Authors thank the reviewers for their comments.

Hereafter, the reviewer's comments are in black and the authors' answers in blue.

General Comments

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.

Suggestion accepted (lines 208-210).

The cluster analysis is carried out within a region that can be defined by the user. So, if this region is large, how can we suppose that the profiles should be similar to each other within the region even below 1500 m? How did the authors chose this criterion? The choice of 1500 m need to be justified, why not deeper? It could be spatial variability that is not easy to distinguish from the salinity drift. I believe that the authors need to prove that below 1500 salinity does not vary.

The 1500 m depth was chosen based on the knowledge of the study area and the analysis of the TS diagrams (Figure 3). The considerations that the user must take into account when using the algorithm has been added (lines 208-210). In addition, the paragraphs that describe the algorithm were rewritten (lines 108-122).

Also, a more detailed explanation of this cluster analysis is needed. The centroid of the groups is considered in salinity? Or in both temperature and salinity ? The cluster analysis bases on iterations to approximate the centroids in data space (what is data space? T-S space?) to their closest centroid. An schematic of the functioning of the algorithm would be very useful for the readers.

An extra sentence was added to the description of K-means (line 100) and the paragraphs that comprise lines 105 to 122 were rewritten hoping to make the explanation clearer.

I suggest to add more information about the processes of this real time quality control as performed by Argo. In this way, the readers can realize in what the proposed technique differs from the one already existent that discards the profiles with salinity drift. Note that the Argo Quality Control is a very detailed process, so it is not easy to justify and propose an alternative method that inspires the confidence of the users. The description and justification of the proposed alternative must be very detailed. I understand that the authors are not saying that

their method is an alternative better than Argo, rather, they propose a solution to have available more data in short time only in the case that they were discarded at first instance by Argo due to a salinity drift. Still, this point should be very clear to the readers.

A more detailed explanation of the quality controls was added in lines 26-39, 84-90 and 255-256.

Figures seem inadequate to me, the first 3 figures could be easily summarized in a single figure, and they are not well referenced in the text. Moreover, some of the figures are explained in a too exhaustive way in my opinion (i.e. figure 3) while others that contain more substantial information are too briefly described (fig. 7)

Suggestion accepted. Figure 3 was removed along with its description (lines 146-155).

The second filter needs further explanation.

A better explanation of what the second filter does was added (lines 120-122). The procedure of the second filter is not very complicated (unlike the first) we hope that this change is enough to clarify it.

How do the authors deal with the data that on top of showing salinity drift, show any other problems? Do they discard them or not?

The aim of this work is to attack salinity drifts that can be detected through the depth at which they occur (more specifically, greater than 1500 m). Although there are other problems in the RTQC, we are not dealing with them in this work.

Specific Comments

Abstract

"In the study area selected as an example, it was possible to recover around 80 % in the case of the first filter and 30 % in the case of the second of the total real time quality control data that are usually discarded due to problems such as salinity drifts"

The paragraph was changed to clarify that users are the ones who normally discard RTQC data due to data problems (line 9).

This sentence is not clear, (a) what is the first and second filter? (b) the Argo quality control is not only based on salinity drift, so can you explain the method a bit more?

(a) A brief description of what each filter does has been added in lines 7 and 8.

(b) We know that there are other problems in the RTQC data, but the aim of this work is to attack salinity drifts.

Introduction:

Line 20-22: It would be useful to say here what percentage of total data are flagged good in average instead of saying that in places with low concentration of profiles the good quality data are (not 'is' as it's written in the text) scarce.

Suggestion accepted (lines 22-24).

Data collection and methods:

Line 47-49: It seems to me that, if you set the range of latitudes and longitudes as described you'll end up with a regular polygon, what happens if the selected polygon by the user is irregular?

Whether the polygon is irregular or not, the procedure of establishing ranges is maintained in order to evaluate fewer profiles with the PIP algorithm.

If that's not the case please explain this paragraph better

This paragraph was shortened according to the reviewers' suggestions (lines 49-64).

Technical corrections

47: change "them" by "the data" for example. The reader won't know to what "them" makes reference.

Correction accepted.

52: The characteristics of the region are relevant for the study presented here? If so please say in which way.

We gave the description of the area to give weight to the amount of data admitted, to mention the water masses and hydrographically justify where we are putting the centroids. It is necessary for the user to know the study area so that he can make a difference between which data have salinity drifts and which do not, to decide which filter to use or whether to use a combination of both (this clarification was added in lines 208-210).

65: What are the corrections of the data according to TEOS-10?. Actually I do not understand well this whole paragraph. The authors used monthly means of T and S from WOA and then the PIP algorithm was used to separate these data in those that are inside and outside of the

polygon, but what happened next? In what consist the comparison between these WOA data and those from DMQC from Argo? Please explain

The corrections applied according to TEOS-10 were specified and an extended description of the comparison process was added to the paragraph (lines 78-82).

60-65: Also, This paragraph describes the statistical mean of each quarter degree (¼) in WOA. It seems that this is shown in Figure 3, why it is not referenced here?

Because in this paragraph we are describing the process of downloading and processing the data, we are not talking about results. In lines 158-163 is the explanation of the results and the reference to the figure.

69: What does “around 30% of the data are part of the RTQC” mean? That are these data flagged good by the RTQC?

30% of the total data. It was corrected (lines 84 and 85).

71: “one being the best and the fourth being the worst. Tests were performed by graphing the TS diagrams using these flags” please review the writing.

Suggestion accepted (lines 86-90).

71: These tests are not related with Figure 5?, if so, please reference it.

