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

Anonymous Referee #1

We wish to thank the Reviewer for their constructive comments that really underline the aspects of the paper that needed to be further developed.

This article (categorized as "review") by Troupin et al. is addressing a multidisciplinary data set collected in the western Mediterranean Sea during the AlborEX campaign. During the campaign in-situ observing devices (ships, floats, gliders, drifters...) have been used (described here) but also satellite data. In the manuscript some aspects of the data set are described. As it stands now I do not recommend publication in ESSD. For the review I followed the ESSD evaluation criteria and also considered the general scope of the journal (as described on the website).

First - Is this a "review" article? ESSD defines review articles as: "...may compare methods or relative merits of data sets, the fitness of individual methods or data sets for specific purposes, or how combinations might be used as more complex methods or reference data collections." As I read it from the manuscript this is not the case. The current version of the manuscript reads more as a copy of data information from individual reports and the data section in scientific publications related to the experiment. As it stands, I do not see the criteria for a "review" type article fulfilled.

We acknowledge the reviewer's comment concerning the nature of the article. We made a mistake during the submission process. Referring to the ESSD web page, we read that "Articles in the data section may pertain to the planning, instrumentation, and execution of experiments or collection of data.", and this is indeed the objective we had when submitting the manuscript. However in the Submission page, the "Manuscript Type" did not offer the possibility to select it, hence we took another one which seemed the closest. We have contacted the editorial office concerning this and the manuscript type was changed on September 18, 2018.

Significance

Three sub-criteria to evaluate:

• Uniqueness: It should not be possible to replicate the experiment or observation on a routine basis. Thus, any data set on a variable supposed or suspected to reflect changes in the Earth system deserves to be considered unique. This is also the case for cost-intensive data sets which will not be replicated due to financial reasons. A new or improved method should not be trivial or obvious. The data set is unique. (rating: 1 Excellent)

Thank you for the appreciation

• Usefulness: It should be plausible that the data, alone or in combination with other data sets, can be used in future interpretations, for the comparison to model output or to verify

other experiments or observations. Other possible uses mentioned by the authors will be considered.

The current manuscript does not provide information that promote the reuse of the data set (it may for subsets). No attempt is made to provide a structured overview about the workflow that is linked to the creation of the data set and, equally important, the QA/QC are not provided in a transparent way. For example, in the netcdf data files I see different QC flags provided – one is for example "SOCIB Quality control Data Protocol". What does that mean? This is not an international standard. A data set description, as envisioned in this ESSD submission, should exactly describe such non-standard QC procedures. Which QA and QC methods were applied (give brief description, DOIs if applicable)?

We agree with the reviewer and to address these issues:

- A new section dedicated to data reuse has been be added (see below) and
- the section *"3.3.2 Quality control"* has been expanded and made more explicit.

Added text:

Data Reuse

Three main types of data reuse are foreseen: 1. model validation, 2. data assimilation (DA) and 3. planning of similar in situ experiments.
With the increase of spatial resolution in operational models, the validation at the smaller scales requires high-resolution observations. Remote-sensing measurements such as SST or chlorophyll-a concentration provides a valuable source of information but are limited to the surface layer. In the case of the present experiment, the position, intensity (gradients) and vertical structure of the front represent challenging features for numerical models, even when data assimilation is applied (Hernandez-Lasheras and Mourre, 2018)).
The AlborEx dataset can be used for DA experiments, for example assimilating the CTD measurements in the model and using the glider measurements as an independent observation dataset. The assimilation of glider observations has already been performed in different regions (e.g. Melet et al., 2012; Mourre and Chiggiato, 2014; Pan et al., 2014) and has been shown to improve the forecast skills. However the assimilation of high-resolution data is not trivial: the the background error covariances tends to smooth the small scale features present in the observations.
Finally, other observing and modeling programs in the Mediterranean Sea can also benefit from the present dataset, for instance the Coherent Lagrangian Pathways from the Surface Ocean to Interior (CALYPSO) in the Southwest Mediterranean Sea (Johnston et al., 2018). Similarly to AlborEx, CALYPSO strives to study a strong ocean front front and the vertical exchanges taking place in the area of interest (see `https://www.onr.navy.mil/Science-Technology/Departments/Code-32/All-Programs/Atmosphere-Research-322/Physical-Oceanography/CALYPSO-DRI` for details).

We also complement the introduction with references to other studies using multi-platform approaches in the same area.

