Ocean Sci. Discuss., 9, C312–C315, 2012
www.ocean-sci-discuss.net/9/C312/2012/

Ocean Science
Discussions

Interactive comment on "The CORA dataset: validation and diagnostics of ocean temperature and salinity in situ measurements" *by* C. Cabanes et al.

Anonymous Referee #1

Received and published: 11 May 2012

Summary
The paper provides a description of how in situ data is collected and processed at the Coriolis Centre and then organised into the COriolis dataset for Re-Analysis (CORA). The quality control procedures to check the assimilated data and the diagnostics are essential information for researchers using such dataset and worthy of publication. However, the language and structure of the paper are not appropriate and the paper needs major revision and editing before it is suitable for publication.

General comments

The paper has huge language problems! Some should have been addressed before submission, e.g. obvious spelling errors in headings and legends. But even then a native English speaker should have edited the paper before submission.
On top of the language issues, the paper is not well structured. The introduction does not give an overview of the different datasets out there and why special datasets are needed. Instead, it sketches already what the CORA dataset is and how data are treated before assimilation.
The next section heading suggests that the CORA dataset evolved from the Coriolis data centre, which makes no sense and indeed, the section discusses the data collection and processing at the Coriolis data centre. The last subsection in here already talks about the data retrieval and organisation of the CORA dataset, which should fit to the next section.
The third section is about data processing and quality control to generate the CORA dataset. The section states that flags done by the Coriolis data centre are kept, but does not explain if they are used to any advantage during the following quality control procedures.
The next section is supposed to deal with diagnostics. However, the biggest part is again about quality control and should have been discussed in the previous section.
The text seems to be a conglomeration of 'copy-and-paste' fragments from different reports and it is hard for the reader to follow the data processing. The reader is confused at the end which quality control procedures were done at the Coriolis level and which were done when collecting data into the CORA dataset. A data flow diagram would help a lot.

Specific points
I will not go into type setting here and correcting the spelling and language problems as this would go beyond the duties of a reviewer. Instead I will highlight some specific points to support the general comments.
There is no consistency in using the acronym 'CORA' and 'CORA3' in the text. The

authors explain that CORA3 is the recent version of the CORA dataset, but they then do not stick with one acronym. I therefore assume that when 'CORA' is used, all versions of CORA are meant, while 'CORA3' means that some method or correction is specific to this latest version only? This needs further explanation within the text.
A lot of acronyms were used, but not explained (e.g. T, S, ATLAS, GODAE, PMEL, GT-SPP, WOA09, SST, SLA, XB, BA and TE, i.e. BATHY and TESAC); others are defined and then never used (e.g. GOSUD) and others defined twice (e.g GSSL).
Not sure about the OS policy about web links within the text. But my feeling is that most of the internet links occurring in the text could have been incorporated as references.
p1274 l23: Which decade?
p1277 l7: What is MyOceanII?
P1277 l17: What determines if the Coriolis centre collects in real-time or in near-real time mode?
P1278 l19: What do quality flags of 0 and 5-9 mean then?
P1278 l25: How large do the residuals have to be to produce an alert?
P1283 l12: I am confused here. The figure 4 shows the percentage of suspicious temperature and salinity profiles in GLORYS2V1 and not in the CORA dataset.
P1283 l16: The authors explain the peak in the number of suspicious salinity profiles around 2000. But what are the consequences? Are the profiles removed from the dataset as they are not conform with the climatology (bad idea) or are they kept within the data collection and flagged as 1 (false alarm)?
P1283 l23: Why do the authors expect the erroneous profiles to be randomly distributed in space? As you say, instruments with defective sensors would generate a trace. Very interesting is the increased number of suspicious salinity profiles at the northern end of the Antarctic Circumpolar Current. Could this be the result of a mismatch between the in-situ profile from e.g. within an eddy or meander to the climatology?
P1284 l9: Why doesn't the check at the Coriolis centre find all duplicates?
p1284 l23: Does it mean that 1.5P1284 l2: Why is the name CORIOLIS in capital

letters here?
P1288 l22: Figure 5 does not show the percentage of profiles with bad positions!
Figure 2: The text introduces the reader to quality flags ranging from 0-9 and explaines flags 1 to 4. However, this figures shows quality flags of 24. Please explain.
Figure 3: the red dots are hard to see and the font size is too small.
Figure 4: This figure is based on the GLORYS2V1 dataset, while the paper describes the CORA dataset. The legend should also explain why the profiles are classified as 'suspicious' or at least refer to the text.

Interactive comment on Ocean Sci. Discuss., 9, 1273, 2012.