

# Interactive comment on "The AlborEX dataset: sampling of submesoscale features in the Alboran Sea"

Anonymous Referee #2

## General comment

Please find below my review of the manuscript entitled "The AlborEX dataset: sampling of submesoscale features in the Alboran Sea" by Troupin et al. I think the data and the paper are relatively well presented. I especially enjoyed that all the files are netCDF format. While the data are limited to a very local application (a 6-day experiment from one sub-region of the Mediterranean Sea), the data are in high-quality and may be useful for process-related studies. Overall, the manuscript may be suitable for publication after moderate reviews. This decision is detailed below.

## Major comments

My major concerns on the actual version of the paper are the following: 1. I think the text is not well organized. Some info on the data is find in Section 2 (AlborEX mission) and in Section 3.3 (Data Processing). This spreading of information makes the search for information through the paper difficult. I would bring Section 3.3. earlier in the paper and avoid to spread the information for each platform in different sections. Some specific comments below are related to this problem (e.g. mention of flags even before introducing them).

The Section 3.3. has been moved earlier in the text, in the Section 2, so that the reader is aware of the processing and Quality Control done of the data. The information is now provided in two subsections:

- "2.4 Processing levels", which has been extended and made clearer following other comments
- "2.5 Quality control", where the general procedure is made explicit.

2. The QC control is a weakness in this manuscript as it suggests that some QC is done, but it is not very clear on which data and how it is done. For some instruments, QC flags and their meaning are embedded in the files (e.g. float and drifters), but some doesn't (glider files). This inconsistency is not so much a problem to me as long as it is clearly stated in the paper which files contains QC flags. These quality flags should however be defined in the text. There are several mentions of "quality flags" in the text and figure caption, but little explanation is provided on these. Figure 12 has 9 quality flags that are not even described (although I see their meaning in drifters and float files). Where the QC is easy to reference

(e.g. "file generated with Socib glider toolbox vX.X", or "File QC done using Socib standard procedure following a procedure described in a certain paper", etc.), it should be mention in the netCDF file as well.

To address these comments:

1. A new table stating the meaning of the quality flag has been created (Table 2).
2. A subsection "QC tests" has been inserted at the end of Section 2 to explain the general procedure for the quality control.
3. In Section 3, for each platform type, a description of the specificities of the QC has been appended.

Concerning the glider data: the toolbox referenced in the manuscript does not apply quality checks on the data in its current version. QC have been implemented but are still in testing phase. Once they are validated, the files will be reprocessed and made available.

More generally, a lot of efforts have been made to ensure that the provided data are of the highest quality, even if that was not reflected in the submitted manuscript. All the SOCIB quality checks are explicitly described in the following document:

QUID\_DCF\_SOCIB-QC-procedures.pdf

SOCIB Quality Control Procedures  
Data Center Facility  
September 2018  
DOI: doi:10.25704/q4zs-tspv

and more tests are progressively developed in the current battery.

3. Why all processing level are not provided? The text suggests that all levels are provided (e.g. Table 3), but at the moment mostly L1 is provided. For gliders, L1 and L2 are provided. For the Float, L1 is provided for Arvor-A3 and Provor-Bio, but L0 for Arvor-C. Why? No explanation for this is provided (I think float data should be provided in L1 and L2 level as well). If some QC is applied on L1, maybe L0 should be provided as well to the future user? For glider L2 data, a choice is made regarding the vertical binning of the profiles. Which size these vertical bins are? This information should be provided somewhere.

Following the definitions adopted at the SOCIB data center, Level 2 only exists for glider measurements: it means that we go from 3-dimensional trajectories to a time series of profiles (the observations are spatially interpolated. The description of the processing levels has been edited and clarified in the new manuscript.

Missing L1 for Arvor-C: this comes from an oversight: the file has been made available in the new version of the dataset. The link to the thredds catalog is: [http://thredds.socib.es/thredds/catalog/drifter/profiler\\_drifter/profiler\\_drifter\\_arvorc001-ime\\_arvorc001/L1/2014/catalog.html?dataset=drifter/profiler\\_drifter/profiler\\_drifter\\_arvorc001-ime\\_arvorc001/L1/2014/dep0001\\_profiler-drifter-arvorc001\\_ime-arvorc001\\_L1\\_2014-05-25.nc](http://thredds.socib.es/thredds/catalog/drifter/profiler_drifter/profiler_drifter_arvorc001-ime_arvorc001/L1/2014/catalog.html?dataset=drifter/profiler_drifter/profiler_drifter_arvorc001-ime_arvorc001/L1/2014/dep0001_profiler-drifter-arvorc001_ime-arvorc001_L1_2014-05-25.nc)

For the glider data gridding (from L1 to L2): the referee is correct, this has to be explained in the manuscript.

