

Earth Syst. Sci. Data Discuss., referee comment RC1
<https://doi.org/10.5194/essd-2022-129-RC1>, 2022
© Author(s) 2022. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on **essd-2022-129**

Anonymous Referee #1

Referee comment on "OceanSODA-MDB: a standardised surface ocean carbonate system dataset for model–data intercomparisons" by Peter Edward Land et al., Earth Syst. Sci. Data Discuss., <https://doi.org/10.5194/essd-2022-129-RC1>, 2022

The authors present a matchup database combining various in situ observations with associated satellite remote sensing and reanalysis products. The challenge of combining these datasets which are collected over vastly different spatiotemporal scales is handled by the novel development of spatiotemporal regions of varying size.

This technique can clearly be applied to other parameters, and the dataset has the potential for reuse.

The dataset is a novel combination of other datasets, some of which have never previously been made publicly available.

Data is accessible as described and easily downloaded without the need of registration. A subset of these data was downloaded for inspection. These netcdf-v4 compatible files were inspected using the ncd4 R library.

the "_FillValue" for the variables consists of 8-bytes which is not a valid type for R and these are coerced to a double precision value, but otherwise no issues. The files are well described in typical netcdf format and adhere to CF conventions.

The files were also tested in a python 3.8 environment using the netCDF4 package.

The dataset is missing earlier years, 1971, 1964-1967 and 1959. I presume this is due to lack of associated in-situ data, but please confirm.

No erroneous data were found and the carbonate system and associated parameters (e.g. nutrients, temperature) are shown to span appropriate ranges and be consistent. At least to the extent of my knowledge of the global distribution of these parameters.

The abstract is a little long. It includes reference to an example application which is not appropriate for an abstract and is also "not shown" - line 30. This appears to have been partially copied from the Introduction section. The article appears otherwise well written.

Each sub-source of data is well described, with collection method, calibration information and uncertainties. However, in the contributions it is mentioned that A. Polukhin assessed the possibility of using certain data due to differences in methods. These differences are not well discussed in the input datasets section so I suggest the inclusion of a short paragraph summarising the rationale for this work.

We note the AMT and Kara Sea datasets start over a decade before the publication of the Dickson et al 2007 best practice document which is given as a reference. I presume the authors are simply saying that the methods are basically the same, but it would be good to clarify.

Line 486 - when and by whom will reassess the appropriateness of the maximum ROI scales? community feedback?

The figures provide a good overview of the nature of the dataset and are of generally good quality.

Figure 1 and 2 could be arranged better, the axis labels very small considering the amount of whitespace between panels