

Interactive comment on “The Malina oceanographic expedition: How do changes in ice cover, permafrost and UV radiation impact biodiversity and biogeochemical fluxes in the Arctic Ocean?” by Philippe Massicotte et al.

Anonymous Referee #2

Received and published: 24 January 2021

General comment This paper provides a good description of an excellent program focused on an environmental system and its likely response to effects of climate change. The large sampling area of multiple along-shelf transects, and sections running from very shallow waters at the river mouth out to the deeper basin provide great coverage. The multi-disciplinary research makes these data particularly important in answering the direct study questions and for use by the wider science community. This is a unique and valuable data set. The publication is well structured and clear. The figures, in particular the map of station location, the Mackenzie River runoff, and water masses were

C1

clear and helpful for understanding the data set and the rest of the paper. The data and code are accessible through the paper's given websites (LEFE-CYBER for raw and meta data, SEANOE for final data, Ecotaxa for UVP5 data, and Zenodo for code). My two main comments are that the size of the main data set and lack of clear variable definitions make it very difficult to use, and the other follows up on Referee1's comment regarding incomplete information on methods and errors. It is a herculean task to bring together over 150 variables from over 50 PIs into a cohesive data set. With more information regarding the methods and errors and reworking the data set this will be a great addition for arctic and climate change scientists.

Specific Comments Methods and errors: The collection and analysis methods are well described for many of the variables, are current state of the art and have had quality control applied. However, as mentioned by Referee 1, the methods and information on errors are not given for all variables. I found it quite helpful to look at the LEFE-CYBER website (http://www.obs-uvfr.fr/proof/php/malina/x_datalist_1.php?xxop=malina&xxcamp=malina) mentioned by the author as home for metadata and raw data, but see that they do not have information documents for all the variables. Looking through the data folders there are some documents that exist but have not been added as links to the site's variable list (ex. CTD documentation in basic files/doc and basic files/report), but in other cases appear missing, for example “Chlorophyll-a and Phaeopigments (concentration)”. It sounds like this will be addressed as mentioned in author's answers A1 and A2 to Referee 1 and information will be added to table 1 as to where to find methods and error information.

Data availability: It would help to have Table 1., the heart of the paper with key information, expanded to state if the listed variables are present in the provided data set. For example, oxygen Winkler data are not present. I assume the list on the LEFE-CYBER site is current and shows the other missing data. Having a table of all the variables collected is very useful and a user can follow up by contacting the PI, however it would

C2

be good to know what is actually provided in the paper's data set.

Dataset: The large amount and various types of data pose a challenge in assembling a data collection. I did not find the current form of the final data posted to the SEANOE website useable due to its unwieldy size, structure and unclear variable names. A. The full data set is given in a single excel file approximately 2400 columns by 600,000 rows. The data set is too large to handle for software and users. My computer could open the file in excel but did not have enough memory to sort the rows. Its not practical for a user to sort through 2400 columns to find the variables of interest. A lookup table would help. This file could be separated into smaller files, in particular those data sets with a substantial number of columns (ANAP, AP and APHY with 500 columns each), and those that are not directly related (i.e. CTD, Foredeck meteorological data, Benthic Boxcore, Zooplankton sampling). Remove other cruise data. There are data from programs in Aug 2008, Feb 2009, and the TARA programs from Jan 2010 to Aug 2012. Remove columns associated with these other programs (i.e. CTD data starting with column CNZ are for the TARA programs)

B. Variable names are not clear enough to determine what they are, which instrument, method or PI they are from, and what their unit of measurement is (i.e. "Average", "345_1_AC"). Some variable names are shared between some data sets ("Cast", "Station", "Temperature", "Bottle"), some are not ("Temperature v. "Temp_Celsius", "Trans" v. "Xmiss", "Bottle" v. "N_btl_fired" v. "Niskin", "Station v. Q_Name").

In addition to clarifying what the variables and units are, it is important to be able to understand which variables connect to the information provided in the paper and Table 1. This currently is not practically possible. This could be addressed in a lookup table or with the variable name and header.

C. The many analyses from a given Niskin bottle do not use the same identifiers making it difficult to match all the data associated with a given water sample. The appended data set use a mix between Cast, Station, Bottle (and its variants), and Depth. At

C3

best these data would be pre-joined, but at least it would be good to have a consistent identifier.

D. The date variable (column T) is a mixture of formats, only some currently understood by excel and thus not currently sortable. Corrections needed w/in data set: The CTD bottle data is one of the key parts to this data set that all water samples are linked back to. This data needs variable name "Depth_dbar" changed to "Pressure_dbar". The CTD+IOP data from the barge have been merged without any identifier to the date, cast, station or location. From the info sheet on the LEFE-CYBER page it looks like two steps are needed – the filename includes the IOP cast # and the infosheet has a table converting IOP cast # to station.

Specific comments on article's content Sea-ice cover The paper mentions the sea ice in Line 49 and 50 (shelf was not ice-free until mid-August) and Line 99 (shelf was ice-free). These appear inconsistent and it would be good to harmonize these lines. Perhaps an image of sea-ice concentration could be added to the paper within the environmental conditions section or a sea-ice edge added to the map of stations. Station order Add a further comment regarding the effect of sampling out of order. Did wind events occur between station sampling that may affect results? Even a statement saying this was not an issue or may be an issue would be helpful for a data user.

Technical Comments/Corrections Figure 8. Add "Surface samples" to the caption for clarity; It would be easier to compare if both sections were on the same plot. There are a number of "?" in the text that appear to be waiting for more information Figure 11a caption "?", Line 270 "?", Line 279 "?", Line 282 "?" Here suggest using the pre-defined (Line 252) SCM instead of DCM for consistency. Figure 14A Would be nice to see A and B use same latitude range Line 380 Looks like this is supposed to be Station 690, not 680 Line 489: Title is missing text. "A 50 % increase in the mass of terrestrial. . ."

Interactive comment on Earth Syst. Sci. Data Discuss., <https://doi.org/10.5194/essd-2020-252>, 2020.

C4