Earth Syst. Sci. Data Discuss., referee comment RC2
https://doi.org/10.5194/essd-2021-171-RC2, 2021
© Author(s) 2021. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on essd-2021-171

Aleksandra Lewandowska (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-RC2, 2021

This is an important and much needed tool with a great potential for further development. The manuscript is very well written, all functions of the database are clearly described and justified. I can only hope that this tool finds its way to the stakeholders and the policy makers in Europe.

I especially admire the functional groups **look-up table and the confidence rating**. It is easy to expand and continuously improve when more data are added. The spatial coverage of the database can be rapidly expanded, especially if SMHI extends access to their time-series from the Baltic Sea. If this happens, it would make sense to expand the lifeforms table by filamentous cyanobacteria to track their blooms in the Baltic Sea. Such information would be highly relevant to policy makers in the Baltic Sea region and some other coastal areas in Europe.

I also appreciate that data from different sources are not aggregated. This gives a lot of **freedom for the users,** who can apply and develop their own statistical techniques to make generalisations.

Figures 5-8 are **wonderful examples how to use the PLET** and what kind of information can be extracted. I do not expect that the database developers will offer such visualisation tool, but this content of the manuscript is a great source of inspiration for the users.

Figures 2-3 on the sampling effort are extremely important from the point of view of statistical diagnostics. It would be great if such figures could be included in the metafile description on the website, so that the user can easily see where are the potential gaps in each dataset and what are the limitations.

Although the manuscript and the database are impressive, below are my suggestions for some improvement.

Regarding the manuscript:

- It might be a good idea to highlight the **advantage of PLET over satellite derived information in the introduction**. There is a short sentence in the discussion about the limitation of bulk indices, such as total chlorophyll a concentration, but I think it would be good to have it earlier in text.
- There is no mention of **current development of plankton trait databases**, such as nutrient utilisation traits database (Edwards et al. 2015 - Ecological Archives), Baltic Sea phytoplankton traits database (Klais et al. 2017 - Functional Ecology), French phytoplankton traits database (Treyture et al. 2020 - Scientific Data). It would be good to place the PLET in their context. Maybe it would be worth adding the links to such trait databases in the future, if they exist for individual datasets, e.g. in the metafile description. This would be especially valued in those cases where taxonomic lists are made available.
- Please add a short information how the PLET is **dealing with synonyms and updates in plankton taxonomy**. Is there an automatic check applied (e.g. with WORMS or AlgaeBase) or does it need to be made manually by data providers? In general, is there a systematic data quality check performed upon submission of the time-series? How often such quality check should be performed? I wonder how to ensure the consistent data quality among PLET database, if the data quality check is the responsibility of data providers. I am sure this is not a problem at the moment, but how to guarantee it in the future when the tool expands?
- Please add a **link to the SMHI portal** in the chapter 3.1.8 similarly as you did for the other time series (https://www.smhi.se/en/services/open-data/national-archive-for-oceanographic-data/download-data-1.153150).

Regarding the PLET:

- The **website performance** needs significant optimisation. I believe the problem is not in PLET, but rather in the host server, but this should be fixed before the tool expands. I tired different browsers and different computers, but the problem persists and the service website jams easily, even when I'm trying to limit my search and download data in small pieces. If this causes problems now, it will grow in the future.
- The short **description of sampling methodologies** (chapter 3.1) is excellent and could be added to the metafile together with the information on sampling effort (see my comment to Fig 2-3). This would make the service more user friendly.
- As the tool is meant for biodiversity assessment, it might be good to add some basic **information on changes in taxonomic resolution**. For example, species accumulation curves for each time series could give a clue on significant change in resolution, which can affect interpretation of the outcomes. As many diversity indices, including the most popular species richness, are sensitive to changes in sample size, this is an important information on data quality. Depending on the visualisation, such curves could have annotations with information on changes in methodology or

instrumentation, which correspond to observed inconsistency.