

#### General comments:

I would like to thank the authors for their thorough manuscript. I applaud them for their effort to compile such an extensive dataset of all available d13C POC measurements across the global ocean over multiple depth layers and monthly and decadal time-scales. Datasets as such are highly valuable and find application in a variety of research and technical fields and are useful to calibrate/validate process-based, mechanistic isotope-enabled models. The dataset and its components are clearly presented and so are major patterns in d13C POC values across space and time. It would be also interesting to see if there are any trends with depth, and how these may vary among ocean areas.

I provide a few – hopefully constructive – general and specific comments below and recommend this manuscript for publication after my comments are addressed or discussed.

#### Issues that I noticed throughout the manuscript:

-data is plural, datum is singular. So, please make sure to use plural forms when speaking about data

-be sure to use present/past tenses and active/passive forms consistently

-be consistent with terminology related to spatial distributions: coarse/fine grid/interpolation.

#### Specific comments:

##### Abstract

Line 1: I think that the correct terminology here is just ‘marine particulate organic carbon stable isotope ratios (d13C POC)’, without -13 here, as by definition the isotope ratio is given by the ratio of the heavy (carbon-13) to the light (carbon-12) isotopes.

Line 13: need commas for statement regarding the Southern Ocean: ‘except for the Southern Ocean, which shows a weaker trend, but contains...’.

##### Introduction

Line 16: Consider changing ‘it is regulating’ with ‘it regulates’

Line 47: there are, not their are

Lines 47-48: I don’t particularly agree with this statement as I think that the factors and processes underlying fractionation during photosynthesis are fairly well-understood, and fractionation can be predicted with confidence from [CO<sub>2</sub>aq], phytoplankton growth rate and community composition (due to different cell sizes and geometries of different taxa). Furthermore, these factors are all governed, to an extent, by temperature, which makes temperature a key predictor for fractionation and d13C POC patterns. Perhaps you should mention that here.

Line 58: concentration of CO<sub>2</sub>aq is also temperature dependent, and so is the distribution of phytoplankton communities. That is, all factors that exert a direct influence on photosynthetic fractionation are ultimately controlled by temperature, which is a major control on fractionation during photosynthesis by phytoplankton. I would stress this a little more in the Introduction.

Lines 59-61: again, I think we have a fairly good understanding of the processes causing variation in photosynthetic fractionation, but a dataset with extensive spatio-temporal coverage is certainly needed to investigate how trends change across space and time and the mechanisms underlying these changes as well as for calibrations/validations of process-based/mechanistic models with no data component.

Line 65: please add citation Magozzi et al. 2017 Ecosphere here. This study models the C isotope fractionation during photosynthesis as a function of a suite of variables provided by the ocean biogeochemical model NEMO-MEDUSA and predicts spatial and temporal patterns in d13C POC across the global ocean over seasonal to decadal time-scales.

Line 66: calibration AND validation. A major issue associated with isotope-enabled biogeochemistry models for the global ocean is the lack of reliable validation datasets, with sufficient spatio-temporal coverage to allow proper validation (sometimes, datasets are so scarce or so scarcely comparable that it almost makes more sense to 'trust' the mechanistic model, based on fairly well-known and understood processes, rather than calibrating the model to the available data)

Line 80: what does 'multilateral' mean?

Line 83: don't need a capital W for we after the semi-column. Please fix this here and throughout the text, as well as in figure captions.

## 2 Data acquisition

Line 89: rephrase this as 'the adjustments that we conducted are described in the following sections' or something.

### 2.2 Adjustments made

Line 122: Isn't this Sackett et al. 1966? With a double t?

## 3 Content and structure of the data set

Line 140: don't need a full stop before reference.

Line 141: why are data presented as anomalies with respect to the global mean d13C POC value? I think that the authors explained this in lines 149-150 but it is not clear to me what they mean.

### 3.1 Raw data file

Lines 149-150: what do the authors mean by 'anomalies contain all relevant information...during first steps of model calibration'?

