Earth Syst. Sci. Data Discuss.,
https://doi.org/10.5194/essd-2017-92-AC2, 2018

Interactive comment on "UDASH – Unified Database for Arctic and Subarctic Hydrography" *by* Axel Behrendt et al.

Axel Behrendt et al.

axel.behrendt@awi.de

Received and published: 12 March 2018

1) Referee comments

1. General comments
This paper presents new hydrographic dataset (temperature and salinity) in the Arctic Ocean and Subarctic sea for the period of 1980 - 2015. The dataset is prepared by combining publically available hydrographic data from various data sources. Proper quality control check and data flagging have been done to secure the quality of dataset. The dataset is available on online, easily downloadable and well documented.

Any field campaigns in the Arctic Ocean require significant logistic preparations and often diplomatic negotiations. After years of preparation, a bad weather, unfavorable sea ice condition, change in political situation can prevent implementation of the scheduled field campaigns. This may be one of reasons why there is no widely recognized publically available hydrographic dataset in the Arctic Ocean. As nicely illustrated by figures 1-3, the quality of WOD 2013 is simply not good enough to conduct any scientific studies. In reality, individual researchers or even organizations own their quality controlled dataset, which is not publically available. One of consequences of this is that more than 10 years old climatological temperature and salinity fields based on WOA 1998, EWG Arctic Ocean Atlas and other sources, as known as PHC climatology (Steele et al., 2001), is still widely used to validate climate model outputs (e.g. IlÄ'scak et al., 2016). The latest version of PHC 3.0 was updated 12 years ago on March 2005. Given the rapid and drastic ocean condition change in the Arctic Ocean from late 1990s (e.g. Proshutinsky et al., 2009; Polyakov et al., 2017), the PHC 3.0 dataset does not represent the ocean condition in present days. Therefore, this UDASH dataset is a milestone to foster various scientific studies in the Arctic Ocean and Subarctic sea.

Having said that, there are three points needs to be improved before it could be published. Once these issues are properly addressed, I would certainly recommend the editor to publish this manuscript in the Earth System Science data journal. First, the dataset is somehow limited for wide exploitation in the climate science community because of the way it is prepared. Second, on-going and future plan needs to be refined. Third, figures and tables need to be improved.

2. Major concerns
2.1. Limitation of dataset exploitation
The dataset is somehow limited for wide exploitation in the climate science community

because of the way it is prepared. There are two issues here. First, it is unfortunate that authors choose to prepare the dataset north of 65N. The choice of 65N somehow limits the potential exploitation of the dataset in future. The physical boundary for the Arctic Mediteraneanin is not 65N, but Greenland-Scotland Ridge (Aagaard et al., 1985; Mauritzen 1996; Dickson et al., 2008). A recent GRACE satellite and data assimilation based study shows that uniform sea level variability both in the Arctic Ocean and the Nordic Seas together on the time scale from weekly to inter-annually (Fukumori et al.,2015). As it is, the dataset chops off the southern part of the Nordic Seas. This will hamper future scientific attempts to address physical interactions between the Arctic Ocean and the Nordic Seas. Thus, I strongly suggest to widen the boundary down to 60°N. Since quality control procedure and nessesaraly scripts are already prepared, I presume this can be done relatively easily.

The other issue is that the UDASH dataset does not provide gridded temperature and salinity fields. Since the gridding procedure requires decent time and effort, in reality, this dataset would not be widely used by numerical modelers. As a result, it is likely that the more than 10 years old PHC 3.0 dataset would still remain as a standard observation based dataset to validate the climate models. I understand that preparation of gridded dataset is out of scope of this paper. However, authors could still try to fill the gap. There are several efforts going on to prepare in situ observation based world gridded temperature and salinity datasets. For instance, EN4 by met office in UK (Good et al., 2013: https://www.metoffice.gov.uk/hadobs/en4/), MIMOC by NOAA in USA (Schmidtko et al., 2013: https://www.pmel.noaa.gov/mimoc/). Other possibilities could be found here (http://www.argo.ucsd.edu/Gridded_fields.html). Authors could contact them directly. There is a chance that they would agree to incorporate the UDASH dataset for their next generation of dataset. In this case, authors could mention this in section 4, on-going and future plan.

2.2. On-going and future plan

On-going and future plan needs to be refined. As it is, it is a wish list rather than a feasible plan. Authors need to be more realistic to define what can be done and what can not be done by identifying amount of work involved for each task. Do authors have funding to update the dataset? If yes, how many months / years it would be? Is updating every year realistic in practice? I believe updating the UDASH database every 3-5 years with same quality control scheme to extend the time series towards present days would be very beneficial to the climate scientific community. Going backwards before 1980 may require additional quality control check.

Again, it is important to create the gridded dataset either directly (i.e. by authors) or indirectly (i.e. by someone else) in order to replace the PHC 3.0 dataset to validate the climate models. Authors should explore any opportunities to make this happen. It is also important that authors promote the UDASH dataset to data assimilation community so that UDASH data will be assimilated to next generation of the data assimilation products. For example, ECCO (http://www.ecco-group.org), ORAS4 by ECMWF (https://www.ecmwf.int/en/research/climate-reanalysis/ocean-reanalysis), and name more (https://climatedataguide.ucar.edu/data-type/oceanic-reanalysis).