The gridding is performed by the function `gridGliderData` ([https://github.com/socib/glider\\_toolbox/blob/master/m/processing\\_tools/gridGliderData.m](https://github.com/socib/glider_toolbox/blob/master/m/processing_tools/gridGliderData.m)), designed to get the glider trajectory data over instantaneous homogeneous regular profiles. By default, the vertical resolution (or step) is set to 1 meter in the present version of the processing, though it can be adapted by the user. For the spatial and temporal coordinates: they are computed as the mean values of the cast readings. For the variables: a binned is performed, taking the mean values of readings in depth intervals centered at selected depth levels.

These explanations are not in the new manuscript in the Section dedicated to the Processing levels.

4. Nowhere the sensor configurations are specified. I think a table gathering this information is worth it. For each platform, the list of sensor should be presented with their configuration (sampling frequency, ADCP ping-per-ensemble, ADCP vertical bin size, etc.). This should include all variables collected, for example, from the ship meteo station from which little information (or none) is present in the text. Same for the glider where there is Chl-a and turbidity data in the files, but these were not mentioned in the text. A table gathering this information would be useful.

We agree with the suggestion and provided this information in the manuscript. Instead of a table, we feel it is better to have the information distributed in each subsection referring to the different platforms. The manuscript has been modified accordingly.

5. A table regrouping all the platform with their basic configuration as well as their number of casts (when it applies) should be provided (sort of extended Table 3).

For each platform, we indicated the basic configuration as well as the number of casts (for CTD, gliders and Argo floats).

## Text-specific comments

- Figure 1 too small (should take page width) - Figure 2 too small (should take page width)

Figures 1 and 2 have been enlarged in the new manuscript

- Figure 2 caption: there is mention of "flag data equal to 1" while these flag are not introduced in the text.

SST is not part of the dataset, we just use them to illustrate the situation during the mission, this is why we did not go into details concerning the flag = 1, which is explicitly described in the caption (good data).

- p.7, L1: The "total number of valid measurement" is not very useful. I would rather put the number of valid casts (see comment above on a new table with this info).

We agree. The number of valid measurements (for the gliders) has been removed and replaced by the number of casts, in the new manuscript.

- p.7, L6: "a spatial interpolation is applied on the original data, leading to the so-called Level-2 data, further described in Sec. 3.3." What does 'spatial interpolation' means? Section 3.3 is not very explicit on this. I know you mean that the glider yos have been separated into downward and upward casts and then assigned to a geographical coordinate, but maybe this should be stated explicitly (and I don't think "spatial interpolation" is an accurate description). Moreover, Is there any vertical interpolation done? Because there are still some NaNs in L2

data.

The referee is right, it is not exactly an interpolation that is performed, but a spatial gridding. The gridding is performed by the function `gridGliderData`, designed to get the glider trajectory data over instantaneous homogeneous regular profiles. By default, the vertical resolution (or step) is set to 1 meter. For the spatial and temporal coordinates: they are computed as the mean values among cast readings. For the variables: a binning is performed, taking the mean values of readings in depth intervals centered at selected depth levels. The NaN are indeed not removed by the binning process, but will be discarded or flagged once the file are re-processed with the new version of the Glider Toolbox.

This has been amended in the new manuscript, in the section that describes the different processing levels.

Added text:

**Level 2 (L2)** : this level is only available for the gliders. It consists of regular, homogeneous and instantaneous profiles obtained by gridding the L1 data. In other words, 3-dimensional trajectories are transformed into a set of instantaneous, homogeneous, regular profiles. For the spatial and temporal coordinates: the new coordinates of the profiles are computed as the mean values of the cast readings. For the variables: a binning is performed, taking the mean values of readings in depth intervals centered at selected depth levels. By default, the vertical resolution (or bin size) is set to 1 meter. This level was created mostly for visualization purposes.

- p.7, L15: "Interestingly, all the drifters exhibit a trajectory close to the front position" → Not clear what "trajectory close to the front means". Moreover, is that really surprising that surface drifter would aggregate on a front?

We remove the "Interestingly", as indeed it is expected and rephrased it to: "All the drifters moved along the front position (deduced from the SST images), until they encounter the Algerian Current".

- Figure 8 caption: "for the duration of the mission" → You mean the ship mission? Or the AlborEX campaign?

We meant for the AlborEX mission; this has been made explicit in the new manuscript. The caption now reads:

Added text:

Surface drifter trajectories. For the sake of simplicity and clarity, the temperature, when available, is only shown for the duration of the AlborEX mission (May 25-31, 2014)

- Figure 10: plots on the right column are of little information here (too low resolution to mean something), I would remove.

We agree that the resolution is not as good as the Arvor-C float, but for completeness we would prefer not to discard them.