Added text:

Similar studies comparing almost synchronous glider and SARAL/AltiKa altimetric data on selected tracks have also been carried between the Balearic Islands and the Algerian coasts (Aulicino et al., 2018; Cotroneo et al., 2016).

I also miss any information how/if this data is disseminated via international data centres and how the data QC and dissemination is coordinate with the respective observing networks (Argo, DBCP, ...). Seadatanet is been mentioned in the text but it is unclear which specific recommendations are given.

(rating: 4 poor)

All the data presented in this paper are open data and can be accessed through the SOCIB Data Center in a few clicks, without any registration. Moreover, the data API (http://api.socib.es) strongly improves the data access to users and the dissemination to national or international data centers, which can easily establish a data transfer if they want to include SOCIB data into their portal.

As of today, many international databases exist and frequently, new ones are created with new projects, making the data landscape complex and the making it tedious to extensively document the data flow between SOCIB data and those databases. For instance:

- all the drifters data are transmitted to the Mediterranean Surface Velocity Programme (MedSVP, http://doga.ogs.trieste.it/sire/medsvp/);
- Most of the data are transmitted to the Mediterranean Operational Network for the Global Ocean Observing System (MONGOOS, http://www.mongoos.eu/data-center);
- MONGOOS sends the data to the In Situ Thematic Assembly Center (INSTAC) of the Copernicus Marine Environment Monitoring Service (CMEMS, http://www.marineinsitu.eu);
- The PROVBIO float is available in OAO database (Villefranche-sur-mer, http://www.oao.obs-vlfr.fr/maps/en/
- The Argo floats and drifters data are transmitted to the CMEMS INSTAC.
- ...

Our approach to guarantee that the data are available to the widest community consists of

1. Having the data easily accessible in a standard format (netCDF) through standard protocols (HTTP, OPEnDAP, ...), and without any registration. This means that any user or entity can download all the files and include them in their portal or database.
2. Providing an efficient data API to make easier the data discovery: the role of the API is really to allow users to make request such as:
   - "provide me all the observations measured by the platform X (glider, drifter)" or
   - "provide me all the observations in the region located in the area Y during a given time period."

The explicit mention to SeaDataNet is made because of their Regional Data Products, which we believe are of crucial importance for the scientific community needing a complete set of historical, in situ data. The data transfer from SOCIB to SeaDataNet is foreseen in the future.

• Completeness: A data set or collection must not be split intentionally, for example, to increase the possible number of publications. It should contain all data that can be reviewed without unnecessary increase of workload and can be reused in another context by a reader.

It is difficult to evaluate this point. However, the nutrient data is not mentioned but is, according to Pascual et al. 2017 part of the AlborEX campaign. I would expect that these data set are described here as well (and respective QC (e.g. GO-SHIP nutrient manual??) and associated uncertainty estimates.

(rating: 2 to 3)

We agree with this suggestion and will add a specific section dedicated to the nutrient data. In relation to these data, we wish to add to the list of co-authors:

- Antonio Tovar-Sánchez, Instituto de Ciencias Marinas de Andalucía, (ICMAN – CSIC), Puerto Real, Spain and
- Eva Alou, SOCIB,

who were responsible for the acquisition and processing of these data during and after the cruise.

We have now included the dissolved inorganic nutrients measured during Alborex in the new file AlborexPerseus2014_LabSamplesNutrients_L1.nc, available at `https://repository.socib.es:8643/repository/entry/show?entryid=07ebf505-bd27-4ae5-aa43-c4d1c85dd500`. The files still has to be included to the general thredds directory of SOCIB.

This text was added to the new manuscript:

Added text:

Samples for nutrient analysis were collected in triplicate from CTD Niskin bottles and immediately frozen for subsequent analysis at the laboratory. Concentrations of dissolved nutrients (Nitrite: NO_2^-, Nitrate: NO_3^- and Phosphate: PO_4^{3-} were determined with an autoanalyzer (Alliance Futura) using colorimetric techniques (Grasshoff et al., 1983). The accuracy of the analysis was established using Coastal Seawater Reference Material for Nutrients (MOOS-1, NRCCNRC), resulting in recoveries of 97%, 95% and 100% for NO_2^-, NO_3^- and PO_4^{3-}, respectively. Detection limits were NO_2^-:0.005 μM, NO_3^-: 0.1 μM and PO_4^{3-}: 0.1 μM.