Concrete and feasible future plan which guides a future direction of the next generation of the UDASH dataset and exploitation plans in the climate science community would be worth to see.

2.3. Figures and tables

Figures and tables need to be improved. Generally speaking, figures and tables describing about the quality control check (figure 1-7 and table 1) are well prepared. However, figures and tables describing about the UDASH dataset itself (figure 8-16 and table 2-5) need to be presented in more systematic way. What readers want to see is when, where, what kind of, and how much data are available in the UDASH dataset. I think figure 11 and figure 12 is a good way to provide such information. In practice, I would suggest following improvements.

(a). Prepare another figure between figure 10 and figure 11about number of profiles available in different months as same format of figure 11 and figure 12. Six distribution maps can be prepared by splitting months into six categories: Jan.-Feb., Mar.-Apr, : : : Nov.-Dec.. Below the six distribution maps, histogram of number of profiles as a function of year (same as figure 11 and figure 12) with the six different categories in different colors can be shown.
(b). Remove figure 13 and figure 14.
(c). For figure 15, prepare it as a same format of figure 11 and figure 12. Below the six distribution maps with different depth ranges, histogram of number of profiles as a function of year with the six different profile depth range categories in different colors can be shown. Double check the fraction of the full dataset. When they are summed up, it should become 100%.
(d). For figure 16, add additional four distribution panels. Temperature and salinity distribution in the Halocline layer (e.g. 50$\pm$20 m) and in intermediate layer (e.g. 1,250$\pm$250 m) would be useful to illustrate the UDASH dataset.
(e). Remove table 4 and 5. Put the fraction of full archive information in figure 11 and figure 12 as done in figure 15.
(f). For figure 8, prepare T-S plots for both Amerasian Basin and Eurasian basin. Plot both before QC and after QC in different colors. The depth range does not have to start from -500m. Choose more appropriate depth range to zoom in the profiles.
(g). Figure 9c. Why this looks different from figure 8a (bottom right, after QC in Amerasian Basin)? They should be identical.

3. Minor concerns

Captions for figures and tables are too short, and often not well explained. This needs to be improved. For example, Figure 3, 4, 5: Specify cruise number, ID, year, month so that the figures can be reproduced. Table 1, 2, 3: captions are too short. Table 1: Most of data source acronym should be in capital letter.

2) Author's response

$4mm]$

Dear Dr. Tsubouchi,

thank you very much for the very helpful comments! Please find our comments to your suggestions below:

2.1 Limitation of dataset exploitation: We are aware that a dataset beginning north of 65°N has its limitations, particularly for the modeller community. However, the region commonly referred to as "the Arctic" starts in the Fram Strait, the Canadian Arctic Archipelago and the Barents Sea on the Atlantic side. This region has been the subject of much scientific study, in particular, in recent years, and has several inherent environmental conditions and processes different to the rest of the world. It is, off course, equally valid to consider a region commonly referred to as the "Arctic Mediterranean", which often includes the Nordic Seas and the Arctic. However, our dataset focuses on the Arctic and connecting passages, and currently does not include the complete Nordic Seas. Including further observations, e.g. from the southern Nordic Seas or the North Atlantic, is out of scope of this work.
In most cases, modellers focusing on large-scale climate issues apply a gridded climatology for their model validation. On the other hand, modelling studies focusing on specific hydrographic situation or regional scale physics (e.g., boundary current, shelf-basin interaction) require non-gridded in-situ profiles. In this sense, our dataset provides a quality checked archive for the latter purpose, while at the same time, provides a basis for a new gridded climatology. Since a requirement for the spatial coverage of a gridded climatology depends on focus of a study, there is no optimal solution except providing a climatology covering the globe. From this point of view,

we would like to spend our effort not on extending the data coverage area, but for merging our dataset as a part of gridded climatology covering the world ocean. As will be mentioned below, this plan is already initiated and ongoing.
In our opinion, it would not be useful to delay further the publication of this quality-controlled dataset, UDASH, including a thorough description. Future plans, however, consider extending the database to include the whole of the Nordic Seas. The thankful suggestion by Dr. Tsubouchi will be archived in the framework of the gridded climatology.

$4mm]Producinggriddedfieldsandprovidinganupdatedclimatology for the Arctic region is defin$

Promotion of UDASH: The dataset was included in the new EU funded project INTAROS (http://www.intaros.eu/) and presented on the second annual INTAROS workshop in Helsinki (January 2018). Furthermore, it will be presented at the Polar Conference of the German Society of Polar Research (DGP) in Rostock (March 2018). The next UDASH presentation will be at the EGU general assembly in Vienna (April 2018). Another opportunity, with a special focus on modelling, would be the FAMOS annual meeting in Bergen (October 2018).

2.2 Ongoing and future plan: We agree that this part has more the character of a wish list and rewrote the section 4.6 to substantiate the plans.

2.3 Figures and tables:
(a) We changed the figure according to the suggestions.

(b) Figures were removed.