- Table 1: "Period" should be replaced by "cycle length" as referred to in the text (Section 2.2.4).

Modified as suggested.

- Table 1: netCDF file for Provor-bio indicates deployment end date 2015-04- 24T12:02:59+00:00,

which is different from this table.

The correct date is indeed 2015-04-24T12:02:59+00:00. The table has been modified accordingly.

- Figure 11 caption: "quality flag" not defined.

Quality flag with a value of 1 (meaning "good data") is specified in the caption. We added a complete description in the text concerning this part.

- Section 3.3.1: A Section on processing levels, but they are not all provided. Why? I think all levels should be provided. This is related to a previous comment.

The origin of the initial decision of not providing the L0 data for all the files is twofold: For some platforms (gliders), the L0 files are rather large and contain many variables related to the platform engineering, not to oceanography. Even if the files were not provided through the Zenodo platform, they are still publicly available using the SOCIB thredds server. In the new version of the manuscript, we adopted a new way to distribute the data (the data catalog), in which the data files corresponding to all the processing levels are made available.

- p.14, Level 2 (L2): "obtained by interpolating the L1 data" → How L2 is obtained by "interpolating" L1? Isn't L1 cut into casts that makes L2?

Correct. It is not an interpolating but a gridding. The explanation of how this gridding is performed has been added to the manuscript.

- p.14, Level 2 (L2): "It is only provided for gliders, mostly for visualization and post-processing purposes: specific tools designed to read and display profiler data can then be used the same way for gliders." → Is there a problem with this sentence? I don't understand it.

We removed the part of the sentence starting with "post-processing purposes"

- Section 3.3.1 / Table 3: Is L1 level for float equivalent to L2 level for glider? For consistency, I think profiling float should have L1 and L2 data as well since these instruments have similarities on the way they profile the water column. . .

The L1 glider data consists of a 3-dimensional trajectories, which means that both the longitude, latitude and depth change with respect to time. The Level 2 aims to have the same data on vertical profiles: the longitude and latitude don't change for a given profile. This is illustrated in the figure below.

- p.12, L1: "This type of current measurements requires a careful processing in order to get meaningful velocities from the raw signal" → Why? What are the limitations that makes this instrument more sensitive compare to other ones?

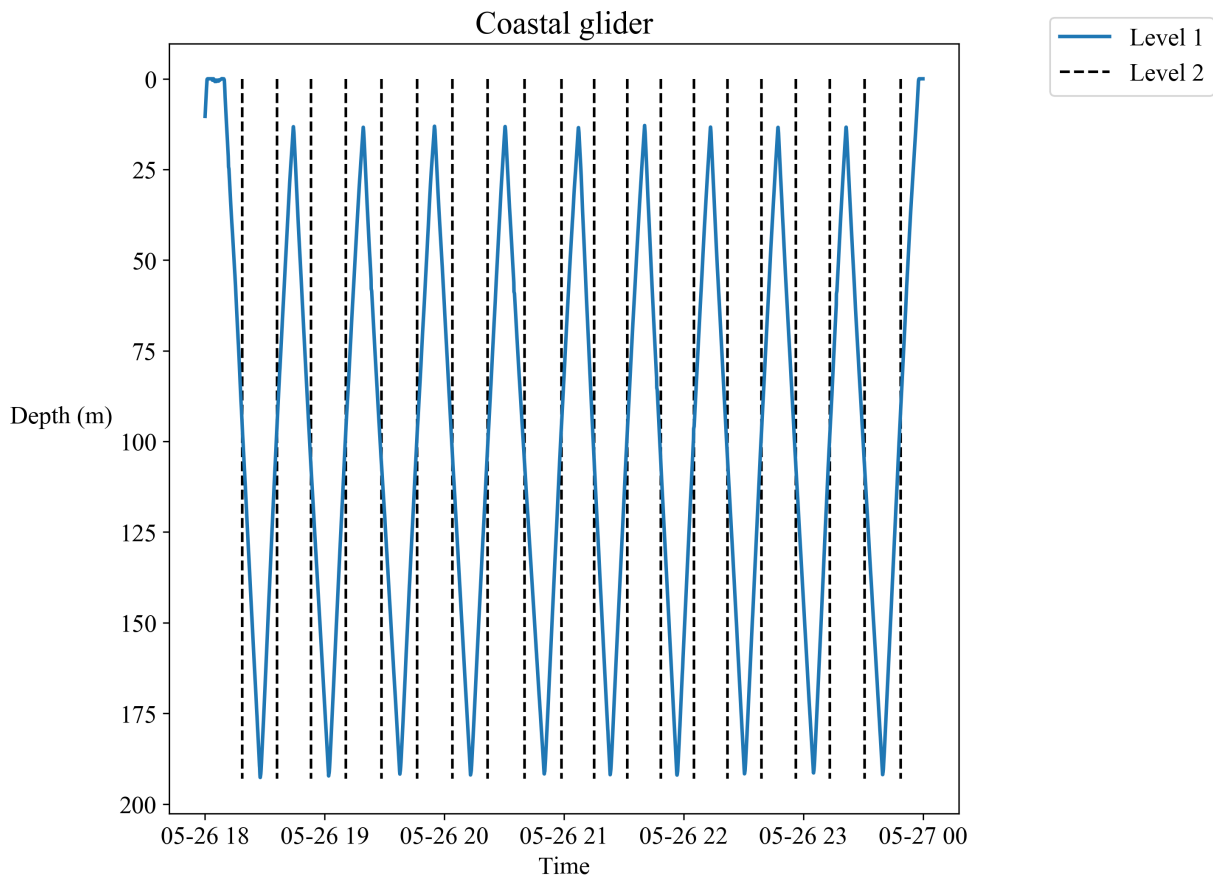
The main reason for this sensitivity is the fact that the vessel's velocity is one or two order or magnitudes greater than the currents that have to be measured. It is thus critical to have good measurements of the vessel heading and velocity.

