Author Response to Referee #2

A global zircon U–Th–Pb geochronology database

Yujing Wu et al.

Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2023-20

RC: Referee Comment, AR: Author Response

Dear referee,

Thank you sincerely for your response, and for dedicating your valuable time and effort to reviewing both the manuscript and the dataset. We deeply appreciate your insightful advice and concerns. We will meticulously address each of your points in the revised manuscript and updated database. Please find our point-by-point reply below.

Kind regards,
Yujing Wu (on behalf of the author team)

Comments and responses:

RC: General Comments:

Wu et al. describe an updated compilation of zircon U-Th-Pb ages from journal articles and dissertations. A previous compilation focussing on Chinese geochronology (Wu et al., 2019) has been expanded to include additional samples from across the world. A recent publication in Earth-Science Reviews (Wu et al., 2022) also describes this database. The dataset is published as two excel spreadsheets on Zenodo, with a third document containing the source references.

Whilst the compilation of data from ~12,000 papers is a commendable effort that could support diverse future research, the database presented here lacks important additional information that would allow quality assessment and control, such as more details on the analytical method and age correction. The original data sources should be included in the manuscript reference list. Beyond the description of the dataset, the manuscript further contains scientific interpretations and discussions that go beyond the scope of Earth System Science Data and would require rigorous, additional scientific review.

I cannot recommend this manuscript for publication in its present form due to several concerns detailed below. I would be willing to review the data description again if these concerns can be addressed, however, I recommend the scientific discussion (Sections 3 to 4.4) be removed from the manuscript.

AR: General Responses:

We greatly appreciate your comments and feedback. We will carefully revise the

manuscript and make the necessary updates to the database to address your concerns. Please find below our general response to your concerns.

Firstly, to ensure quality assessment and control, we will enhance our Zenodo repository by including zircon reference materials such as 91500 and GJ-1, which were used for age correction. Additionally, we will provide isotopic ratios ($^{206}Pb/^{238}U$, $^{207}Pb/^{235}U$, $^{207}Pb/^{206}Pb$, and $^{208}Pb/^{232}Th$) along with their associated uncertainties. By doing so, readers can use this zircon database to calculate new ages as needed. In addition, it is important to acknowledge that null values may exist due to the absence of provided information in the original literature. Despite these null values, the remaining data remains useful. In the revised manuscript, we will include detailed statements regarding the presence of null values to ensure transparency and clarity for readers.

Secondly, we would like to clarify that we prefer to include the source references in the Zenodo repository rather than in the manuscript itself. Because the original data sources encompass more than 10,000 papers, and including them in the manuscript's reference list would extend it to over 400 pages. By placing the source references in the Zenodo repository, we can still provide readers with access to the references and in the meantime respect previous academic achievements.

Thirdly, we are willing to make some reductions in the out-of-scope scientific interpretations and discussions. Our initial goal was to introduce the characteristics and potential value of our database, but we may have provided excessive detail. To address this, we will remove Sections 4.1 to 4.4 and instead summarize the content in one or two paragraphs. Additionally, we will make efforts to condense Section 3 as much as possible to ensure that data description is the main purpose of the paper. However, we still want to retain an abridged version of Section 3 in the manuscript since it presents the fundamental characteristics of the zircon data clearly and intuitively.

RC: Specific Comments:

RC: It is unclear how much of the presented data compilation was already included in Wu et al. (2019) and Wu et al. (2022) vs. what has been added since. What is the added value of this present database that it should not just be an updated version of the previous publications?

AR: Thank you for your comments. We will detail the parts of the database used in previous publications and clarify the newly added components in the revised manuscript. Since previous publications were conducted during the data compilation without disclosing the data, we treat this database as the initial publicly available version, rather than an update.

RC: Quality assessment/quality control: there is very little description of the curatorial procedure during compilation of the dataset; e.g. information on the recalculation of uncertainties (if any) where sources are inconsistent or on how lithologies were assigned (curatorial decision or is this information contained within the data sources?). Furthermore, very little metadata is provided that would allow others

(including myself) to assess data quality.

AR: Great point! In the revised manuscript, we will include a dedicated section that provides a detailed overview of the curatorial procedure. Please find below the response to the questions you listed:

Firstly, we did recalculate some uncertainties to normalize the errors to standard deviation. The original references contained uncertainties in various forms, such as relative uncertainty and 2 standard deviations. We have developed specific processing methods to handle these different forms, and we will provide comprehensive details on these methods in the revised manuscript.

Secondly, we categorized the lithology into three groups (sedimentary, igneous, and metamorphic) based on the information from the original literature. In instances where the data sources did not provide any lithology information, we left the "Lithology" field empty. We will explicitly state the proportion of null values in the revised main text.

Thirdly, regarding quality control, we have checked twice to ensure that the collected data was consistent with the information provided in the original literature during the database construction. The zircon data were directly collected from the original literature, and we made no changes to the data itself, only standardized the forms to preserve the original content as much as possible. We are confident about the authenticity and consistency of our data, and welcome others to do quality check by comparing the data in the database with the data in the original reference. The information of the references is provided in the Zenodo repository. However, the quality of the geochronology data itself (e.g. how age correction was applied and how uncertainty was derived) depends on the original references. Interested researchers can conduct in-depth studies based on the reference file we provided. As mentioned earlier, we will update our database to include isotopic ratios and uncertainties (including $^{206}Pb/^{238}U$, $^{207}Pb/^{235}U$, $^{207}Pb/^{206}Pb$, and $^{208}Pb/^{232}Th$), as well as zircon reference materials (such as 91500 and GJ-1) for age correction. This will enable researchers to have the option to recalculate zircon ages or do necessary corrections in their own way instead of relying solely on the ages provided in the original literature. This aligns with our goal of providing researchers with a comprehensive data compilation for further investigation.

