting samples. In addition, the south-west part of the study area is mountainous with few lakes, hence our modern pollen dataset still has a geographical gap in the south-west part. We explain the reason for the uneven distribution in the new version. We also explain how we attempted to minimise the long-distance pollen transport by selecting small lakes. Generally, pollen grains are identified and counted under an optical microscope. To ensure a reliable representation of the entire pollen assemblage by the counted pollen data (reducing the error of pollen analysis), an adequate number of pollen grains and *Lycopodium* spores should be counted for each sample, which we now discuss.

Line 89-98:

"*To reduce the influence of long-distance pollen transported by wind and rivers, small and shallow lakes (or pools) with less than 100-m radius and without long inflow rivers (n=117) (locally sourced pollen grains are the dominant components for small lakes; Sugita, 1993) were selected to collect pollen samples (Figure 1). To reduce the influence of the lake-shore vegetation component, the lake surface-sediment samples were collected from the central part of each lake, with the top 2 cm of lake sediment forming the sample (Tian et al., 2008). Although the selected lakes generally have an even distribution, there is still a gap in the south-west part of study area because of a lack of lake and road access (Figure 1).*"

Line 108-111:

"More than 500 terrestrial pollen grains were counted for each sample, and more than 200 Lycopodium spores were counted for most of the samples (mean=270 grains; median=480 grains), both of which ensure a reliable representation of the entire pollen assemblage by the counted pollen data."

Next, it would make the MS easier to read if subsection "3.3 Data Description" were its own section (e.g., 4 Results). Finally, while overall the use of English is good, there are minor yet reoccurring instances of run on sentences that should be addressed.

Response: We modified the subsection "3.3 Data Description" as an independent section "4 Data Description" in the new version. And the written English was polished by a native English speaker.

Main Comments (Data)

The database is easily accessed and downloaded. The database itself is a two page Excel file that clearly presents the individual counts of pollen as well as their overall percentage at each site. Providing both counts and percentages is a good touch. However, adding errors to the sample meta data (Elevation, Mt_{co}, Mt_{wa}, T_{ann}, and P_{ann}) would be useful for future users of the dataset.

Response: The climate data were obtained from the Chinese Meteorological Forcing Dataset (He et al., 2020, Scientific Data) and in the original publication, He et al. describe data quality and error. Hence, we decided not to add that into our manuscript.

Minor comments

line 154: Appendix 2 does not exist, it seems to be referencing Appendix B, please clarify.

Response: Corrected. "Appendix 2" should be "Appendix B".

Line 154-155:

"*Our final RF model includes 19 pollen taxa (Appendix B), which all make a positive contribution to the precipitation distribution.*"

Please double check section numbering once structure is fixed.

Response: Done.

Figures:

Figure 1) Please provide rain fall scale for isohyet map. Make 'a' and 'b' more clearly visible. Also Line 67 references Figure 1, but Figure 1 does not have relevance to elevation discussion in this sentence. A 3rd subfigure in Figure 1 with elevation of the region would help.

Response: Done. We have modified the two subfigures and add a new one with an elevation map.

Figure 2) The choice of colors within the figure are not clear, please explain in figure caption.

Response: Done.

Page 9:

"***Figure 2*** *Pollen diagram showing the major taxa (percentage; %) of the 117 samples arranged by mean annual precipitation (P_{ann}; mm). Pollen with red bars are positively related to P_{ann}, those with blue bars are negatively related to P_{ann}, while the relationship is insignificant for those with green bars.*"