

The authors thank the reviewers for their comments.

From now on, the reviewer's comments will be in black and the authors' responses in blue. The lines number refer to the document "Author's tracked changes".

General comments

My major concern is still the applicability of the filters proposed by the authors to different study areas. The algorithm bases on the fact that salinity is constant below 1500 m and so, the salinity drift can be computed and corrected. While this might be true for certain regions it is not true for all of them and the authors need to be clear on the fact that the user should know if the salinity varies below that level within the desired study area.

No, the algorithm is not based on the fact that salinity is constant below 1500 m. It is based on the fact that there are depths where there is little dispersion in the salinity data where we can apply group analysis to filter the data. In the study area called TPCM, which was previously studied before using the filter, it was found that at 1500 m there was this little dispersion of the data and therefore this depth was used, however in lines 104 (now 101), 199 and 259 it is specified that the depth is by default, and in lines 188-191 and 258-260 it is mentioned that it is the responsibility of the users to make decisions when using this method based on their knowledge of the study area. We added in lines 190 and 191 that you must also decide whether to use the default value of the depth or not. And we rewrote lines 99-105 to try to be more explicit when mentioning that the 1500 m depth is by default and the user can modify it.

We also added five examples of study areas with different hydrographic characteristics, extensions and locations (lines 113-119, 194-210), in which the default value worked quite well except for the study area named Alboran Sean, where it was suggested to use shallower depths.

Giving an option to the user to choose the reference level when using the filter, if possible, would be great, even if the default can still be the 1500 m.

Sure, that's the idea mentioned on lines 104 (now 101), 199 and 259.

I still think that some figures are not very useful. Particularly the ones that show only region delimitation like 2 and 8. Keeping only one of them would be a better option in my opinion.

Suggestion accepted, we removed Figure 2.

Particular comments

Lines 4-5: The authors say in the abstract "This work shows a methodology to filter profiles within a given polygon using the odd-even algorithm, this allows analysis of a study area, regardless of size, shape or location". As I said, this is not precise, here the authors should make clear that the user need to ensure that in the study area salinity doesn't change below 1500 (or let the user to choose the reference salinity)

We rewrote lines 99-105 to try to be more explicit when mentioning that the 1500 m depth is by default and the user can modify it.

In my previous review I made this comment: “I find that the aim of this study is interesting and the tool the authors provide is useful. However, it presents one main problem that is the applicability of the tool in any polygon of any area chosen by the user. I don’t think that the tool can be used globally and so, its limits need to be specified by the authors.”

And the authors responded: “Suggestion accepted (lines 208-210)”

...But I don’t see the answer to my concern in these lines. Is this a mistake in the line numbering?

We add a response and a personalized document for each of the comments of the preprint, these lines refer to the document that we sent you as a response. We did not know that the reviewers were only going to check the final document, we apologize. These lines are now 188-191.

Line 6: sentence too long and unclear. Please rewrite

Suggestion accepted (lines 4-6).

110-115: And where can this analysis be seen? The authors talk about this later in the paper and I think that they should focus in the later occurrence of this topic and delete it from here (or just mention it quickly). By reading this paragraph, the reader wonders what I asked at the beginning of this comment, which I made before reading the whole paper.

You can see it in the results, here in the methodology we are briefly describing the study areas and what was done.

117: This paragraph fits better in the section “web application”

The "Web application" section is part of the results, not the methodology. We decided that a section to separate three paragraphs in the methodology is not necessary.

Lines 148-150: The authors said: “Taking into consideration that at depths greater than 1500 m, the variations in salinity and temperatures are imperceptible, the cluster analysis was performed with the salinity data measured at depths greater than 1500 m”

Again, this is not true. I had asked the authors to show this is true or otherwise specify very clearly in the paper that the selected region must be checked by the user in order to ensure that below 1500 m the salinity differences within a given region are not important enough, and that the algorithm can be applied.

We specified very clearly in the document that the user must know their study area on lines 188-191 and 257-260. However, just before this paragraph we are describing our study area, which we take as an example. We believed that the sequence of the writing would suggest that we are talking about this study area and its characteristics, we already saw that it is not. We modified this sentence to specify that we are talking about the study area (lines 153-154).

Figure 3: The difference between the two yellow and the two purple colors is imperceptible. Please choose more different tones of color or two different colors. Also, Argo data seem deeper (colder) than WOA, which is unlikely since Argo only reaches 2000 m and WOA data contains deeper data, so I think that the legend is the other way around (yellow for Argo and purple for WOA), please check this.