RC: Inconsistency of data:

1. The “Method” field mixes analytical methods with instruments; sometimes only a reference is cited. These should be separated and you should use a controlled list for both the analytical methods and the instruments: for example, there are >10 different spellings for ICP-MS. What is the difference between null values and those labelled “unmentioned”?
2. Fig 1, and the text in general, gives the impression that you have location information for all records. However, coordinates are missing for many entries in the data sheets.
3. The reference file should also include DOI, title, name of co-authors to guarantee unique identification of the data source. These citations should be

included in the reference list to this manuscript.

AR: Thank you for your comments. Please find below a point-by-point response to the questions you listed:

1. Sorry for the confusion. This confusion arises due to that the initials for both the analytical method ("Mass Spectrometry") and instrument ("Mass Spectrometer") are "MS". We will replace the "Method" field with "Instrument". Second, there are different spellings for ICP-MS because the original literature wrote that way. We would like to keep these different spellings ("different" instruments) to provide researchers with more options. Third, we will further clean the "Instrument" field, such as addressing the "unmentioned" label and cited references.

2. We will give a clear statement on the null values of location information. Null values are inevitable because the original literature didn't provide associated information. However, the rest information is still helpful for some studies.

3. We will add DOI and title in the reference file. We need to declare that some dissertations and old papers (especially papers in Chinese journals) don't have a DOI. But, the unique identification of the data source can still be guaranteed by other information we provided. We prefer to put the source references in supplementary materials or the Zenodo repository instead of the manuscript. Because the original data sources have more than 10,000 papers, which will take more than 400 pages if put in the manuscript reference list.

RC: Sustainability of the database: is this a curated database that will be maintained and updated? If so, over what timeframe will it be maintained? If not, have there been any attempts to integrate your work with existing, curated compilations such as those of EarthChem (https://earthchem.org/), GEOROC (https://georoc.eu/), Martin et al. (2022, https://doi.org/10.1038/s41597-022-01730-7 and https://doi.org/10.25625/FWQ7DT)?

AR: This is a good point. At present, our Zircon database is not a curated database, and we do not have a specific maintenance and update plan for it. While the maintenance and update work is a possibility that may be considered in future projects, we believe that the current format of the database does not hinder its research potential.

We have not considered the integration of our database with other compilations yet because we prefer that our database maintains its independent existence. The EarthChem and GEOROC are undoubtedly great curated compilations. However, the Zenodo repository is also a good platform for promoting open science and sharing our database with the research community.

RC: Incomplete referencing:

1. Of other zircon geochronology compilations (e.g. EarthChem, GEOROC, Martin et al., 2022). How much overlap exists to these previous compilations? Equally, how many data are missing?

2. Of scientific literature, including statistical treatment of oversampling/sampling

bias, which should be applied to your database before any geological interpretations are drawn (e.g. Keller & Schoene, 2012: https://doi.org/10.1038/nature11024; Mehra et al., 2021: https://doi.org/10.1130/GSATG484A.1)

AR: We appreciate your comments. Please find below a point-by-point response to the questions you listed:

1. Our zircon database is independent of other compilations. We began the data construction in 2017 and we didn't refer to other compilations when constructing our own. Since the forms of the databases are different, it is difficult to compare them one by one to check the overlap or missing data. There might be some overlap since we might collect the same literature. Nonetheless, our database does have unique advantages. For example, we collected a large amount of data in Chinese literature, which is difficult for non-Chinese scholars to obtain. We believe the diversity of databases can provide more options for future research.

2. We will address the sampling bias in the revised manuscript and use the new results for geological interpretations. In the meantime, we want to keep the results using raw data in the main text or supplementary materials for reference because they display the original characteristics of the zircon data. It is also possible that researchers can use our raw data to explore more advanced ways to deal with the biased sampling issue in the future.

RC: The discussion & scientific interpretation are very superficial, with language that is both too informal and very pompous. Previous work on this topic is not discussed in sufficient detail. As this is a submission to ESSD, I believe that much of Sections 3 and 4 goes beyond the scope of a data journal and could be removed. My recommendation would be to instead focus primarily on Section 4.5 and ensure that discussion of previous literature in this section is comprehensive, detailed and accurate.

AR: Sorry for our sketchy discussion and poor language. We didn't write the discussion and scientific interpretation in detail to avoid extensive interpretations of data. Perhaps there was a little deviation in our understanding of the scope of ESSD. As you suggested, we will focus on Section 4.5 in the revised manuscript, adding comprehensive, detailed, and accurate discussion. We will remove Sections 4.1 to 4.4 and summarize the content in one or two paragraphs instead. Additionally, we will abridge Section 3 as much as possible to ensure that data description is the main purpose of the paper. However, we want to keep an abridged Section 3 in the manuscript because this section is necessary, which intuitively presents the basic characteristics of the zircon data. Finally, we will find an advanced editing service to improve the language.