**Review of the manuscript: "A dataset for investigating socio-ecological changed in Arctic Fjords."**

This work manuscript presents the compilation of various data from Arctic Fjords, to investigate the socio-ecological changes in 7 Arctic Fjords. The manuscript is very clear and well written, the dataset is well presented and useful for several type of studies. However, this dataset is just a compilation of datasets already available in several databases, although I think that much more data are available but maybe not publicly. I suggest publishing it after major revisions. Please find my comments beneath.

**Main comment:**

My main comment is about the added value of this dataset. I understand that this is a compilation of already publicly available datasets. Although of interest for some studies I think such a compilation of data should aim for more than 7 fjords in the arctic, and rather try to compile all available data for Greenland and Svalbard fjords for example. A lot of data is available in a lot of fjords in Greenland, but maybe not easily available indeed. However, considering those data will add a lot of value to this dataset, that could be used to assess for example the impact of climate change on the Arctic Fjords.

**Minor comments:**

How do the authors define the Arctic? Is it the Arctic circle? I am a bit surprised that a fjord in Northern Norway is included in the manuscript. Please clarify.

L30: 'range of habitats for many important species': do you have any reference for that?

L34: Please delete 'extreme'

L37: This sentence argues that most of the sampling in the Arctic is performed by large cruise ships. I think it should be clarified here that you are talking about the coastal Arctic and I guess about surface data. In my knowledge, the central Arctic Ocean is mainly occupied by research ship so far, and most of the data come from autonomous platforms.

l.134: The manuscript explained the source of the data and the type of data, but for the ocean temperature for example, how are the data organized? Are they daily averaged or is it just a compilation of all available data in the fjord? For example, in Kongfjorden, several temperature datasets are available, moorings but also CTD from ships for example… Is that indicated in the dataset?

L.183: Did the author look at the arctic data center for more datasets? Is it on purpose that the fjords are only located on the Eurasian side of the Arctic. I could expect that quite a few data are also available in the Canadian Archipelago.

l.210: 'The first of these is the UNIS database': please add a reference

l.210: 'which is a collection of all the moorings': Is this database only composed of moorings? To my knowledge this database is also composed of repeated ship CTD transects.

l.250: in the datasets, instead of correcting the already available data, I think it will be better to flag those data.

l.299: How is a 4km resolution ice cover representative of the sea ice cover in a fjord? Is this resolution good enough for the size of a fjord?

l.301: in addition to the sea ice cover, it will be of great interest to get the sea ice thickness. Combined with the sea ice cover, this will give an estimate of the sea ice volume, the real indicator of the sea ice loss in the Arctic. Does this dataset gather this information too?

l.335: why is the unit of the nutrients umol/L? umol/kg will be consistent with the units of TA and DIC.

l.337: I am surprised that there are no data for the biology drivers in Storfjorden. There has been several cruises in the region that collected biology drivers. How do you decide which data are merged in your dataset?

Table 3: what is Q?

Figure 5: I am not sure about the interest of this figure. There is indeed correlation in between those variables, but this does not imply causality, and there is no explanation of the correlation between the variables in the manuscript.

l.419: I don't really see the point of section5, as it just shows some correlations but does not explain them really.