### 3.2 Interpolated data sets

Please make sure that you use the terms coarse/fine and grid/interpolation consistently throughout the manuscript.

Line 165; '...seven different files, where six files contained an individual decade each' or something.

Lines 165-168: here you should maybe say that: on the coarse grid, data were interpolated independent on time, averaged across depths; interpolation on the fine grid only included data with complete spatio-temporal information, averaged across times and depths.

Lines 168-175: why did you interpolated data onto two different grids? Couldn't you just interpolate on the fine grid and resample if you needed values interpolated on a coarser grid?

Line 176: would be helpful if you could please add a statement to say, in poor words, what the Ferret interpolation does and how that works; essentially a sentence that explains the eqns., briefly.

## 4 Main data set characteristics

Heading: you always use data set, not dataset. Make sure to be consistent.

### 4.1 Range and outlier values

Line 199: before and after the two outer modes, respectively? Density declines to 0 at d13C POC values < than the more negative mode, and at values > than the more positive mode. Is that what you mean here, right?

Lines 204-205: it would be helpful to mention where these locations are. Also, please change 'the smallest outlier' with the 'most negative/the lowest outlier' or something, as it is not straightforward what you mean here with smallest outlier

Lines 206-2010: again, I think it would be helpful if you could please add where these coordinates are

#### 4.2 Sampling method

Line 220: What does it mean that different sampling method could be attributed to 67% of the data as meta information? That 67% of the data had associated sampling method information?

Line 229: double brackets in reference

Lines 232-233: how did you account for spatial sampling bias? What do you mean here with 'by comparing with regions'? Simply that you compared d13C POC data obtained with different method within each region (Figs. 3b-d)?

Line 233: are they? It looks to me that in some cases different sampling methods provide different d13C POC values, e.g., between 30-60 N, bottle values are lower than values obtained with the other sampling methods

Line 236-237: please rephrase this sentence as it is very complex and it is not clear to me what you mean

Line 239: closely aligned

In general, how do read from Fig. 3 the % of data collected with each method in each area?

Line 240: rephrase with 'the variance... is approx. 3 per mil lower than the variance of all d13C POC values, which is approx. 5 per mil, the highest value observed here', or something.

Line 242: show a pronounced, remove clearly. Also second is repeated twice. Also show a clear individual maximum, remove mostly.

#### 5 Spatial distribution

Line 245: please consider rephrasing this, e.g., 'we show the spatial distribution of d13C POC measurements across the global ocean surface and depths'. Data is plural, therefore 'most d13C POC data have been measured', please make sure to be consistent with use of plural for data throughout the manuscript.

##### 5.1 Vertical distribution of the data set

Line 250: if 80% of the data have associated depth info, depth is a fairly well-recovered metadatum, isn't it? Does that mean that most datapoints don't have associated T and sampling method info?

Line 254: 'within the first 130 m'

Line 255: remove already

Line 255-256: '200 d13C POC values are available in the depth interval [...]'

Line 257: add respectively at the end of this sentence

In addition to how many data points there are for each depth layer etc, it would be interesting to see a plot showing trends in d13C POC values with depth, similar to Figs. 6, 8 and 10 for biomes, months and years...

##### 5.2 Horizontal distribution of the data set

Out of curiosity: does the dataset include any d13C POC data for the Mediterranean Sea?

Line 260: I tend to prefer the use of 'grid' over 'interpolation', as the interpolation is essentially the spatial/horizontal distribution of values which can be done over a grid, isn't it?

Line 261: to set context for next sentence, mention here that data are averaged across all depths'.

Line 263, don't need a full stop after Figure 5, but comma.

Line 264: not sure what 'also, data locations of ... occur' means.

Line 267: lowest?

Line 269: substitute 'with' with 'of' XX per mil

### 5.3 Meridional trend of d13C POC values

Line 276: again, make sure to be consistent with use of coarse/fine grid/interpolation.

Line 279-280: description of colors and lines should be given in figure caption, not in the main text.

