Réponse anonymous reviewer #3:

We would like to thank the anonymous reviewer #3 for his/her wise comments that helped us to improve the manuscript.

Little was presented in this paper corroborating the quality control choices that were made. The presentation of a "Global ocean indicator" moved in this direction. But the attempt did not seem very robust, and in my opinion should be discussed in more detail in a future paper or expanded upon.

We have rewritten the part 'Global ocean indicator', in particular we have described more precisely what the comparison shown in figure 12 was and we have softened our conclusions. It is true that this part needs to be expanded upon but we prefer to hold it for a future paper.

Specific Comments:

Page 1276, line 11: It would be more appropriate to state "...models equipped with FSI sensors and built by Woods Hole Oceanographic Institution.....". The majority of WHOI built floats are deployed by WHOI, but not all.

We have changed it in the text.

Page 1278, line 16: While perhaps obvious from the text, the Wong et al citation does add some uncertainty. Does all types of data go through the same Argo program checks? If so it might be a help to state that the automatic QC procedure designed for Argo is applied to all data types (explicit).

Data downloaded onto the Coriolis database are going trough the same automatic checks as those designed for Argo floats (Argo Quality Control Manual, Wong et al., 2012). This is the case for all data types except for Argo floats handled by other DACs than Coriolis that have already qualified their own floats and for sea mammal data already qualified by CEBC- the Chizé Centre for Biological Studies).

However, some of the automatic tests are only specific to Argo floats (e.g. Digit Rollover Test) and are not applied to the other data types. Moreover, for the "Impossible speed test", thresholds defined for Argo floats are refined for other platform types.

We have expanded section 3 that now provides a description of the main automated tests. We have added a table that gives the complete list of the 18 tests, points out the tests specific to Argo floats and any differences compared to the description of the tests provided in the last version of the Argo Quality Control Manual, Wong et al., 2012.

Page 1278, paragraph with line 25: In this paragraph a number of netcdf format files are presented. It would aid the reader if it was stated whether these data types are only in-house types or are more commonly available from other data depositories.

We have rewritten this paragraph:

The data are classified in 7 types depending mainly on the data sources and resolution. Most of these types correspond to the data types defined for the GTSPP program (PF, CT, XB, BA, TE) others are 'in-house' types (OC and MO). All the data are stored in netcdf files with the same format (the same as the one defined for the Argo program) and a naming convention that indicates the data type.

Page 1286, line 2 : The QC flags used in the Argo program have been adopted, and at some point in the text a citation towards Argo documentation is given. At one point in the text, the flags 1 and 4 are defined, however elsewhere in the text, flags 2 (page 1291, line 4) and 3 (page 1286, line 2) are mentioned without definition. It would be helpful to either shortly define these extra flags, or even better add a small table of common flags.

We have added a table with the complete definition of flags.

Page 1286, lines 4-15 : The description of XBT corrections is the only section where I felt not enough information was given to the reader. First is there enough reference profile data within 2 degree lat/lon and 15 days of a XBT? If there isn't what is done? Could "error of immersion" be described? The consideration of bathymetry makes sense...why not extend that to areas of strong geostrophic currents?

As suggested by the reviewer #2 we have chosen to shorten the description of the method because it is the same as the one described in Hamon et al. (2012). However, we have better explained why we have not used the coefficients given in table 2 of Hamon et al, 2012 but instead, re-computed them (it was not clear at all in the first version of the text).

One of the main reasons was that in Hamon et al 2012, the corrections were computed from the comparison of collocated XBT-CTD pairs. But in the 2000s their coefficients were based on much less collocated pairs than for the years before (see their Figure 5). We have then decided to not only use CTD profiles as in Hamon et al. 2012 but also Argo, drifting buoys, and mooring buoys data (only those with quality flags 1 or 2) to get more reference profiles co-localized with XBTs. We then obtain between 6000 and 16000 collocated pairs each year between 2002 and 2010 which is much more than if only CTDs were used as reference profiles. We have considered that it was enough for our statistics.

As in Hamon et al, 2012, we have considered the bathymetry to avoid potential biases resulting from cross-shelf fronts. It is true that we could have extended that to strong geostrophic currents. This can be an improvement for further versions.

Page 1288, line 7-8 : Towards the beginning of the document (page 1278 line 20-21) it was stated that outside QC flags are kept. But here it says that flags are set by Coriolis and CORA. In general, the paper seems unclear on whether outside flags were used, or were ignored. Another instance...page 1289 line 22...” these data have been requalified....” This could mean that all flags were reexamined from previous values or flags were created from scratch. The difference between re-certifying flags and starting from scratch seems to me to be an important.

