Ocean Sci. Discuss., 9, C342–C347, 2012
www.ocean-sci-discuss.net/9/C342/2012/

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 #2

Received and published: 17 May 2012

The paper describes the production of the COriolis dataset for Re-Analysis (CORA) and also presents some statistics describing the data and its quality. The paper will be a very valuable resource for those wishing to use the dataset and I certainly recommend its publication. However, there is scope to improve the manuscript by reorganising it, expanding it in places and by clarifying the text.

First, my main general comment about the paper is that it could be made into a much improved resource for readers by reorganising it and changing its emphasis. The authors seem to have attempted to describe CORA in terms of additional things that are done to data extracted from the Coriolis database. However, I believe that users of

the dataset would prefer a more comprehensive description of the dataset. This would also improve the flow of the paper e.g. section 4 would then not need to discuss quality control (QC) checks done when data are ingested into the Coriolis database as these could be described briefly in an expanded section 3.

Second, there are a number of things in the paper that are stated without indication of the reasoning behind them. For example the spatial and temporal criteria that are used in the duplicate check and the thresholds used in quality control checks. I would like to see more indication of why these things were chosen.

Third, care should be taken to define all acronyms and initialisms at first use (e.g. CLIVAR, TAO, PMEL), and to better allow for the fact that readers will not have expertise in all the areas covered by the paper (e.g. what is the MyOceanII project, what is ETOP05, what is the Hanawa XBT correction and why should readers be concerned about whether it has been applied)?

Finally, it is indicated (page 1287, lines 20-22 and lines 24-25) that the CORA dataset is missing data before the early 2000s. Why were these gaps not filled by sourcing data from other places other than the Coriolis database? This appears to be a major negative to using the CORA dataset and some comment from the authors to address this concern would be welcome.

More specific comments are detailed below:

- The introduction tends to be too detailed and would benefit from being reworked e.g. much of the paragraph about Argo data should be kept for the later section. It would also benefit from more text that places CORA into context with the other ocean datasets. Why is it preferable for a user to make use of CORA in place of World Ocean Database, for example? Also, the first sentence does not make sense and needs rephrasing.

- Page 1275, line 18 - ’quick updates of the dataset required by reanalysis projects’ -

what is the requirement and does CORA fulfil this?

- Page 1276, line 5 - the list of references should include Gouretski and Koltermann, 2007, GRL 34, L01610 since this paper first highlighted the issue.

- Page 1278, line 16 - what are the QC checks that are done and what is the process of checking the data visually? This is important information that should be included in an expanded section 3.

- Page 1278, line 18 - what do all the numbers 0 to 9 mean (only 1 and 4 are defined here)?

- Page 1279, line 8 - are these pressure adjustments applied in the CORA data as this is not stated clearly?

- Page 1279, line 18 - why is the period 1990 - 2010 chosen?

- Page 1279, line 18 - if some of the data included in CORA3 were downloaded as long ago as 2010, won't this have missed some of the delayed mode updates to Argo data?

- Page 1280, line 4 - if the paper is to refer to the CORA documentation in this way, it might be useful for this to be attached to this paper as a supplementary file.

- Page 1280, line 10 - are any adjustments performed on MBT data? Are these included in the CORA dataset?

- Page 1280, line 25 - is this meant to say level rather than profile? This distinction needs to be made clearer throughout the manuscript - which tests reject a whole profile and which reject individual levels? Are there QC flags provided for the whole profiles?

- Page 1281, line 1 - what are the depth and region dependent thresholds?

- Page 1281, line 3 - why were these thresholds chosen?

- Page 1281, lines 4-6 - the discussion of this test is too brief and needs expanding. Why use the annual fields rather than e.g. monthly? Why is the threshold 10 times the

standard deviation? Is the climatology interpolated to the profile location and levels? Are the objectively analysed World Ocean Atlas fields used (the statistical mean fields might be more appropriate given the use of the standard deviation)?