Line 283: what does 'but 14' mean? Biomes were numbered from 9 to 17, where 15-17 had to be cut to the given lateral range. Also, consider using longitudinal rather than lateral here.

Line 284: I think there is a mistake here, location of biomes in the Atlantic is shown in Fig. 6c, not Fig. 10.

Lines 285-287: please insert per mil symbol. Also consider changing 'the final biomes' with 'the biomes with more positive d13C POC values' or something.

## 6 Temporal distribution of the dataset

Line 293: Fig. 7 here, not 5.

Line 295-296: latest data are (plural), and again in the following line

### 6.1 Seasonal trends

You don't describe Figs. 8b,d but I think they're informative as they show the seasonal trend in d13C POC values.

Also, please make sure the distinction between winter/summer is clear for the /S hemisphere in this paragraph.

### 6.2 Multi-decadal trends

Line 328-329: remove 'both'. Also second is repeated twice.

Lines 329: main maximum shift or the shift in main maxima. Remove 'with every decade lower'.

Lines 334-339: I would remove these lines, if you really want to keep Figs. 10c,d in. Or you could also remove the figures and just say that there are not enough data in the SO to investigate multi-decadal trends.

## 7 Conclusions

It would be nice to have a paragraph in Conclusions with examples of research and technical questions that could be tackled/answered with datasets as such. These applications should link back to themes presented in the Introduction.

Additionally – and this may reflect my own research interests – I think that the authors should stress the importance of their dataset for calibration/validation of process-based, mechanistic models. A major issue related with the application of these models in ecology, for instance, has been the lack of suitable calibration/validation datasets, resulting in large and mostly unknown uncertainties (models trusted more than data, as they're based on fairly well-understood mechanisms whereas data are scarce and often incomparable). Datasets like this one provide a validation tool for mechanistic model, and potential for the development of data-based models of the spatio-temporal distributions of stable isotopes in marine ecosystems. An approach that has been successfully used to develop data-based isoscapes is the INLA method (St John Glew et al. 2019 MEE, St John Glew et al. 2020 ESSOAr), which allows separating spatial from non-spatial components of isotope variance when predicting spatial isotope patterns. This dataset could be suitable for such approach as it contains some meta information (e.g., sampling method, depth,

month, decade, etc.) which can be included as factor to estimate non-spatial variance when predicting spatial variation from environmental covariate sets.

## Tables

Table 1 & others: don't need a capital letter after semi-column. The second column lists in which ..., without comma after lists. The third and fourth columns, plural; also unnecessary comma between show and from; also show from what values to which values (or something).

Table4: change inspired with based on, or something. Not sure what the sentence starting with 'below 50 m...' means, why only below 50 m? In the last sentence, depth range not depths range.

## Figures

Figure 1: in caption don't need capital V for values after semi-column. Please fix this here and in other figs' captions and throughout text.

Figure 6b: can you plot mean lat for each biome on the x-axis, rather than biome number? Or at least an arrow N to S below the x-axis? You need to make this fig as much self-explanatory as possible.

Also, b can you plot some confidence intervals around means in panel b, given by variance of KDE? Alternatively, you could plot boxplots of for each KDE, without black line connecting means.

Figure 8b,d: Similar comments to Fig. 6b.

Figure 9: in caption, 'grid-locations of d13C POC data, colored by sampling decades' or something. Find a clear way to say that the grids of sample locations are shown here, colored by the decades in which the samples were collected. Aren't there any grids with multiple samples collected in different decades?

Figure 10b,d: Similar comments to Figs. 6b and 8b,d.

Also, wouldn't show panels c,d for Southern ocean, but just mention in the text that the SO was excluded from analysis as available data are sufficient to derive KDE for only three decades. If you really want to keep panels c,d in for consistency and as justification for insufficient data in the SO, then don't describe patterns in the main text.

Panel b: why does the y-axis go down to -30 per mil, when the minimum mean d13C POC value > -24 per mil?

## Dataset

I have seen that dataset is stored in Pangaea; do you also plan to submit it to Isobank?