

Referee Comment of:

An internally consistent data product for the world ocean: the Global Ocean Data Analysis Project, version 2 (GLODAPv2)
by Olsen et al.

The paper describes the GLODAPv2 data product starting from its origins, the observing programs and predecessors products/projects that allowed the creation of a unique data set from mainly ship based surveys arising from the 70's. The main objective of GLODAPv2 was to provide to the scientific community a unique global data set of consistent measurements applying an objective methodology to quality assess, document and publish the data. Twelve core parameters (salinity, oxygen, nitrate, silicate, phosphate, TCO₂, TAlk, pH, and the four halogenated transient tracer species) passed through primary Quality Control (QC) that identified possible outliers and assigned defined QC flags. Most of them were also subjected to secondary QC analyses to detect and correct biases through computed adjustments (internal consistency analysis). The secondary QC methodology developed within the framework of three predecessor projects (GLODAPv1.1, CARINA and PACIFICA) and the strategy applied consisted in the complete re-processing of the entire data set including also new surveys, making use of all the experience/knowledge gained during the previous years of data analysis. The secondary QC applied automated crossover and inversion scripts to define data adjustments but preserving the real long-term signals like time trends. This represents the foundation for climate studies.

The manuscript presents an overview of the secondary QC results by parameters and the detailed online adjustment table where the user might find information per cruise on the recommended adjustments to the twelve core parameters and comments regarding their justification. Proof of the final data set increased internal consistency is also supplied.

Two kinds of products are made available: 1) the original data updated to WODE exchange format; 2) the bias corrected datasets (one global plus 4 regional sub-sets) as ASCII or binary format files.

Finally the authors give useful recommendations for data usage and some lesson learned.

General comments on the overall quality of the discussion paper

The paper presents a unique scientist driven data synthesis generating well documented, qualified and internally consistent data products.

However the manuscripts appears very long and full of details that compromise its readability. Sometimes the excess of details distracts the reader from the main focus. This happens mainly in the first part of the manuscripts where the structure is confused since mainly driven by temporal progression of facts. A clear schema with all the historical background, the objective and a description of the paper outline in the introduction, would be preferable.

The development of the secondary QC methodology applied to GLODAPv2 product is mixed within the description of the project predecessors, including information that is more appropriate to a project report. Also the description of the production of GLODAPv2 might be shortened and better organized.

I understand the huge work behind the production of GLODAPv2 and the effort to put together all this documentation, however the paper now looks confused.

I recommend the paper for publication after a minor revision. It will follow a series of suggestions that might ameliorate it.

Specific Comments and technical corrections

I suggest including a table with the ACRONYMS.

Introduction

Page 4, Line 12: Start a new line "The main goal..."

At the end of the introduction I would insert the outline of the paper.

Page 5, Line 22: Start a new line "GLODAPv2 is primarily..."

Page 6, Line 16: the phrase "the GLODAPv1.1...(SIO)." is not clear to the reviewer. It is very specific, is it necessary? If yes please rephrase.

Page 7, Line 24: What is Table 1, what's its meaning? I would insert a phrase to introduce it.

Page 8, Line 13-20: Here you describe the CARINA effort, are these details necessary here?

Page 9, Line 18: is it table 4 necessary? You already refer to the special issue, I believe it is redundant. You provide 13 Tables that are a lot.

Page 10, Lines 9-19: Are these details necessary here?

Page 10, Lines 19: is it table 5 necessary? 13 Tables that are a lot, you might put this in the Supplements.

Page 10, last line: the initial minimum adjustment limits of Table 6 are those introduced in CARINA (page 9, line10)? Please specify here and in the table caption.

Section 3

I recommend synthesizing the strategy steps, including the references to the sub-sections and leaving all the details to the sub-sections. Example:

- 1. Identify and ingest data not included in GLODAPv1.1, CARINA, or PACIFICA (GLODAPv2NEW, Section 3.1)*
- 2. Re-evaluate GLODAPv1.1 using CARINA analysis tools (GLODAPv1.2, Section 3.2)*
- 3. Combine GLODAPv1.2 with CARINA and PACIFICA (Section 3.3)*

...