- Page 1281, lines 7-15 - this is an interesting test that deserves more discussion. There are some issues with the current way it is described: the word bias is used, which implies that the offset calculated using this equation is an error. However, it may be real and I suggest using a different word. Why is it the standard deviation that is averaged to give the threshold (rather than average the variance) and why is 3 times the standard deviation used? Finally, I am slightly unconvinced about the example given in Figure 1. Were any checks done against other profiles in the area that confirms this profile to be wrong?

- Page 1282, first paragraph - this needs expanding. What are the previous systematic checks? How are the profiles verified?

- Page 1282, line 9 - the GLORYS renalysis only covers a limited time span of CORA3. Does this affect the quality of data outside this span? Could an alternative reanalysis be used that covers the whole period?

- Page 1282, line 18 - can the authors present any evidence to say that the assumption of a Gaussian distribution is justified?

- Page 1283, line 5 - how was this threshold decided upon?

- Page 1283, lines 20-22 - is anything done about the fact that too many observations are rejected during ENSO events?

- Page 1284, lines 4-5 - what happens to the profiles whose quality is difficult to evaluate (are they rejected or accepted)?

- Page 1284, line 12 - why is the window increased to 24 hours? Is any checking done on two profiles that are selected as duplicate to see if they are identical?

- Pages 1285-1286 - the section about XBT bias corrections should either be made more comprehensive (e.g. to give a non-expert reader an understanding of what the Hanawa correction is) or should be shortened to be a brief summary that refers back to Hamon et al. (2012) for the details of the method. Overall, I was left confused about what has been done about applying the XBT bias corrections. Were the Hamon et al. (2012) corrections recalculated for the CORA3 dataset and if so, why? Why were some of the details changed e.g. using the average temperature in the top 400m to separate profiles rather than 200m in Hamon et al. (2012) and what is the effect of making the different choice about applying the Hanawa corrections first?

- Page 1287, line 17 - what is an ATLAS buoy and what is a next generation atlas mooring?

- Page 1288, lines 8-14 - does this mean that an entire profile is rejected if 75% of levels are rejected, or is this being done only for the purpose of the figure?

- Page 1288, line 18 - ETOP05 needs to be defined and referenced. How is this dataset used (e.g. is it interpolated to the profile position)?

- Page 1288, line 27 - I don't understand what a theoretical position is, please define.

- Page 1289, lines 5-6 - what is meant by 'checking if the date and time are sensitive'? Please also define what the maximum allowed speeds are.

- Page 1289, lines 8-10 - please comment on why the percentage of bad salinity profiles changes after 2003.

- Page 1290, 1291 - these pages will be difficult to understand for readers not very familiar with the Argo project. Please rewrite with this in mind and give explanations about what the different types of floats are, what is a controller etc? On page 1291, line 15 what is the '27%' referring to?

- Page 1292, lines 1-3 - please provide a reference for this.

- Section 4.3 - I found this section rather confusing and it also needs to be made more robust. Please could it be rewritten to state more clearly what the comparison is that is being shown in Figure 12. For example, the figure has a time series marked as 'ARIVO' but this is never mentioned in the section (nor is ARIVO ever defined or a proper reference for it given). Also, since the section ends with pointing out that there are differences in trends and in the time series, it seems odd to have stated that the comparison shows good agreement and that this means that CORA does not miss too much bad data. I also don't understand what 'Differences of the 6-yr trends remain in the error bar estimation' means?

- Table 1 - please provide references for the Argo float numbers.

- Fig 1 - please state what the envelope is (10 standard deviations?)

- Fig 2 - what do the flag values mean?

- Figure 10 and section about Argo data - why the higher numbers of floats with position errors in 2004-06?

Technical points:

- Page 1275, line 22 - is e.g. meant rather than i.e.?

- Page 1278, line 6 - National Museum of Natural History needs to be defined better (e.g. by providing a web address).

- Page 1287, line 29 - could a website reference for NDBC be added?

- Page 1289, line 28 - please be consistent on the use of GDAC or global DAC.

- Figures - Unknown and gliders are spelled incorrectly in many of the figures.

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