

Authors reply to: What an incredible solution! (Comment on essd-2021-171)

Todd OBrien (Referee)

Referee comment on "The Plankton Lifeform Extraction Tool: A digital tool to increase the discoverability and usability of plankton time-series data" by Clare Ostle et al., *Earth Syst. Sci. Data Discuss.*, <https://doi.org/10.5194/essd-2021-171-RC1>, 2021

Dear Authors,

Wow! What an excellent tool/dataset and a super clever solution to the challenge of comparing disparate time series programs and their data! By aggregating species into a cluster of "lifeform" groupings, you solve methodological intercomparison challenges, while also creating a product that is more readily understandable to ecosystem- and policy- level users. You have been very careful and thorough in your design, and I especially appreciate the confidence ratings that you applied to the different lifeform categories. The PLET trait lookup table by itself is super valuable and useful, and being able to apply it directly to time series via the BASSH/PLET tool is even better. Overall, everything about your manuscript was excellent, and it was a pleasure to read. You provided a really well written paper and methodology, and I had only a tiny few questions after reading through it.

Dear Todd O'Brien, thank you for taking the time to assess our paper and provide your extremely helpful and positive review, it is greatly appreciated. Below we address each of your points, including your text in black font, and our response in blue font.

First Question: You mention that each time series data set is preserved via a DOI. Is this doi/data the "original raw unaggregated data" (e.g., species counts by month by year **before being translated into lifeform categories**) or is it the data after it has been aggregated into lifeform categories (and individual species data is likely removed from that doi)? I ask because the reason IGMETS (and ICES WGZE/WGPME) only worked with totals (total copepods, total diatoms) was/is because some time series holders were hesitant to share their full raw species data. If you are sharing only the aggregated data, that would sooth most contributors (by not releasing the full raw data itself) and greatly increase/encourage more participation. That is an excellent solution to this ongoing challenge, and I think you should talk to ICES WGZE and ICES WGPME about getting more data sets into your tool.

This is a very valid point, and one that we have discussed as a group many times. We ended up making it the data provider's choice, and we ask them to select an option to provide raw data or not on the 'Permission Agreement Form', we ask this form to be completed upon submitting data to the lifeform tool. If the data providers have selected not to make the raw data available then the DOI links to their preferred citation which should be used to reference the monthly lifeform output, however if they have made the raw data available then the DOI

landing page will also provide a link to download that raw data. The lead-author is an admirer of your work, and we would very much like to collaborate with the working groups that you suggest, we see the PLET as something that will grow and expand, with the hope of becoming a global tool.

Second Question: *I really like Figures 5, 6, 7, and 8. Is it possible to get the BASSH/PLET tool to automatically generate those? (If it already does, I could not get figure it out.) Or perhaps you can pre-generate them, for the fixed site time series at least? This is very useful summary information about the time series, with or without the interactive tool component.*

This is a great suggestion and one that we hope to implement in a future version of PLET. We are hoping that we can provide a link to this manuscript when it is published, to act as a resource for the use of PLET and the current datasets. We have spent a lot of time discussing the potential issues of automating outputted comparisons, one of the issues being if coverage were not sufficient in some regions (depending on the dataset). If users want to extract lifeforms from PLET in a particular region and compare the outputs from different organisations they can apply their own method of normalisation to do so (e.g. The z-score figures within the manuscript), but we are not then accountable for any misinterpretation/misuse of the data.

Third Question: *The PLET Trait Look-Up Table is probably most important part of the entire database/tool "ecosystem". With what frequency do you hope to maintain and expand that table? On a related note, while you say you are marine-focused, adding Baltic Sea species would greatly expand your area of coverage in Europe. Surely HELCOM has most of the Trait info you need to make this expansion in the Look-up Table?*

Yes, we are very keen to keep maintaining the PLET trait look-up table, and very much hope to incorporate Baltic Sea species with the help of experts in that area, as well as other coastal species that exist in brackish waters. Currently this process is fairly ad-hoc and this is linked to the availability of funding, essentially when we have a funded project that is linked to the use of PLET, then we are able to work on the developments of the tool and the look-up table.

Fourth Question: *While you say (in the manuscript) you can't really compare different time series, you actually did .. in Figures 5,6,7,8, by using the Z-score. If you add these graphics to PLET somewhere (Question 2 above), multi-site comparisons or overviews should also be possible. Maybe not "live" (via the tool), but perhaps as pre-generated products elsewhere in the BASSH/PLET web page?*

What we meant by this statement is that caution should be taken in combining outputs from different datasets. By outputting monthly lifeforms the datasets become comparable to a degree, but there are always nuances or caveats

associated with different sampling techniques that need to be acknowledged. By keeping each data providers data separate but outputted in a unified way for comparison, we respect the differences and acknowledge the data provider, but can also infer changes or trends that may be a function of the sampling technique/regime instead of the actual changes in plankton. We like your suggestion of 'overviews' and this is something we would hope to implement in a future version of PLET. Please also see our response to your second question.

I do not have anything negative to say, but two suggestions:

Suggestion #1: *For me, the BASSH/PLET tool will usually "timeout" on the CPR data unless I subset the geographic region and/or time period. (This is not a problem with the single site time series, as they are much much smaller.) Are you using raw, full- geographic-resolution CPR data (i.e., the original silk locations)? For performance, you may want to subset those into geographic average boxes, perhaps 0.5 x 0.5 or 1.0 x 1.0 degree boxes. It would greatly reduce the number of data points the tool would have to process "on the fly".*

Thank you for your suggestion. We have developed a caching system whereby if a user has already requested an area/period than that data is cached and outputted to speed up processing, but there are still queries that have not been cached and it can hang on the larger data requests. We have not aggregated the CPR data into degree boxes yet simply because of wanting to keep the spatial resolution for some of the finer-scale regional boundaries used in assessments, but this could be a longer-term solution we consider. We have also written some scripts using curl/wget to run through iterations of data queries to pull out lifeform outputs in bulk, and for the CPR data these outputs can be grouped by degree boxes. We hope to publish these scripts on the PLET site once they have been tested and finalized.

Suggestion #2: *Table A1 is super long ... as in 29 pages in the review PDF. Since I am guessing that listing will change fairly regularly, why not make it an online file and only give a one page example of its content in the manuscript? I find the 29 pages distracting as I am trying to get down to Table A2 ... I am really excited to see where this will go! Please reach out to ICES WGZE/WGPME to expand the coverage of this tool!*
Todd O'Brien

We agree with this suggestion and have therefore deleted Table A1 and referred to the Plankton Lifeform Traits Master List instead as the information that was within Table A1 is also included in the look-up table.

Thank you again for your thoughtful and thorough review, we hope you find our responses satisfactory, and we will certainly aim to reach out to the working groups suggested in the near future.
Clare Ostle and co-authors.