A new figure with the schematic of GLODAPv2 production is welcome!

Page 12, Lines 18-end of page: are these details important? First phrase is a repetition.

Page 13, First Line: you refer to a Supplement, which one? I looked here <http://www.earth-syst-sci-data-discuss.net/essd-2015-42/> and I got this, please check.



DOI Not Found

10.5194/essd-2015-42-supplement

This DOI cannot be found in the DOI System. Possible reasons are:

- The DOI is incorrect in your source. Search for the item by name, title, or other metadata using a search engine.
- The DOI was copied incorrectly. Check to see that the string includes all the characters before and after the slash and no sentence punctuation marks.
- The DOI has not been activated yet. Please try again later, and report the problem if the error continues.

You may report this error to the responsible DOI Registration Agency using the form below. Include your email address to receive confirmation and feedback.

DOI:

URL of Web Page Listing the DOI:

End of page 15: The 7 distinct scenarios are well summarized in Table 9, why don't you refer to it here and shorten the text? Why don't you insert here the table and the results of the salinity and oxygen pre-calibration (page 27 lines 10-21, page 28 lines 9-19)? It could become section 3.5 Pre-calibration of Salinity and Oxygen Data, 3.5 becomes 3.6 and 3.6 becomes 3.7. Is it pre-calibration part of a primary or secondary QC?

Page 16, Lines 11-21: Might this fit better in the Lesson Learned section?

Page 17, Lines 5-15: At the end of section 3.4, after 5 pages of description of your strategy, you concluded to reset the entire database and carry out secondary QC on unadjusted data. Isn't it better to start from this? I would emphasize this since at the beginning thus I suggest revising and shortening Section 3 considering my previous comments.

Page 20 Line 25: "As an example our method..." please correct "As an example of our method..."

Page 21, Line 5: "...the bias of these these data." Please take out the repetition.

Page 22, line 20: I suggest to end the phrase after "...incomprehensible."

Page 22, point 1: Why do you assign different values (0 or 1) to good quality data? Moreover this list could be ameliorated taking out unnecessary words and highlighting the flags assigned(-888,-666,-777...). A Table would be preferable.

Page 22, Line 24: I would add here "...in the table. A comment was not always entered when the data appeared unbiased. Some of the comments might also refer to workshops where the magnitudes of adjustments were discussed and decided.

"The GEOMAR Adjustment Table gives the dataset source of each cruise: CARINA, PACIFICA, GLODAPv1.2, or GLODAPv2 (NEW). When accessing the table be aware of the following:

- for CARINA cruises

- for PACIFICA...
- for GLODAPv1.2..
- for GLODAPv2 (NEW)...

Comments for CFC and PH parameters are either inherited from CARINA or from the data processing described in sections 3.5 and 3.6."

I would take out last item at the beginning of page 24.

Page 22, last line (same at page 23): Why GLODAPv1.2 appears as a dataset source? Shouldn't it be GLODAPv1.1? Here the reader might be confused. Please write specify it at page 23 line 15 instead of referring only to the sections.

Page 23: I would recommend using the work "re-processing" instead of re-analysis, which is usually used for model data.

Page 25, line 8: what is sigma-4?

Section 4

I suggest having only 2 sub-sections:

- 4.1 The Adjustment Table
- 4.2 The Secondary QC summary

and leave the text at pages 20-21 as introduction to sub-sections.

I would leave in **The Secondary QC summary section** only the results of the secondary QC! The salinity and oxygen calibration (page 27 lines 10-21, page 28 lines 9-19) might fit in Section 3, as already written above. Figure 5 and Table 10 should be introduced/explained here in the text and cited in the following sub-sections.

I suggest to put only the parameter names in the sub-sections (now inconsistent)

Page 32, Line 5: please insert the reference to Table 6.

Figure 6 and 7 are just cited in the text and they have very long captions. I would prefer to have an introduction/justification/explanation of the figure in the text and a short caption.

References in advance to sub-section 4.4 (page 28, 29, 30) suggests that its content might fit into the The Secondary QC summary section before the description by parameter together with Table 11 explanation.

Page 38, point 5: It is not clear to me, could you explain better? Are thos values excluded from product files or included and flagged 9?