A sentence has been inserted at the beginning of that paragraph and we removed the sentence "*hence it is relevant to have a quality flag (QF) assigned to each measurement*".

- p.12, L4: "Figure 12 shows the QF during the whole mission." → How QF are calculated?

The QC procedure for the VM-ADCP is complex as it involves tests on a large number of variables such as:

Bottom Track Direction



Bottom Track Velocity

Bottom Track error on velocity

Bottom Track Depth from beam

Sea water noise amplitude

...

with dependencies between them but also variables related to the vessel position and behavior (pitch, roll, speed, ...). The tests adopted are listed in the reference QUID document:

[QUID\\_DCF\\_SOCIB-QC-procedures.pdf](#)

SOCIB Quality Control Procedures

Data Center Facility

September 2018

DOI: [doi:10.25704/q4zs-tspv](https://doi.org/10.25704/q4zs-tspv)

and the new manuscript now contains a summary of the ADCP QC procedure.

Added text:

The vessel's velocity is one or two order or magnitudes greater than the currents that have to be measured, hence this type of current measurements requires a careful processing in order to get meaningful velocities from the raw signal. The QC procedure for the VM-ADCP is complex as it involves tests on more than 40 technical and geophysical variables (SOCIB Data Center, 2018). The different tests are based on the technical reports of Cowley et al. (2009) and Bender and DiMarco (2009), which aim primarily at ADCP mounted on moorings. The procedure can be summarised as follows:

1. Technical variables: valid ranges are checked for each of these variables: if the measurement is outside the range, the QF is set to 4 (bad data). Example of technical variables are: bottom track depth, sea water noise amplitude, correlation magnitude.
2. Vessel behaviour: its pitch, roll and orientation angles are checked and QF are assigned based on specific ranges. In addition the vessel velocity is checked and anomalously high values are also flagged as bad.
3. Velocities: valid ranges are provided for the computed current velocities: up to 2 m/s, velocities considered as good; between 2 and 3 m/s, probably good, and above 3 m/s, bad.

- Figure 12: Too small.

the figure has been enlarged in the new manuscript.

- Figure 12 and text below: 9 different quality flag are presented without any introduction on how they are calculated. The new paragraph in the same section (see comment before) now explains how the quality flag are assigned.

- Section 3.3.2 is very short. Should be re-worked following comments above. We agree that the section dedicated to the Quality Control was too short. The QC are now described as follows: A general description in Section "2.5.2 QC tests" and Specific explanations of the tests performed for each platform, making that part more self-contained.

## Comments on data files

The dataset consists of a relatively large number of files. I did my best but it was nearly impossible to review them all in details.

We really appreciate your time to extensively check of the files.

