

Review of manuscript [essd-2023-479](#) from [Dugenne *et al.*](#) with the title “[First release of the Pelagic Size Structure database: Global datasets of marine size spectra obtained from plankton imaging devices](#)”

2nd Round of the review.

General: I commend the authors for their responses and the changes made to the manuscript. I have no further comments on the revised manuscript.

However, I need to restate a point missed from my first comment.

Major comments:

previous comment: Point 1 (critical): I’ve encountered some potential errors in the size distribution datasets. The five largest biovolume size classes in the Scanner and UVP datasets (6.14×10^{12} to 4.91×10^{13}) have been constructed using a different size step than the rest. I.e. $V_{n+1}/V_n \neq 2$ or $\log_2(V_{n+1}) - \log_2(V_n) \neq 1$ (where V = biovolume size class) for said classes. Is this possibly due to something going wrong when merging size-fractionated data? Additionally, there are also some issues in the lower or upper size ends of some spectra. There are some cases when the largest class/classes should have been dropped (according to what the authors describe in lines 375 – 377) but they didn’t (maybe this is somehow related to the above-mentioned issue?) (see [i]). Also, according to authors, the size bin of the maximum biovolume/abundance was chosen as the lower size limit (lines 379 – 380), but there are cases where this is not true (see [ii]). These are all critical issues that can potentially heavily influence the spectrum parameters and should be addressed before publication. Regrettably, this means that the corresponding graphs and analyses in the results should also be checked and corrected if necessary (but, see also Point 2).

authors’ response: We thank reviewer 2 for checking carefully the products and providing detailed examples of faulty datasets below. We have identified the cause for the discrepancies: the thresholding function described on lines 322-337, and have now corrected it, checking the expected datasets provided in [i]. As a consequence, we have now updated the products on Zenodo and all the product links, statistics and figures provided in the manuscript. Although we had to re-generate the products, we note that none of the updated statistics and trends presented or discussed in the manuscript differed significantly from the previous estimates (mostly the 2nd decimals). The low biovolume/abundance in the smallest size classes observed in [ii] results from aggregating and averaging normalized biovolume/abundances in product 1a. As mentioned in lines 284-292, size spectra parameter calculations were performed on weekly, $0.5^\circ \times 0.5^\circ$ grid cells, and afterwards, both the size spectra themselves and their parameters were averaged in monthly, $1^\circ \times 1^\circ$ grid cells. Thus, it is likely that some size classes were not present in all sub-bin estimates that were averaged in the monthly 1-degree cells, resulting in the pattern observed by the reviewer. We have included a clarification regarding this pattern in lines 334-336: “It is important to clarify that this

thresholding is applied to the weekly, 0.5°x0.5° bins, so it is possible that 1a products present low normalized/abundance values at the lower end if the smallest size class is present in only some sub-bins”.

response to authors: The authors missed one of the three issues I’ve raised in my first point. Perhaps this was due to bad wording of my original comment, because, otherwise, they thoroughly and carefully addressed this point. The issue not addressed is the fact that some of the larger size classes in the UVP and Scanner datasets are not consistent with the rest of the size classes, i.e. $V_{n+1}/V_n \neq 2$. To demonstrate this, let’s look at the biovolume size class ratio (V_{n+1}/V_n) of the UVP and Scanner datasets:

biovolume size class	BV/ratio (V_{n+1}/V_n)
...	...
1.08e+14	2.00
2.16e+14	2.00
4.30e+14	1.99
8.61e+14	2.00
1.73e+15	2.01
3.45e+15	1.99
6.14e+15	1.78
1.03e+16	1.68
1.74e+16	1.68
2.92e+16	1.68

It is immediately apparent that there is something wrong with the larger size classes. In any case, I’ve pinpointed the source of the problem:

In the github repository (<https://github.com/jessluo/PSSdb>): When the NBSSs are computed (*4_compute_NBSS.py*), the *size_binning_func()* is called (included in *funcs_NBS.py*) which then calls the *ecopart_size_bins.tsv* (PSSdb/ancillary) through the *configuration_masterfile.yaml*. The *ecopart_size_bins.tsv* contains the ESD and biovolume size bins and is indeed the source of this error. In both ESD and biovolume, the ratio $n+1/n$ (n =size bin) should be constant for all n . But from line 45 and after, the $n+1/n$ ratio changes (from ~ 1.26 to ~ 1.19 for ESD, and from ~ 2 to ~ 1.68 for biovolume). Correcting this file should fix this issue. The dataset should also be updated accordingly.