

Earth Syst. Sci. Data Discuss., referee comment RC2
<https://doi.org/10.5194/essd-2020-308-RC2>, 2021
© Author(s) 2021. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on **essd-2020-308**

Anonymous Referee #2

Referee comment on "Climate-Biogeochemistry Interactions in the Tropical Ocean: Data collection and legacy" by Gerd Krahnemann et al., Earth Syst. Sci. Data Discuss., <https://doi.org/10.5194/essd-2020-308-RC2>, 2021

General comment

This manuscript gives an overview of, and describes the scientific activities and resulting datasets of the SFB754. The report contains the metadata of all cruise conducted within this project, as well as details on the methods used and the treatment of the data.

This overview is a good entry point into the vast array of data that were collected within 2008-2019 in the frame of this project and contains important information for data re-use.

Specific Comments

Page 13, line 95: Link to <https://www.sfb754.de/sfb754-osis> is misleading, as it shows everything in OSIS and one has to go to "context" – SFB754 to get the real subset, is there no direct link to SFB754 in OSIS? If OSIS is able to store the data I am missing the explanation why the data was additionally published in other data centres.

Page 13, line 111 and following (=Table 2):

Link <https://doi.org/10.1594/PANGAEA.926545> does not work, the dataset is still in review (<https://doi.pangaea.de/10.1594/PANGAEA.926545>) and it is the only one which starts with "SFB 754" (with empty space) instead of SFB754, please correct. Also in this collection quite a few datasets appear as "unpublished" and are access restricted. I would like to have more transparency on how many datasets are actually published/accessible and why some are not. (even those that are 5 years old, see e.g

<https://doi.pangaea.de/10.1594/PANGAEA.861224>)

A few collections could maybe be combined or more clearly ordered, I am not sure e.g. why the collection <https://doi.pangaea.de/10.1594/PANGAEA.926794> contains those two entries, they don't seem to contain similar data. Also, one of the entries in itself is a collection (<https://doi.pangaea.de/10.1594/PANGAEA.903023>) which makes the whole thing really confusing. Maybe reduce the number of collections and make it a little bit more general (e.g. only one collection for all datasets of the BIGO lander).

Page 26, line 398: watercontent should be "water content"

Page 27, line 425 and page 31, line 540: "analyzed" – in all other cases you write "analysed"

Page 40, line 749: "rhizone" probably is rhizome?

Page 41: In my opinion, a chapter (after chapter 4) on the most important results (with links to the respective publications) and an outlook is missing. The following questions are unanswered:

- What are the most prominent results of this project, which datasets have been combined to obtain these results?
- Did the project reach its goal, did it surpass it or is there anything left unclear?
- How can the data best be used, can it be combined?

Reference list: in the citation of the datasets you always include "Dataset.". As far as I can see this is not part of the original citation, is this intentionally done to separate them from the other references?