

Earth Syst. Sci. Data Discuss., referee comment RC2
<https://doi.org/10.5194/essd-2021-159-RC2>, 2021
© Author(s) 2021. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on **essd-2021-159**

Anonymous Referee #2

Referee comment on "Description of a global marine particulate organic carbon-13 isotope data set" by Maria-Theresia Verwega et al., Earth Syst. Sci. Data Discuss., <https://doi.org/10.5194/essd-2021-159-RC2>, 2021

General Comments:

The authors present a new compilation of global measurements of POC $\delta^{13}\text{C}$ values and a description and overview of the data set. This is an important goal: such a data set is currently lacking and could be an important tool for observing temporal changes, validating current ocean biogeochemical models that incorporate $\delta^{13}\text{C}$ -POC, and generally exploring ocean carbon dynamics in the particulate phase.

After downloading and considering the data set in addition to the summary manuscript, this effort leaves some questions. The creation of such a data set should be forward-thinking and demonstrate a clear vision for how it will grow. Some improvements are needed for this data set to be truly useful and forward-thinking.

Specific Comments:

- Presentation of data as anomalies from a mean does not seem logical. As more data are added, the mean will change; thus, the data set needs to be presented as actual values. In addition, it is unclear why anomalies are reported to so many decimal places. The original data are likely all reported to only one decimal place, possibly two, which is the maximum practical precision of typical isotope ratio instrumentation. The mean itself and the anomalies should not be presented to a precision exceeding 1-2 decimal places, depending on calculation of uncertainty (see detailed comments below).
- Many currently available sources of data are not included. I refer the authors to Close

and Henderson (2020) for one example of a different list of publications containing oceanic POC $\delta^{13}\text{C}$ data, which they incorporated in a recent depth-resolved global POC $\delta^{13}\text{C}$ assessment, as well as a list of publications which they did not include for reasons specific to their assessment. They included more than 300 data points from Pedrosa-Pamies et al. 2018; Bishop et al. 1977; Jeffrey et al. 1983; Hurley et al. 2019; O'Leary et al. 2001; Trull et al. 2008; Saino 1992; Minagawa et al. 2001; Hernes and Benner 2002; Druffel et al. 2003. Additional data sources they did not include but listed and collectively contain hundreds more data points: Williams and Gordon 1970; Eadie and Jeffrey 1973; Druffel et al. 1996; Benner et al. 1997; Trull and Armand 2001; Hernes and Benner 2006; Close et al. 2014; Krishna et al. 2018; Liu et al. 2018; Griffith et al. 2012; Xiang and Lam 2020. Most of these data sources are not included in the data set presented here, and most of them are not archived in PANGAEA. There are currently many different databases currently used for isotope data in the oceans, particularly across different national research agencies such as BCO-DMO in the U.S., National Institute of Oceanography in India, JAMSTEC's time-series data sets, Japan Oceanographic Data Center, etc. Importantly, the current manuscript mentions a lack of data from the northern Pacific, but it has missed some publications containing such data, perhaps because European/Atlantic research results are disproportionately represented in PANGAEA.

- Seeing as this initial data set is missing much existing data, how do the authors propose to keep the data set up to date as new data are produced and published, or not published but instead entered into other databases? In addition to the publications listed above, there are new data sources since 2020, such as from the Arabian Sea by Silori et al., from the Arctic by Xiang and Lam, and South China Sea by Yang et al.

Some additional major questions:

- What is the reasoning behind including data without 3-dimensional spatial coordinates (latitude, longitude, depth)? 128 data points lack lat/lon data, and 837 data points lack depth data. Of what use are data points without 3-dimensional spatial information?
- There is an opportunity missed here to also include details of analytical methods that would serve as a quality control measure. Namely, did the original data sources describe acidifying the samples (i.e., can we be sure the data are POC rather than total PC?), using what acidification technique, and did they include a blank correction? Older data may have been produced using closed-tube combustion/dual inlet IRMS, whereas newer data were likely produced using EA-IRMS. Because sampling method is included, the lack of analytical method is notable.
- Defining POC: For some of the sampling methods, a size fraction is not specified. For instance, in situ pump and MULVFS samples are often size fractionated, but there is no data field specified here for size fraction. Similarly, there are zooplankton net results in the current data set. Do the contents of a zooplankton net belong in a data set of POC? Many of the other collection methods listed here exclude zooplankton as components of passive POC, such as pre-screening of sediment trap samples through a 250-350 micron mesh to exclude zooplankton.

Technical Corrections:

Data set---

Analytical uncertainties are not reproduced here but should be. They should have been included in the original data sources. Often the uncertainty for bulk measurements is consistent across samples so may be mentioned only once in the source texts rather than being tabulated for each data point.