Once data are downloaded onto the Coriolis database, they are going through real time automatic tests (except for Argo floats handled by other DACs than Coriolis and for sea mammal data already qualified by CEBC. In this case originator flags are kept).These automatic checks set quality flags for each measurement (i.e. for pressure/depth temperature and salinity at each observed level of a profile) as well as for position and date of the profile.

Once the CORA3 files are extracted from the Coriolis database, data are going through delayed time validation procedures. Only data considered as good or probably good (flag 1 or 2) after real and near-real time tests or data that have never been checked (flag 0) are further verified in delayed mode (note however that the CORA3 dataset contains all the data, even those qualified as bad during the previous real time tests)

Page 1290, line 4 : Is an Argo profile bad if only a single level is bad? It is unclear.

Quality flags are set for each measurement (i.e. for pressure/depth, temperature or salinity at each observed level of a profile) as well as for the position and the date of the profile. For the purpose of the plot we have considered that the temperature or salinity profile was bad if at least 75% of the temperature or salinity measurements at observed levels were bad. We have made it clearer in the text.

Page 1290, line 8 : I think it would be more accurate to say “mainly WHOI SOLO floats with FSI CTD sensors “

Yes, we have made the correction.

Page 1290, line 10 : In the aggregate, the WHOI FSI floats resulted in a cold bias, however individual profiles, or regions of profiles from float sub-types could have been warm. It is an important distinction.

This has been clarified in the text.

Page 1291, lines 8-12 : It is stated that “13% of TNPD are flagged as bad” This is quite a low number. However, the reader is unsure if this is all the profiles that should be flagged bad by Argo guidelines. Without giving some idea of how many should be flagged bad, this statistic is not very meaningful.

In delayed mode, the float PIs are asked to flag the data (TEMP, PRES and PSAL) of TNPD floats as bad (flag 4) when float data show observable T/S anomalies that are consistent with

increasingly negative pressure drift and to flag the data of TNPD floats as probably good (flag 2) otherwise. Thus, all the TNPD floats are not necessarily flagged as bad by the PIs.
It is difficult to say how many should be flag bad. However, only severe negative pressure drifts show observable T/S anomalies (an error of −20 dbar will cause a positive salinity error of approximately 0.01 PSS-78). Barker et al (2011) estimated a median error of -3 dbar for all TNPD profiles that can be compared with a close good profile. As a consequence they recommended that all the TNPD floats be excluded from studies of oceanic heat content and decadal changes. In the CORA3 dataset 13% of TNPD floats are flagged as 4, meaning that probably only the most severe negative drifts have been caught.

Page 1291, line 16-18 : Is it Argo policy to delay-mode quality control the salinity profile while not quality controlling pressure? If so this needs to be brought to the readers attention.

Delayed-mode pressure adjustment procedure for APEX floats has been added in the Argo quality control manual in Fev. 2009, once a problem with the Druck pressure sensor was found. During the year 2009, PIs started to apply these pressure adjustments in delayed mode before computing the salinity adjustments. However, delayed mode processing is a long term task and some floats have not been reprocessed by the PIs yet.

Page 1292, line 3 : To what does (>50%) refer?

We have changed the text that was unclear: "Among them, about 27% are not corrected and 23% have a correction equal to zero."

Page 1292, lines 28-29 : The authors give one possible interpretation of the near agreement in figure 12. I don't think it is the only reason. Since no further evidence of the other variables involved, perhaps this statement should be softened.

Yes, this is true and we have softened this statement. One of the possible reasons is that the method used to compute the GSSL is robust and not very sensitive to some remaining bad data. Another reason is the possible compensatory effect of some residual positive and negative biases in our GSSL estimates.

Page 1293, lines 1 : How is the error bars in figure 12 determined?

Error bars are calculated as described in Von Schuckman and Le Traon, (2011).

$$E^2_{total}(t) = E^2(t) + E^2_{c\lim}(t)$$

This total error includes the uncertainties on the averaged parameter in every 5°x10°x3months box and the choice of the reference climatology, but it does not take into account possible unknown systematic measurement errors.

Page 1293, line 11 : reads "Any validation system is perfect..." Is this as intended?

No, it was an error. We have replaced it by “No validation system is perfect….”

Page 1293, line 17 : reads “...to do not flag a profile as bad if we had some doubts.” This is unclear what the meaning is.

Our general approach was not to flag a profile as bad if we had some doubts, meaning that if the visual checked performed on the profile (comparison to climatology and neighbouring profiles) were not sufficient to decide if the profile was good or not, we let the flags unchanged.

Figure 1 : It would be helpful to define the ’envelope’ used in the caption.

The envelope is 10 standard deviations; this has been added to the legend.

Figure 10 : The 2 colors for APEX and (with TNPD) are very close, perhaps different colors could be chosen.

We have changed this figure according to your comment.