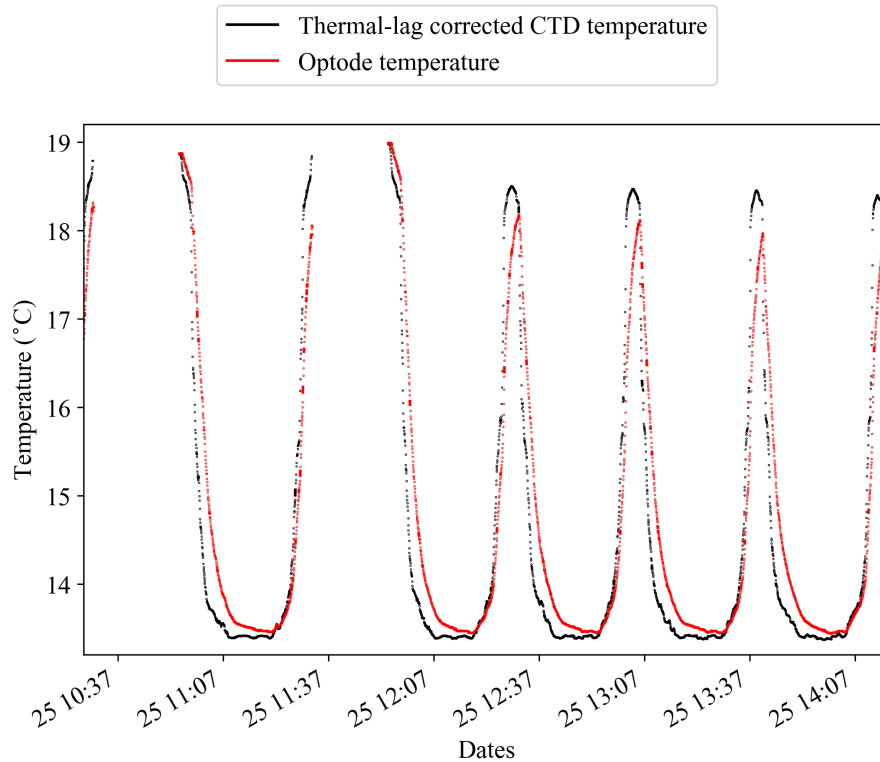
Here are some comments: - There are very large spikes in deep glider turbidity

yes, as the provided datasets for gliders have not undergone the quality checks (yet), there are still spikes and bad values for some of the variables. The text has been modified accordingly.

- There are missing data for about 10h in deep glider data between May 25-26. Unless I missed it, no explanation for this are provided.

The referee is right, some data are missing because the glider payload suffered an issue with the data logging software, resulting in no data acquisition during a few hours, during which the problem was being fixed. After that the data acquisition could be resumed.

This explanation has been added to the corresponding section in the new manuscript.



Added text:

On May 25 at 19:24 (UTC), the deep glider payload suffered an issue with the data logging software, resulting in no data acquisition during a few hours, during which the problem was being fixed. After this event, the data acquisition could be resumed on May 26 at 08:50 (UTC).

- Oxygen data for both glider seems to suffer from thermal lag problems

Yes it is true, we have reached the same conclusion when checking the oxygen data. The issue comes from the sensitivity of the optode to the temperature and the time response of the temperature sensor.

Comparing the temperature obtained with the glider CTD and the temperature of the oxygen sensor (next Figure) also highlights the lag existing between the 2.

To the extent of our knowledge, there is not yet an agreement from the community on how to correct this lag. Nicholson and Feen (2017) proposed a calibration based on the measurements made with the glider optode of the oxygen partial pressure of the atmosphere. Such a procedure can be contemplated in the near future.

Added text:

Finally, oxygen concentration measurements (not shown here) seem to exhibit a lag. According to Bittig et al. (2014), this issue is also related to the time response of oxygen optodes. As far as we know, there is not yet an agreement from the community on how to correct this lag, this is why the data are kept as they are in the present version, though we don't discard an improvement of the glider toolbox to address this specific issue.

- Provor-bio datafile contains levels down to over 7000 m. Some problems are found: 1. Why such long level dimension?



The 7000 comes is the depth dimension, as shown by the "ncdump -h" output:

```
dimensions:  
  time = UNLIMITED ; // (71 currently)  
  depth = 7118 ;  
  name_strlen = 49 ;
```

But it does not mean that the maximal depth is actually 7000 m or deeper, as it depends on the vertical resolution. Here the deepest measurements are on the order of 1000 m. The profiles from PROVIO are shown in the next 2 figures.

2. No good data is found below 325m, although Table 1 suggest that the float is profiling to 1000m

We confirm that the float acquired data up to approx. 2000 m, even though the vertical resolution is not as high as near the surface. We reproduce (see below) the Figure 10 from the manuscript, this time without limiting the depth range, in order to confirm the availability of data at that depth.

- Arvor A3 data file suffers from similar problem: file contains data only down to 115m while Table 1 says 2000m

For the Arvor A3 we confirm that profiles are available up to approx. 2000 m. The "115" mentioned above are in fact the number of vertical levels provided in the file, not the final depth. Also see the figure above for the data availability.

- Arvor-C data file (only L0 provided) do not contain metadata (no file attributes, etc.). In addition, missing data (at least for temperature) appears to me as very large numbers (9.969210e+36) that makes them difficult to manipulate.