Underway data is reported as 0 m depth, but ship seawater intakes are usually several meters below the surface. For those interested in the surface microlayer, the distinction between 0 m and 5 m would be an important one.

What is "biological sample"? This does not sound like POC.

How are CTD/rosette, CTD, bottle, and Niskin bottle different methods?

Text---

Introduction: contains unsupported claims: please include references.

Line 10: The reason for the focus on the Atlantic should be clarified (data coverage)

Line 30: awkward phrase. Maybe "some particulate organic carbon" or "parts of the particulate organic carbon pool"

Line 32: please use an earlier reference for the soft tissue pump

Line 40: omit the factor of 1000 to adhere to generally accepted $\delta^{13}\text{C}$ terminology (see Coplen TB. 2011. Guidelines and recommended terms for expression of stable-isotope-

ratio and gas-ratio measurement results. Rapid Commun Mass Spectrom. 25: 2538-2560).

Lines 108 and 110: repetitive about referring to this as the Tuerena dataset.

Line 115: this sentence about depth intervals & timeframes is confusing

Line 116: Longitude values were converted

Line 123-124: The last part of this sentence is awkward

Line 125: Need to elaborate on why some values were considered suspicious.

Line 127: Run on sentence (comma splice)

Line 127-128: Confusing sentence. An example would be helpful. What was done if this criteria was not met?

Lines 135-137: This information might fit better where longitude transformation was first discussed.

Lines 142-143: How did these datasets & interpolations address depth?

Equation 3: A mean should not be this much more precise than the precision of the individual measurements. The analytical uncertainty in individual measurements is likely between 0.05-0.2 per mil, and the original data were likely presented to a precision of 1-2 decimal places (per mil notation). The propagated uncertainty in the mean would likely be somewhere around 0.1-0.3 per mil. Therefore, the precision of the mean likely should not exceed 1-2 decimal places, purely from a standpoint of analytical realities.

Line 153: unnecessary second comma

Line 154: Meaning of first sentence is unclear

Line 172: The depth of the euphotic zone varies by location, so this statement is not wholly accurate.

Lines 202-205: would be useful to mention at what depths these outlier data points were collected. Is there any context to these locations or depths that would hint at the reason for these low values?

Line 206-209: Values higher than -10 per mil at depths between 636-901 m are very strange for a station where the total water depth is around 3000 m (this appears to be Line P, somewhere near Station 11-13). This may be a case where checking the acidification method and/or contacting the authors would be appropriate.

Line 242: extra "second"

Line 263-264: Not a full sentence; could combine with previous.

Line 267: Confusing sentence: discussing high values but mentioning explicitly low values, <30 per mil.

Line 283-284: Difficult to understand this sentence

Line 293: Figure 7

Section 6.1: Somewhat confusing. Grouping data by season may help.

Lines 300-301: Specifying the percentage of data is in this decade would be clearer

Lines 301-302: How many data points is enough? How do you test this?

Line 304: Possible data coverage issue here. The Southern hemisphere could have lower

values due to more data coverage in the high-latitude Southern Ocean compared to the high-latitude Arctic.

Line 305: How is this determined? Doesn't July have high values?

Line 306: Winter in which hemisphere?

Line 309: Lowest what?

Line 316: The phrase about the data points in 1960s and 1980s suggests that there might be similar numbers of data points; however, the 1980s have much more data but the spatial coverage is similarly poor.

Line 333: why is the median for the 2000s not shown on the figure?

Line 375: interest

Tables---

Table 1 caption: unnecessary comma in second sentence.

Table 1 Eadie and Jeffrey entry under column "to": is this added from the source itself?

Table 2 footnote 2: how did you address this for sample durations *not* entirely within an explicit month and year?

Table 3: entry for depth under "coverage": typo in the total number of data points

Table 4: What does this first sentence mean? I would specify: Larger number of surface ocean depth layers than included in the model grid (right?)

Figures---

Many figures have "change" in the title because they are showing change over time, space, etc., -- these figure titles could be more descriptive. For instance, titles for Figures 8b and d would be more appropriate as "median d13CPOC value in the ___ hemisphere across months" or similar.

Figure 5: Better figure resolution or larger size will be helpful for the reader. Colors are difficult to discern at this resolution.

Figure 6: Please name the biomes in the caption so the reader does not have to refer to the original publication. The colors in panel A are difficult to discern, so it is difficult for the reader to relate the colors in A to the colors in C, and then also to transpose that information to the numbers in panel B. Including a north-south indicator on panel B would be helpful, and including the numbers on panel C would be helpful. A more appropriate title for panel B might be "d13CPOC mean across biomes."

Figure 10a and b titles: "data density by year in..." would be more descriptive.