Data quality

The data must be presented readily and accessible for inspection and analysis to make the reviewer's task possible. Even if a data set submitted is the first ever published (on a parameter, in a region, etc.), its claimed accuracy, the instrumentation employed, and methods of processing should reflect the "state of the art" or "best practices". Considering all conditions and influences presented in the article, these claims and factors must be mutually consistent. The reviewer will then apply his or her expert knowledge and operational experience in the specific field to perform tests (e.g. statistical tests) and cast judgement on whether the claimed findings and its factors – individually and as a whole – are plausible and do not contain detectable faults.

I touched on that already under "Usefulness". In the manuscript no transparent QC assessment is presented. What were the methods of processing (provide key steps, DOI at least). What were, including quantification of uncertainties and qualification via flags, the results of the QA/QC procedures? Which were the major shortcomings of the data acquisition process and what could be done better in the future? For example, has the drifter data included in the European E–SurfMar data base and also in the DBCP global drifter data sets? Have the recommendations (Best Practices, Protocols) from E–SurfMar / DBCP considered? It looks like no commonly agreed standard has been used for some parameters – as "SOCIB Quality control Data Protocol" suggest? (rating: 3)

The QC procedure is described in the document

QUID_DCF_SOCIB-QC-procedures.pdf

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

The procedure in based on the commonly agreed standards.

The article has been re-organised and for each type of platform, a description of the quality checks performed on the corresponding data has been added.

Which were the major shortcomings of the data acquisition process and what could be done better in the future?

Possibly the glider sampling strategy could be improve by increasing the relative frequency of surfacing, in order to have more information on the variables near the surface.

Presentation quality

Long articles are not expected. Regarding the style, the aim is to develop stereotypical wording so that unambiguous meaning can be expressed and understood without much effort. The article should express clearly what has been found, where, when, and how. The article text and references should contain all information necessary to evaluate all claims about the data set or collection, whether the claims are explicitly written down in the article, or implicit, through the data being published or their metadata. The authors should point to suitable software or services for simple visualization and analysis, keeping in mind that neither the reviewer nor the casual "reader" will install or pay for it.

mostly OK (given the limitation outlined in the previous points). It would be useful to include a brief introduction into the "design of the experiment. Visualisation tools are not given. (rating: 2-3)

A section "*Design of the experiment has been added*" in Section 2, after the "*General oceanographic context*" References to existing visualisations tools have been provided in a new section "*4.3 Data reading and visualisation*". It is worth mentioning here that a set of Python functions are provided to read, process and visualise the content of type of file.