(c) We changed the figure according to the suggestions. We added a sentence to the figure caption to explain why the percentages sum up to more than 100% (a profile may belong to more than one category).

(d) We changed the figure according to the suggestions.

(e) Tables were removed and figures were changed accordingly.

(f) We do not understand exactly the difference to what we did. We changed the depth ranges but kept the colors. The words "before QC" and "after QC" were moved.

(g) The reason for the difference is that figure 8 shows data from all the sources and figure 9 data from only WOD. The different shape results mainly from ITP data in the Beaufort Sea, which were taken mainly from the source WHOI.

3 Minor concerns: We added some more information to figures 3, 4 and 5. Captions of tables 1, 2 and 3 were improved. The data source acronyms in table 1 are our own abbreviations and were used in the same way for the data files. We therefore didn't change them.

Best regards,
The authors

2 March 2018

3) Manuscript changes

$4mm]$

2.2 Ongoing and future plan:
Section: 4.6 Ongoing and future work
Updates of the archive are planned on a yearly basis and are funded for at least the next three years. An update of approximately 4 500 profiles from R/V Polarstern was already included in the published archive. Once the data were reformatted, the QC procedures for this update were finished within one week. Yearly updates are therefore a realistic goal. The publication of the next updated and improved version of UDASH is planned for late 2018.

Ongoing work:
We are currently writing an extensive data documentation, which will be available on the PANGAEA webpage in summer 2018. The document contains a large number of additional maps, diagrams, statistics and helpful details which facilitate the work with the database. Estimated amount of work left: 1–2 weeks.
Furthermore, we developed a quick-access tool that enables the user to extract a subset of the data by making selections, such as geographic region, instrument and time period. A preliminary version of the software already exists. It will be equipped with a graphical user interface (GUI) and provided with the first update in late 2018. Estimated amount of work for GUI programming and set up: 4 weeks.

Further goals for 2018:
Almost 40 000 profiles (mainly from WOD) could not be associated with a certain platform type (Table 5). We will attempt to identify the platforms from the metadata record and provide the information with the next update. Most of these profiles

likely originate from ARGO floats and are still recognizable by a number in column 2 (cruise/platform name) of the data files.We are already in contact with experts to identify the ARGO floats by their ID numbers. Most of the ITP profiles within UDASH still have the same cruise/platform name (e.g. itpmerged) in column 2. In one of the next steps, the single ITP profiles will be identified and renamed by using the WOD metadata. The changes will be included in the forthcoming update.
The software for exploring the WOD metadata has already been set up. Estimated amount of work therefore: 2–3 weeks.

Longer term goals and perspectives:
UDASH includes ITP data of different quality levels (level-2 and level-3 (Krishfield et al., 2008)). This information is not yet included in the metadata. We aim at including as many level-3 data as possible and will provide the information in one of the next updates.
The list of DOIs is not yet complete. Finding DOIs for individual profiles is a complex task and requires time for research. However, we will attempt to complete the list and to provide more DOIs in the next versions of UDASH.
Cooperation with WOD: UDASH contains approximately 15 000 profiles which are not yet part of WOD. However, most of these data will likely enter WOD in the coming years. The WOD is now the designated Center for Marine Meteorology and Oceanographic Climate data (CMOC) in the Marine Climate Data System (MCDS) of the Joint WMO-IOC Technical Commission for Oceanography and Marine Meteorology (JCOMM). This will ensure that all available profile data will flow into WOD and that users will find the most complete record in one place (Tim Boyer (NOAA), personal communication, 2018). UDASH would benefit from the completeness of WOD and could more focus more on additional quality control and data presentation.

3 Minor concerns:

Caption figure 3: Sample output of the cruise track test. a) Cruise track in the southern Barents Sea. Source: wod13. Cruise ID: SU-8114 (bottle data). Position 28 (red circle) represents a cruise track outlier. Outlier position date: 7 January 1989. b) Cruise track in the southern Beaufort Sea. Source: wod13. Cruise/Ship: ARCTIC IVIC. Cruise ID: CA-12540 (bottle data). Position 13 (red circle) has a wrong time stamp. Error position date: 12 September 1986.

Caption figure 4: Sample output of the cruise track test. a) Cruise track in the Barents Sea (August 1988) with possible position errors and errors in the time stamps. Source: wod13. Cruise ID: SU-89247 (MBT data). b) Cruise track in the Barents Sea (July 1989). Source: wod13. Cruise: MI0847. Cruise ID: RU-118 (bottle data).

Caption figure 5: CTD temperature profile measured on 13 July 1998 near the Lofoten Islands. Source: ices. Ship: Johan Hjort. Station: 0434. Two outliers (red dots) were detected at around 400 m depth (artificially induced) and 800 m depth (real). See section 3 for a description of the detection algorithm.

Caption Table 1: Data sources and their abbreviations used in the data files. Also shown are time periods and URLs of the downloaded data.

Caption Table 2: Format of the variables and metadata in different columns of the data files.

Caption Table 3: Quality flags used to mark erroneous or suspicious data detected during the QC routines.