

Earth Syst. Sci. Data Discuss., referee comment RC2  
<https://doi.org/10.5194/essd-2022-41-RC2>, 2022  
© Author(s) 2022. This work is distributed under  
the Creative Commons Attribution 4.0 License.

## Comment on **essd-2022-41**

Emilia Trudnowska (Referee)

---

Referee comment on "The Green Edge cruise: investigating the marginal ice zone processes during late spring and early summer to understand the fate of the Arctic phytoplankton bloom" by Flavienne Bruyant et al., Earth Syst. Sci. Data Discuss., <https://doi.org/10.5194/essd-2022-41-RC2>, 2022

---

This article presents impressive, really and definitely impressive project and consequently its datasets. Moreover it concerns a very important and valuable topic – retreating ice in the Arctic and its impact on the primary production and its fate, thus the consequences for the whole ecosystem. The number of measurements, parameters, devices, new technologies, as well as the resolution of sampling are remarkable. I am sure that plenty important articles may emerge from this dataset. Moreover it can support many other studies as a great reference. In general the dataset as well as the description were made with caution and mostly with sufficient details, only some very minor suggestions I can provide for better utilization by other users:

Title: actually this dataset includes much more information and parameters than just phytoplankton spring bloom. I would suggest something oscillating around the processes in the marginal ice zone. Also 'spring' is not that accurate for the June-July period

Abstract: I think you could add a little bit more information (e.g.; list of most important measurements, parameters performed, the number of stations etc.), while the name and the abbreviation of the ship are of lesser importance in my opinion. I think it is also important to mention that not only data from cruise-based stations are available, but also high resolution measurements of gliders, MVP, ship-track water monitoring, BGC-Argo floats etc..

L86: You can consider referring here to an article about the export of the marine snow produced by the phytoplankton bloom that was based on the GreenEdge dataset : Trudnowska E., Lacour L., Ardyna M., Rogge A., Irisson J-O., Waite A., Babin M., Stemmann L. (2021) Marine snow morphology illuminates the evolution of phytoplankton blooms and determines their subsequent vertical export. Nature Communications, 12, 2816. In general it would be great if you could gather all the existing articles that

published results related to GreenEdge. However, I am aware that it is not a review, so up to you and editor to decide, if that remark is relevant.

L119: for understanding how such complicated dataset can be used to track successive crossing of the MIZ you just refer to Randelhoff et al., 2019, but I think one sentence explaining the idea of arranging stations according to OWD (open water days) would be very helpful for potential data users. Now I see that you did so in 4.1 section, good!

Table 1 – I think it is incomplete (e.g. no mention of the UVP, LISST, sediment traps etc.). It would be great if you could list all the equipment used. Moreover, 'Hydrobios' is just a company producing many equipment, please be specific what you mean by this. Potentially you could also consider adding an extra row about the high resolution measurements so as to include all the devices applied in one place. While I think that Table 3 is great (impressive) with the full list of parameters, Table 1 disappoints with scattered information. In my opinion it should not be that much devoted to the 'operations' but rather to equipment used, because only the cruise participants may know what stays behind a specific 'operation', while the second-hand users prefer to know what was done with which device.

L133-135 I think those sentences about Argo are not suitable for this article, especially to this part.

L152: This header is not strictly true for the content. You say nothing about data processing, but a lot about the unification of formats. Therefore I suggest to change it accordingly, e.g. Time-for-space formatting & data quality control

L172: In my opinion you mostly provide methodology than data description. Maybe you should put a header as e.g., Description of data collection.

L176 & others: Maybe you could consider giving numbers to the parameters included in Table 3, and then to refer to those numbers also when listing them ?

L389: what are those further analyses?

In general Figures are of good quality, just a few suggestions:

Fig. 6. What is the black line representing? Is that a ship route ?

Fig. 10. Please correct a figure caption to be more specific/informative, as "Relative proportions of the different taxa within the different groups of samples," by this we do not now that those are mostly bacteria presented etc.

Fig. 15. It can supplement Fig. 2 as e.g., b) panel

Datasets: I was not able to check all of the files provided, but within the ones I investigated at <https://www.seanoe.org/data/00487/59892/> I found some problems:

- Not only data from the Cruise are provided (as indicated in an article), but also data from the ice camps (2015, 2016), e.g. CDOM absorption dataset, Bacterial biomarkers dataset, Camlum (Ed) dataset, Carbon stock (DOC/TDN) any many many other files
- Erogenous latitude (e.g. 200 Lat in CDOM absorption file)
- NAs – rows with NA content in such crucial cells as 'Station', 'Depth' should be deleted as are useless (e.g., Particulate and phytoplankton absorption dataset, Bacterial biomarkers dataset)
- Different data formatting within one file – at some point the number of columns and their content starts not to agree with the other rows - e.g., Bacterial biomarkers dataset; Days of open water (DOW); Pigments, nutrients, Chlorophyll a and Phaeopigments (concentration); Sediment traps (chl<sub>a</sub>, faecal pellets flux, Particulate mass, etc.)
- Headers – it is very important to name them in an identical manner, e.g. in Table 4 you claim that each .csv includes the same metadata columns. Even if that is true, but is not always, then the naming should be exactly the same, e.g. you indicate in a Table "operation code", while it occurs by different variations of names (e.g., "code\_operation" in Bacterial biomarkers dataset, Nutrients etc.), which is the same for human, but makes a huge difference for coding and automatic merging various datasets.
- Zooplankton – while there are no taxonomical results provided based on net sampling, the UVP dataset (Underwater Vision Profiler (UVP) – zooplankton) includes also other than zooplankton categories (e.g., detritus, artefact)

I think you could supplement Table 3 with an additional column saying in which file this parameter can be found.

As far as I understand the files provided at <https://www.seanoe.org/data/00487/59892/>

are 'ready to use', but many parameters described in the paper are not provided there

Whereas the raw data is provided at [http://www.obs-ivlfr.fr/proof/php/GREENEDGE/x\\_datalist\\_1.php?xxop=greenedge&xxcamp=amundsen](http://www.obs-ivlfr.fr/proof/php/GREENEDGE/x_datalist_1.php?xxop=greenedge&xxcamp=amundsen)

But the access to this data is mostly restricted to the owners of the specific account?

My replies to the crucial questions of the review:

Is the article itself appropriate to support the publication of a data set? – Yes

Is the data set significant – unique, useful, and complete? – Unique, Useful but with some mess and problems with access to some types of data

Is the data set itself of high quality? – Yes

Is the data set publication, as submitted, of high quality? – Yes