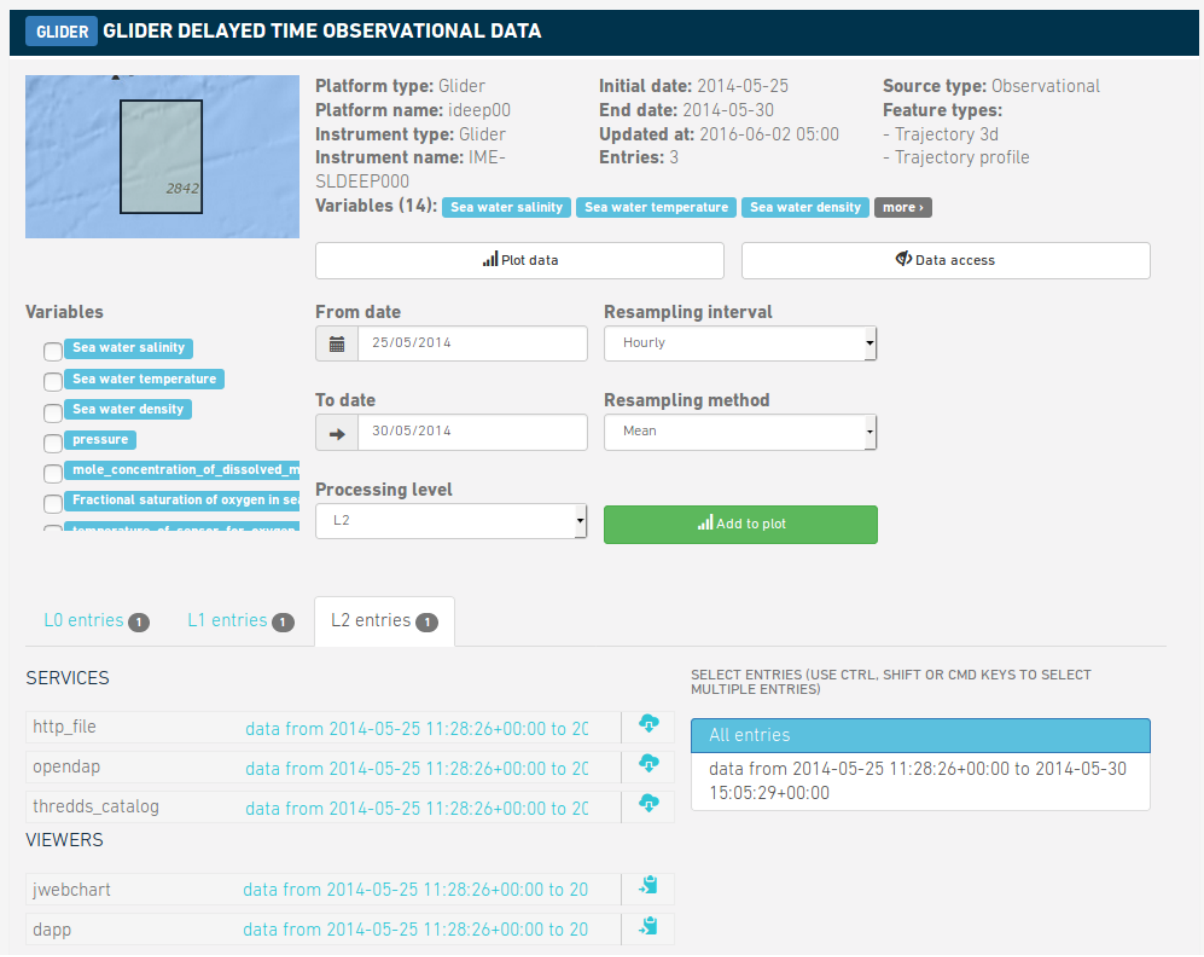

Figure 1: Access to the deep glider data: the in-house viewers are listed in the bottom left corner.

Added text:

Design of the experiment

The deployment of in situ systems was based on the remote-sensing observations described in the previous Section. Two high-resolution grids were sampled with the research vessel, covering an approximative region of 40 km $\times$ 40 km. At each station, one CTD cast and water samples for chlorophyll concentrations and nutrients analysis were collected. The thermosalinograph observations were also used in order to assess the front position.
One deep glider and one coastal glider were deployed in the same area with the idea to have butterfly-like track across the front. These idealised trajectories turned out to be impossible considering the strong currents occurring in the region of interest at the time of the mission. The 25 drifters were released close to the frontal area with the objective to detect convergence and divergence zones. Their release locations were separated by a few kilometers.

Also, when accessing the data through the catalog (doi:10.25704/z5y2-qpye), users have access to different viewers (depending on the type of data), in one click, as shown in the figure below.

The following paragraph has been added:

Added text:

When accessing the data catalog, users are provided a list of in-house visualisation tools designed to offer quick visualisation of the file content. The visualisation tools depend on the type of data: *JWebChart* is used for time series; *Dapp* displays the trajectory of a moving platform on a map; the *profile-viewer* allows the user to select locations on the map and view the corresponding profiles.

Specific comments

P2/l.4: I do not agree with the statement: "a perfect observational system would consist in dense array of sensors present at many geographical locations, many depths and measuring almost continuously a wide range of parameters. . . " – this "generalization" is trivial and useless. From an observing design point of view a "perfect" observing system must follow a design that will record only the observations that are needed to analyse the problem. As such the perfect observational system always depends on motivation for the experiment (or the problem in more general words) - in some cases a "perfect observing system" may comprise only one single sensor at one single depth at different locations if this has been found a sufficient approach for solving the problem (e.g. estimating global warming through a global tomography array). Please reformulate the statement along those lines.

We agree that this formulation was not adequate and rephrased this part following this comment, as follows:

Added text:

To properly capture and understand these small-scale features, one cannot settle for only observations of temperature and salinity profiles acquired at different times and positions, but rather has to combine the information from diverse sensors and platforms acquiring data at different scales and at the same time, similarly to the approach described in Delaney and Barga (2009). This also follows the recommendation for the Marine Observatory in Crise et al. (2018), especially the co-localization and synopticity of observations and the multi-platform, adaptive sampling strategy. We will refer to this as multi-platform systems, by opposition to experiments articulated only around the observations made using a research vessel. Further details can be found in Tintoré et al. (2013).