Suggestion accepted; we changed the colors of the data at depths greater than 1500 m. On the other hand, the data has been rechecked and the legend in the figure is correct. In this study area the WOA18 data reaches a depth of 1500 m.

157: “To manually avoid indicating the number of k centroids” change for “To avoid indicating the number of k centroids manually”

Suggestion accepted (line 162).

180: I suggest to avoid referring to the user as “the researcher” since many Argo data user are not researchers. “ the user” would be maybe a better option.

Suggestion accepted (lines 9, 178, 184-186, 238, 250, 253, 255-256, 258 and 279).

205: I still think that too much attention is being paid to the PIP which is a very simple task. The number of vertices is really irrelevant when choosing data inside a closed contour in Matlab or Python. However, I agree that being able to select an irregular polygon is useful for the web interface.

Suggestion accepted (lines 213-214).

The authors thank the reviewers for their comments.

From now on, the reviewer's comments will be in black and the authors' responses in blue. The lines number refer to the document "Author's tracked changes".

General comments

I have added the suggestions for revision in the pdf.

We copied and pasted the pdf comments to this document for easy review.

The major concern are the data in Figure 4 and if only data with qc=1 are shown

We rewrote the sentence that describes the methodology (lines 79-82).

Particular comments

Lines 23-24: the R-files in the Argo programme just don't have an expert quality control performed on them yet. It can not be said in general that they are not of such good quality

Suggestion accepted (line 24).

Lines 47-48: 1) the sentence needs revision. Which 'data access platforms' are you referring to?

2) And are suggesting that these platforms offer the user your tool?

1) We modified the sentence to be more specific with the type of platform (line 43). The list is seen in the discussion, at this point it is not relevant to mention them.

2) Yes, we are suggesting it. That is why we wrote about the usefulness it can have if it were integrated.

Line 68: a current is not transported

Suggestion accepted (lines 52-55).

Line 69: this word can not be correct, please check

Suggestion accepted (lines 52-55).

Lines 95-96: I would suggest to rephrase:

...during the RTQC only gross errors can be identified by the automated procedures and to detect more subtle issues with sensor drift can only be carried out in delayed mode-

Suggestion accepted (lines 75-79).

Line 98: Is it really true that the TS diagram in Fig. 4 shows only RT data with QC 1?

I doubt this. The supplementary material provided shows a similar figure color coded by qc code. And there it appears as if most of the erroneous data have real time qc flags of 3 and 4.

Please confirm.

No, it is not true, this is a verb conjugation error. We rewrote the sentence (lines 79-82).

Lines 119-120: I don't understand why the depth of the profiles is important in the algorithm

The clustering technique used in this work is based on the distance that exists between the data (lines 88-94), using the depth to delimit which data to use with this technique. This allows the use of a subset of data with less dispersion than if the full data set were used (lines 98-112).

Lines 122-124: what do you mean by kept the group where the dmqc data were found in regard to filtering the RTQC data?

The DMQC data is the data that concluded the Argo quality control process and therefore it is good data, for this reason the algorithm always preserves the group that contains DMQC data. It never filters DMQC data, it only uses it to group RTQC data that has patterns similar to the DMQC (lines 93-105 and 153-158).

Lines 129-135: does the reader has access to this data. Should you refer to the supplement?

The reader has access to the data through the Argo snapshot from December 2020. We now added the supplementary material for reference so that readers can see the figures and results (<https://doi.org/10.6084/m9.figshare.14999613.v1>). The reference was added to line 196.

Line 139: What is meant here by Ifremer synchronization service? Can you provide a web link or web page address?

Suggestion accepted (line 124).

Table 2 caption: Can the user see the corresponding figures from the supplement somewhere? The table makes no sense in the manuscript if you don't see corresponding plots-

We now added the supplementary material for reference so that readers can see the figures and results (<https://doi.org/10.6084/m9.figshare.14999613.v1>). The reference was added to line 196.

Line 225: Why should sensor drift be only contained in specific month? If a sensor on a float goes bad it will drift continuously.

Same question for the other study areas.

The data are evaluated separated by month for the seasonality they have and thus avoid combining data from hot months with cold months. For this reason, the algorithm filters and shows data with salinity drifts per month, it is true that the drifts are continuous, but the months mentioned here are the months in which the algorithm failed to completely filter the data. In this case, as it is written in the text, it is recommended to use the second filter (lines 199-201).