The L0 file with the metadata and the L1 file have been prepared and are now available. The link to the thredds catalog are provided below: L0: [http://thredds.socib.es/thredds/catalog/drifter/profiler\\_drifter/profiler\\_drifter\\_arvorc001-ime\\_arvorc001/L0/2014/catalog.html?dataset=drifter/profiler\\_drifter/profiler\\_drifter\\_arvorc001-ime\\_arvorc001/L0/2014/dep0001\\_profiler-drifter-arvorc001\\_ime-arvorc001\\_L0\\_2014-05-25.nc](http://thredds.socib.es/thredds/catalog/drifter/profiler_drifter/profiler_drifter_arvorc001-ime_arvorc001/L0/2014/catalog.html?dataset=drifter/profiler_drifter/profiler_drifter_arvorc001-ime_arvorc001/L0/2014/dep0001_profiler-drifter-arvorc001_ime-arvorc001_L0_2014-05-25.nc) L1: [http://thredds.socib.es/thredds/catalog/drifter/profiler\\_drifter/profiler\\_drifter\\_arvorc001-ime\\_arvorc001/L1/2014/catalog.html?dataset=drifter/profiler\\_drifter/profiler\\_drifter\\_arvorc001-ime\\_arvorc001/L1/2014/dep0001\\_profiler-drifter-arvorc001\\_ime-arvorc001\\_L1\\_2014-05-25.nc](http://thredds.socib.es/thredds/catalog/drifter/profiler_drifter/profiler_drifter_arvorc001-ime_arvorc001/L1/2014/catalog.html?dataset=drifter/profiler_drifter/profiler_drifter_arvorc001-ime_arvorc001/L1/2014/dep0001_profiler-drifter-arvorc001_ime-arvorc001_L1_2014-05-25.nc)

R/V Socib CTD and thermosalinograph files say that units of temperature are "C". I prefer the convention from glider files which uses "Celsius".

We take note of the suggestion and will perform the modification in a new release of the data files, as it involves a re-processing of several files from other missions). The referee is totally right, as the Unidata documentation (<https://www.unidata.ucar.edu/software/netcdf/netcdf/Units.html>) states that "Celsius" should be used, "C" meaning "Coulomb".

## Minor comments

- p.2; L23: "makes it possible" → makes possible

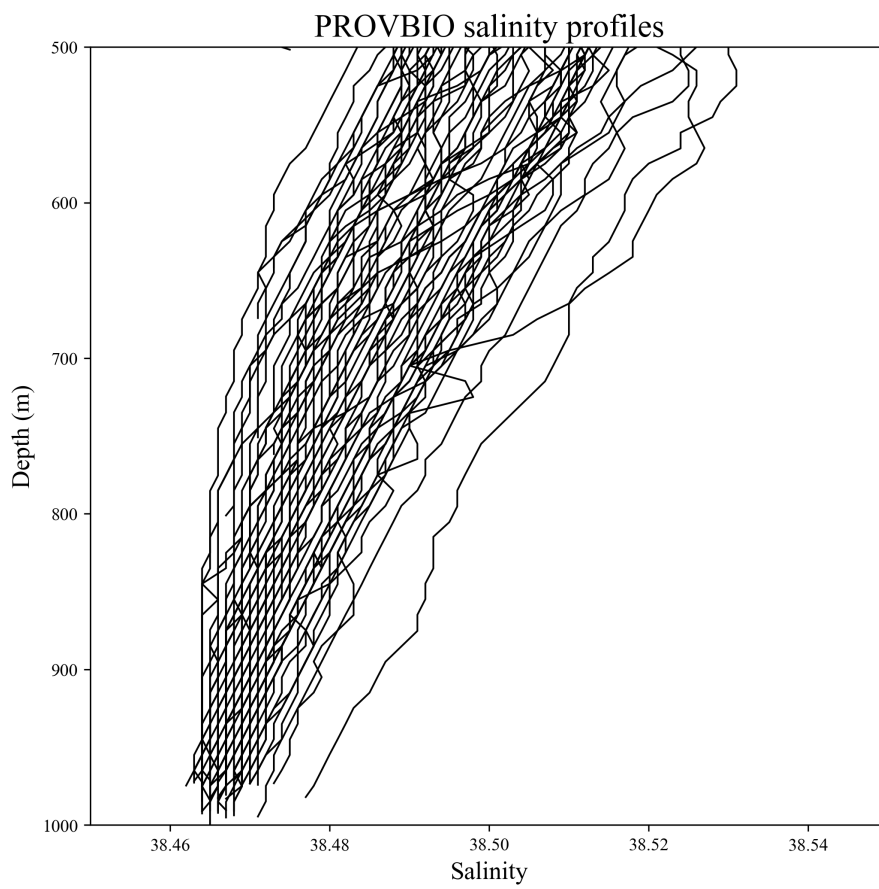
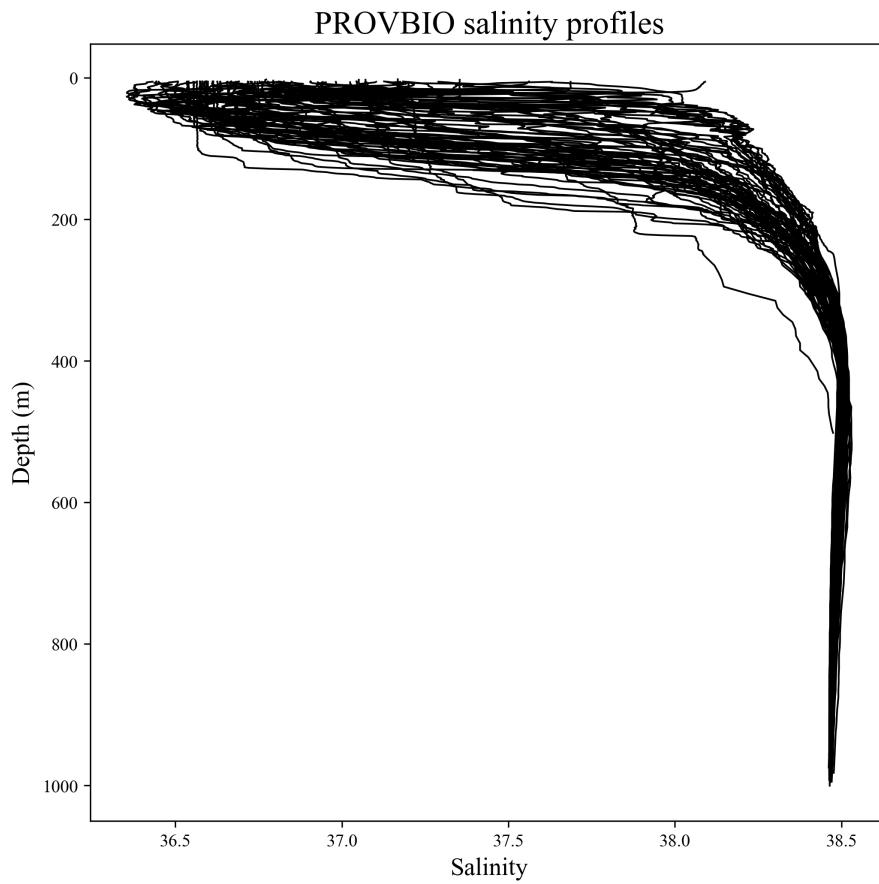


