Earth System Science Data Discussions
Open Access

Earth Syst. Sci. Data Discuss., referee comment RC2
https://doi.org/10.5194/essd-2022-129-RC2, 2022

Comment on essd-2022-129

Anonymous Referee #2

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-RC2, 2022

General Comments

Over the past couple decades, large datasets of *in situ* carbonate system measurements (e.g., SOCAT, Bakker et al., 2016) have been compiled and quality controlled, and are frequently used to train models for prediction and/or spatiotemporal gap filling of carbonate chemistry using proxy variables. However, models trained on sparse and irregularly spaced data can suffer from biases that favor more highly sampled regions or time periods. These biases can be partially alleviated by aggregating data points to bins of constant spatial and temporal extent, but binning in this way can still result in sub-optimal data division due to elongation of bins near the poles, unintentional splitting of clustered data, and unaccounted for interactions with coastlines.

To address these shortcomings, Land and coauthors compile and describe a database of *in situ* surface carbonate chemistry measurements built around regions of interest (ROIs), which are constructed so as to include as many *in situ* measurements as possible within a maximum timespan of 10 days and a maximum diameter of 100 km. This database — OceanSODA-MDB — also includes spatiotemporal matchups with satellite, model, and reanalysis datasets corresponding to each ROI. Land et al. display the utility of this newly compiled database by re-training a global algorithm from Takahashi et al. (2014) to predict potential alkalinity (PA) from sea surface salinity, displaying a global reduction in the root mean squared error between measured and predicted PA: from 15 to 12 µmol/kg in marine waters and 32 to 23 µmol/kg in coastal waters.

The ROI strategy is a compelling and creative way to group *in situ* data and match the grouped data with other datasets for algorithm training. The strategy is explained well in this manuscript, and certainly has potential advantages over fixed spatiotemporal binning. OceanSODA-MDB is easily accessible as a series of NetCDF files. It provides co-located satellite, model, reanalysis, and *in situ* data that can surely be used to re-train various algorithms that are in use today. This manuscript is a valuable contribution to the literature in its current form, but I’ve added a few comments and suggested corrections in the following sections.

Specific Comments

CO_2 system calculations only performed from C_T-A_T:

In lines 112–113 it is stated that carbonate chemistry calculations were only performed when C_T and A_T were available. Are there a significant number of cases for which other carbonate chemistry pairs were available (e.g., C_T-pH, A_T-$p$$CO_2$, etc.)? I'd be interested in the reasoning for only making carbonate chemistry calculations with the C_T-A_T pair. Similar accuracy in calculated parameters should be obtainable by pairing either C_T or A_T with either pH or $p$$CO_2$ when the measurements are of sufficient quality (e.g., Orr et al., 2018).

Creating the radial in situ data:

I think the explanation of how the data grouping is performed is great. It is clearly a complex process but breaking it down step-by-step in section 3.2 is very helpful. I am curious about the prospect of expanding this methodology to subsurface data and also how easily the methodology might be adapted to form finer resolution ROIs. I hope the authors might consider making their code publicly available at some point in the future to facilitate adaptions of their methodology to other datasets and/or spatiotemporal resolutions.

Updates to OceanSODA-MDB:

It is indicated in section 6 how the MDB could be updated. Are they any concrete plans to make updates to OceanSODA-MDB at regular intervals in the future?

Minor Comments and Technical Corrections

Line 130: The references in this line are repeated.

Line 135: Should this be dataset no. 4? Additionally, I'd be interested in knowing how the Argo data were acquired. A monthly snapshot should have an associated DOI (e.g., doi.org/10.17882/42182#95967 for Sep. 2022). If files were individually downloaded, were individual profile data files used, or interpolated Sprof files? I think this would be very helpful information for anyone who wants to replicate your methodology.

Line 141: "Dickson et al." should be outside the parentheses here.

Lines 157–161: Is it safe to say that this paragraph and the entries for these datasets in Table 1 should be eliminated since they are added to OceanSODA-MDB along with CODAP-NA?

Lines 275–277: Are discrete pCO$_2$ measurements treated any differently that underway p CO$_2$ measurements? I'd image that in many cases discrete pCO$_2$ data points would spatiotemporally match with those in the SOCAT database, but may be discarded based on the criteria noted here.

Lines 381–382: I don't think it's obvious why the advent of Bio-Argo floats would cause the mean number of pH measurements per ROI to decrease rather than increase. I think I can infer: the 10-day sampling cycle of the floats and the fact that only one pH measurement is obtained in the upper 10 meters means one individual ROI is generally created for each float profile? Regardless, since this result seems counterintuitive on first glance, I think it should be explained here.

Figure 16: Should the green line in the legend be T14 fit, rather than TS13?

References: Lauvset et al., 2018 is not in the reference list.

References

Bakker, D.C.E., et al. A multi-decade record of high-quality fCO$_2$ data in version 3 of the Surface Ocean CO$_2$ Atlas (SOCAT), Earth System Science Data, 8, 383–413, 2016.

Orr, J.C., Epitalon, J.-M., Dickson, A.G., and Gattuso, J.-P. Routine uncertainty propagation for the marine carbon dioxide system, Marine Chemistry, 207, 84–107, 2018.

Takahashi, T., et al. Climatological distributions of pH, pCO_2, total CO_2, alkalinity, and $CaCO_3$ saturation in the global surface ocean, and temporal changes at selected locations, Marine Chemistry, 164, 95–125, 2014.