Additional references

Aulicino, G., Cotroneo, Y., Ruiz, S., Sánchez Román, A., Pascual, A., Fusco, G., Tintoré, J., and Budillon, G.: Monitoring the Algerian Basin through glider observations, satellite altimetry and numerical simulations along a SARAL/AltiKa track, Journal of Marine Systems, 179, 55–71, doi:10.1016/j.jmarsys.2017.11.006, URL https://www.sciencedirect.com/science/article/pii/S0924796317302658, 2018.

Cotroneo, Y., Aulicino, G., Ruiz, S., Pascual, A., Budillon, G., Fusco, G., and Tintoré, J.: Glider and satellite high resolution monitoring of a mesoscale eddy in the algerian basin: Effects on the mixed layer depth and biochemistry, Journal of Marine Systems, 162, 73–88, doi:10.1016/j.jmarsys.2015.12.004, URL https://www.sciencedirect.com/science/article/pii/S0924796315002298, 2016.

Crise, A., Ribera d'Alcalà, M., Mariani, P., Petihakis, G., Robidart, J., Iudicone, D., Bachmayer, R., and Malfatti, F.: A Conceptual Framework for Developing the Next Generation of Marine OBservatories (MOBs) for Science and Society, Frontiers in Marine Science, 5, 1–8, doi:10.3389/fmars.2018.00318, URL https://www.frontiersin.org/articles/10.3389/fmars.2018.00318/full, 2018.

Delaney, J. R. and Barga, R. S.: Observing the Oceans - A 2020 Vision for Ocean Science, pp. 27–38, Microsoft Research, URL https://www.microsoft.com/en-us/research/publication/observing-the-oceans-a-2020-vision-for-ocean-science/, 2009.

Grasshoff, K., Kremling, K., and (Eds), M. E., eds.: Methods of Seawater Analysis, , doi:10.1002/9783527613984, URL https://onlinelibrary.wiley.com/doi/book/10.1002/9783527613984, 1983.

Hernandez-Lasheras, J. and Mourre, B.: Dense CTD survey versus glider fleet sampling: comparing data assimilation performance in a regional ocean model west of Sardinia, Ocean Science, 14, 1069–1084, doi:10.5194/os-14-1069-2018, URL http://dx.doi.org/10.5194/os-14-1069-2018, 2018.

Johnston, T. M. S., Rudnick, D. L., Tintoré, J., and Wirth, N.: Coherent Lagrangian Pathways from the Surface Ocean to Interior (CALYPSO): Pilot Cruise report, Tech. rep., Scripps Institution of Oceanography (SIO), URL http://scrippsscholars.ucsd.edu/tmsjohnston/files/calypssocibcruisereport2018.pdf, last accessed: December 17, 2018, 2018.

Melet, A., Verron, J., and Brankart, J.-M.: Potential outcomes of glider data assimilation in the Solomon Sea: Control of the water mass properties and parameter estimation, Journal of Marine Systems, 94, 232–246, doi:10.1016/j.jmarsys.2011.12.003, URL http://dx.doi.org/10.1016/j.jmarsys.2011.12.003, 2012.

Mourre, B. and Chiggiato, J.: A comparison of the performance of the 3-D super-ensemble and an ensemble Kalman filter for short-range regional ocean prediction, Tellus A: Dynamic Meteorology and Oceanography, 66, 21 640, doi:10.3402/tellusa.v66.21640, URL http://dx.doi.org/10.3402/tellusa.v66.21640, 2014.

Pan, C., Zheng, L., Weisberg, R. H., Liu, Y., and Lembke, C. E.: Comparisons of different ensemble schemes for glider data assimilation on West Florida Shelf, Ocean Modelling, 81, 13–24, doi:10.1016/j.ocemod.2014.06.005, URL http://dx.doi.org/10.1016/j.ocemod.2014.06.005, 2014.

Tintoré, J., Vizoso, G., Casas, B., Heslop, E., Pascual, A., Orfila, A., Ruiz, S., Martínez-Ledesma, M., Torner, M., Cusí, S., and et al.: SOCIB: The Balearic Islands Coastal Ocean Observing and Forecasting System Responding to Science, Technology and Society Needs, Marine Technology Society Journal, 47, 101–117, doi:10.4031/mtsj.47.1.10, URL http://www.ingentaconnect.com/content/mts/mtsj/2013/00000047/00000001/art00010;jsessionid=2cbcvta0m97c.x-ic-live-02, 2013.