Figure 1: Salinity profiles acquired by the PROVBIO float. The 2nd panel depicts the profile in the 500–1000 m layer.

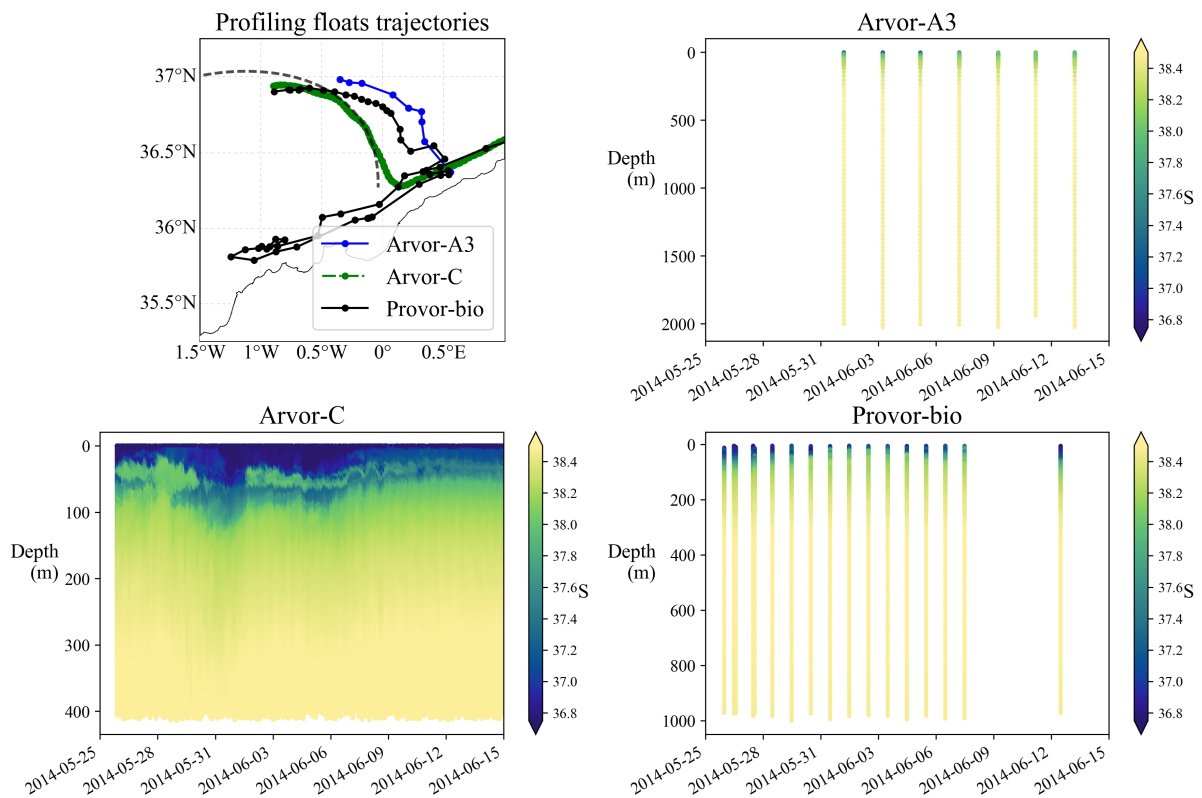


Figure 2: Adapted figure 10 of the manuscript, with the maximal depth of the profiles displayed.

corrected

- p.2; L23: "creation and publication of aggregated datasets covering the Mediterranean Sea" → SeaDataNet is not only about the Mediterranean replaced by "covering different European regional seas, including the Mediterranean Sea"

- p.2; L32: "thanks due to" → thanks to

corrected (removed "due")

- Section 2.2.1: "CTD surveys" or CTD legs?

corrected (legs)

- Glider L1 files (e.g. dep0012\_idEEP00\_ime-sldeep000\_L1\_2014-05-25\_data\_dt.nc) say that the project is "PERSEUS". Is that right? There is no mention of the AlborEX project in the file header.

Correct, AlborEx was the Subtask 3.3.4 of PERSEUS project, but in this case AlborEx was not explicitly mentioned in the file header. This will be added during the next re-processing of the data files.

- p.10, L1: problems with latitude longitude degree symbol.

corrected

- p.10, L5: temperature, salinity and T,S is use on the same line. Please homogenize.

replaced by "In addition to these variables"

- p.12, L17: "Network Common Data Form (netCDF, <https://doi.org/http://doi.org/10.5065/D6H70CW6>, last accessed on August) 3, 2018)" Is there a mis-placed parenthesis?

Corrected, the "(" after .org has been removed.

- p.13, L2: problem with file name (too long for page)

Corrected (new line added).

- p.16, L25: How stable in time the python codes made available on Github will be?

Generally, reading netCDF files with Python is an easy task, as it is with other languages (MATLAB, Julia, R), so we do not expect any difficulties for the data users. Here what we did is to provide a set of the Python codes written to show how to read the data and reproduce the plots of the papers, as we think it might save time if somebody wants to create something similar, or even reproduce the paper plot.

With Python it is relatively straightforward to use virtual environment, which allows one to work with specific version python modules. If a user works with a virtual environment which has the same packages versions as those specified on GitHub (file requirements.txt), then the code will run (since the netCDF files will be the same).

Even if issues occur, we think that providing the codes employed to manipulate the data files, along with the data, is a step toward the reproducibility of the results.

## Additional references

Bender, L. and DiMarco, S.: Quality Control and Analysis of Acoustic Doppler Current Profiler Data Collected on Offshore Platforms of the Gulf of Mexico, Tech. rep., U.S. Dept. of the Interior, Minerals Mgmt. Service, Gulf of Mexico OCS Region, New Orleans, LA, 66 pp., 2009.

Bittig, H. C., Fiedler, B., Scholz, R., Krahnemann, G., and Körtzinger, A.: Time response of oxygen optodes on profiling platforms and its dependence on flow speed and temperature, *Limnology and Oceanography: Methods*, 12, 617–636, doi:10.4319/lom.2014.12.617, URL <https://aslopubs.onlinelibrary.wiley.com/doi/abs/10.4319/lom.2014.12.617>, 2014.

Cowley, R., Heaney, B., Wijffels, S., Pender, L., Sprintall, J., Kawamoto, S., and Molcard, R.: INSTANT Sunda Data Report Description and Quality Control, Tech. rep., CSIRO, 2009.

Nicholson, D. P. and Feen, M. L.: Air calibration of an oxygen optode on an underwater glider, *Limnology and Oceanography: Methods*, 15, 495–502, doi:10.1002/lom3.10177, URL <https://aslopubs.onlinelibrary.wiley.com/doi/abs/10.1002/lom3.10177>, 2017.

SOCIB Data Center: SOCIB Quality Control Procedures, Tech. rep., Balearic Islands Coastal Observing and Forecasting System, Palma de Mallorca, Spain, doi:10.25704/q4zs-tspv, URL <http://repository.socib.es/repository/entry/show?entryid=a85d659d-b469-4340-ae88-c361333c68b6>, 2018.