

Comment on Lawler et al 2021

Anonymous Referee #2

Referee comment on "The Southern Ocean Radiolarian (SO-RAD) dataset: a new compilation of modern radiolarian census data" by Kelly-Anne Lawler et al., Earth Syst. Sci. Data Discuss., <https://doi.org/10.5194/essd-2021-148-RC2>, 2021

Lawler et al. present a compilation of published and newly generated dataset radiolarian assemblages from 228 Southern Ocean sites. The dataset is significant, and it will be very useful for the micropaleontological community focused on biogeography, paleoecology and transfer functions in the Southern Ocean.

However, there are some inconsistencies, incompleteness and lack of transparency in some cases. For example, there is a reasonable argument about the criteria used for sample inclusion, which includes 25% of tropical subtropical samples, but only for some regions. Why Abelmann et al 1999 was included, if it was not counted/recounted by the authors, and others potential datasets with full census counts were not included (e.g. Piasias et al 1997; Hollis and Neil 2005? There is missing information about taxonomic harmonization for some datasets (Abelmann et al 1999), missing environmental data and erroneous information about sample preparation (and subsequent inconsistencies in the fraction used for slide preparation). For this, I recommend major revisions, and I hope the authors can improve the quality of the data set for an article in a journal of the standing of ESSD.

Major comments:

If samples from Indian Ocean were recounted to ensure taxonomic consistency, why the 44 sites were published by Abelmann et al. (1999), who is not in the author list, were not recounted too? How the authors are sure that there is a taxonomic consistency? Did the authors check somehow that taxonomy and nomenclature was consistent among the post-2000's datasets/publications? The authors should explain why they included an incomplete and apparently non-harmonized dataset in their compilation. The final dataset has 74 samples which have <100% (31% of the total dataset); 13 less than 60%. Some of the criteria used by Abelmann et al. (1999) to remove species can be justified and reasonable (rare taxa below certain threshold), although it is not optimal when you are trying to assemble a robust modern dataset. Other criteria, such as depth habitat or clarity of identification, are not so clear, and can lead to substantial biases when Abelmann et al. (1999) is combined with the other datasets which have all the species identified and sum up to 100%. If having 100% of the assemblage is not a major criteria to include samples, why the authors did not decide to include other studies which cover gaps in the distribution of surface sediment samples?

- Section 2.1.2

The authors say that 'For both the previously published and unpublished sites, slides were prepared using the procedure described in Cortese and Prebble (2015).', and describe the use of a 45µm mesh-size sieve. This is not correct. Rogers and De Deckker (2007) indicate in their material and methods section that 'samples were washed through a 63

µm sieve'. Please correct this, and discuss the implications of merging data which belong to different size fractions.

- Complementary data

The authors mention that they extracted different environmental variables at the location of the samples, but these are not included as supplementary material. Moreover, the choice of 100 m depth for extracting those variables seems a bit arbitrary. Although I can accept that higher radiolarian abundances are found at the subsurface, that is not likely true for all the regions included in the dataset. In some cases, higher abundances can even occur deeper in the water column. This section should include the raw environmental data at several standard depths, not just one. This information could be added to Table 1 or included as an additional table.

- 2.3 Structure of the SO-RAD dataset

The authors should include a column in Table 1 which indicates the source of the data (publication or new data). Is there any information about the abundance (Radiolarian/gram) or preservation (qualitative estimation) of the radiolarian assemblages at each location? That kind of information is very relevant in paleoceanography.

- 3.2 Environmental data

As indicated above, the environmental data is not included in the submission, so it can not be evaluated.

L. 184. This sentence is confusing: ...' are well removed from the direct influence of seasonal sea ice.'

- 2.4 Software

According to the ESSD guidelines, 'ESSD data products and data descriptions should include all codes, libraries, statistical or interpolation routines, model versions'. In the case of Lawler et al, they should include the R code used to generate the figures and summary statistics.

- 4.1 General limitations

One particular issue which is not even discussed in the manuscript is the quantification of the uncertainties inherited from source data, environmental and radiolarian census counts.

Minor comments:

L. 38. Subtropical Southern hemisphere assemblages are not dominated by siliceous microfossils. Subtropical (25° to 40° S).

Paragraph starting at L. 42. There are multiple sentences about feeding strategies, size and habitat in that lack a valid reference. Although radiolarian experts are familiar with those statements, new users may appreciate if the authors could include references to support those statements.

L. 53. Missing reference about carbonate preservation

L. 54. Sentence about the importance of siliceous microfossils is confusing. Importance for what, for the reconstruction of past conditions? Importance for the carbon cycle and biological pump? If that is the case, that is a very simplistic view of the impact of anthropogenic warming on World's oceans. Other studies have observed a poleward expansion of calcareous nanoplankton species relative to non-calcifying groups of phytoplankton (e.g. diatoms) (Winter et al 2014).

L. 135-138: The authors mention some naming adjustments, so-called taxonomic harmonization, updating of species concept and species grouping. However, there is no any table with information about original names vs. adjusted ones.

L. 225: There is a considerable number of samples which could be considered as tropical and subtropical. If the dataset is called 'Southern Ocean RADiolarian' dataset (although 'Subtropical is included in the discussion, L. 226), it should only include samples from the Southern Ocean (which has its northern limit at 60° S latitude). In its current state, 25% of the samples correspond to the tropical-subtropical domain. I understand the motivation of including lower-latitude samples; to cover potential scenarios of influence of tropical waters in mid-to-high latitudes locations in the past, although I doubt the assemblages in

those cases would resemble its composition to modern tropical locations. Perhaps the authors should consider renaming the dataset and using the term Southern Hemisphere, which would be more accurate. Moreover, if the authors have decided to include tropical-subtropical samples, why they did not include samples from other sectors in the Pacific Ocean (which are currently underrepresented)? Publicly available datasets by Pisias et al (1997), Hollis and Neil (2005) or the recent compilation by Hernandez-Almeida et al (2020) have samples which, geographically, would fit in the scope of the current compilation. Although perhaps they are taxonomically different, merging would be possible, as it seems it has been made for the dataset by Abelman et al (1999).

L. 280: Even if Abelman et al (1999) explain in the original publication the rules for the exclusion of certain species, would be interesting to include those here, to help to understand why Abelman's dataset was included and not other available datasets.

L. 332: I would never recommend to subset dataset by taxa living at a particular depth, in particular in the Southern Ocean where there is so little information about depth habitats of radiolarians, which may vary zonally, or in the seasonal or multi-annual time-scale. Instead it would be more convenient to include environmental data for multiple depths.

References:

- Hernandez-Almeida, I., Boltovskoy, D., Kruglikova, S. B., & Cortese, G. (2020). A new radiolarian transfer function for the Pacific Ocean and application to fossil records: Assessing potential and limitations for the last glacial-interglacial cycle. *Global and Planetary Change*, 190, 103186.
- Hollis, C.J. and Neil, H.L. 2005 Sedimentary record of radiolarian biogeography offshore eastern New Zealand. *New Zealand Journal of Marine and Freshwater Research*, 39:165–192
- Pisias, N. G., Roelofs, A., & Weber, M. (1997). Radiolarian-based transfer functions for estimating mean surface ocean temperatures and seasonal range. *Paleoceanography*, 12(3), 365-379.
- Winter, A., Henderiks, J., Beaufort, L., Rickaby, R. E., & Brown, C. W. (2014). Poleward expansion of the coccolithophore *Emiliana huxleyi*. *Journal of Plankton Research*, 36(2), 316-325.