Yes, it is related because we are describing the methodology that we follow, but not the results. The figure (now Figure 4) does not show a methodology, it shows the results. For this reason, the reference is made to the figure in the results (lines 167 and 168).

80: The centroids are computed in the T-S space?

This paragraph describes how the k-means algorithm works, we are not talking about TS diagrams here. "The data space" refers to the data space of any problem and not the one we present in this work. A sentence was added to this paragraph (line 100) in order to be clearer in the general description of the algorithm.

88: “it calculates the mid-ranges of each quality control”… The mid ranges of what variable? Please describe what mid-ranges is.

Variable specification added in the line 110.

(a) In line 85 The authors say that Algorithm 1 is used to automate the process of the enumeration of the groups to be searched for the k centroids. However then, in line 93 the

authors claim that “The procedure described above is the first filter of the RTQC data”. (b) I did not understand well how the explained algorithm is a filter and how it relates with the stated from line 85. I would suggests a better explanation of this process since it seems to be key for the paper.

(a) A more detailed explanation of the purpose of automation was added (lines 106 and 107).

(b) A non-technical summary was added to clarify the filtering process (lines 116-122).

93-94: The second filter need further explanation (already said in general comments)

A better explanation of what the second filter does was added (lines 120-122). The procedure of the second filter is not very complicated (unlike the first), we hope that this change is enough to clarify it.

100: “In Figure 2, the blue line delimits the EEZ of Mexico and the yellow box delimits the TPCM.” it is the opposite way

Mistake corrected (lines 133 and 134).

115: In my opinion the authors describe too in detail the processes of selecting the data that are inside the polygon. It is not a complicated task to accomplish and I don’t think that it deserves a whole figure with two panels to show the same thing.

Redundancy was removed from the paragraph that was already explained in the methodology (lines 146-155).

120: the DMQC in Argo and the same DMQC in WOA18? Please specify, since the WOA data also have quality flags

This was already explained as part of the methodology during the correction of the comment on line 65.

Figure 4 needs a legend indicating what the two colors are (and same for Fig 5 and 7)

Figure 4 already had the legend (like 5 and 7) indicating each color in the upper left panel. The number of figures has been changed and the legend was changed to the bottom of Figures 3 and 4 to make it visually easier to see.

121-122: That seems to be true, but in the figure we cannot see where the 1500 m depth limit is

Colors of Figures 3 and 4 were modified. Dark purple and dark yellow for RTQC and DMQC data greater than 1500 m respectively, light purple and light yellow for RTQC and DMQC data less than 1500 m.

125: So...data that have been qualified as good by Argo in their RTQC present salinity drift? Why were they qualified as good then? Probably the Argo system knows that they can be corrected? That's why I recommend to include in this paper some more information of the more relevant choices of the Argo QC.

The explanation of why they were rated good was added on lines 86-90. Also, more information about Argo was added on lines 36-39.

128: And here the authors say that these data with drift are labeled as erroneous...I don't understand this contradiction with previous lines. Is it a mistake or am I confused?

It's a mistake. The extra sentence was removed (line 167).

134: The authors need to show that this is true and where it is true. This is a major shortcoming of the study

The methodology described in this work was carried out in five different study areas with differences in extension, profile density and hydrographic characteristics, with the aim of seeing the results of this method in other parts of the world (lines 123-129, 213-229 and 277-279). The results are now shown in Table 2 and the generated graphs have been added in "Supplementary Material #1" at the end of this document.

144: one of the blue profiles shows evident salinity drift, why is that? Argo error? Authors mistake in plotting the profiles with different colors?

Figure 6 graphically shows how the proposed algorithm separates the data into two groups in each iteration. The first iteration contains data with salinity drift, this is correct, but for this reason the algorithm did not finish its execution in this iteration, until the third iteration (Fig. 6c).

145: discarded by who? By Argo RTQC or but the authors of this study? This figure needs more explanation

Specification added on line 185. A more extended explanation was added in the methodology about this procedure (lines 116-122) and a more extended explanation of the figure on lines 182-187.

146: “both groups contain data in DMQC ”? Is this true? Or mainly both groups contain data that match those of DMQC? I’m not sure I’m understanding. And under which criterion is the matching defined? Also, shouldn’t panel c plots be in blue color for Consistency?

A better explanation was added to the paragraph to resolve these doubts (lines 185-187).

151-153: The results of the second filter seem quite good and promising. However, I insist that this second filter is not explained enough. Please provide a more detailed explanation in the methods section.

The explanation of the second filter was rewritten in lines 120-122. Also, a sentence was added to the paragraph (line 193).

Table 1: What does “meas” mean?

The clarification was added in the paragraph where the table is referenced (line 195).

164: “the researcher may simply not use the data from those months”. I strongly advise to delete this sentence for two reasons: (I) the problem is probably not with these months but with the data and if we change the region, the wrong data would be in different months. (ii) it can be not easy for researchers to go and look for the data that are wrong and decide if the are wrong enough to discard them. The I’d propose to the researchers to only use the data from the second filter (and using data from the first one would be on their on risk).

We cannot delete this paragraph because it talks about the feasibility of using filter one or filter two, or a combination of both if necessary. The paragraph was extended with a clarification that the results depend on the extension and the hydrographic characteristics of the study area (lines 208-210). We hope it is clearer.

171: I would start a new subsection here, something like: “Web application”

Suggestion accepted (line 230).

171-173: This figure and result is very similar to those in fig 3 and which I have already advised to reduce. Now I insist that the authors could joint together Fig 9 and 3 and summarize the description. Choosing data that belong to a given polygon (even if it has a complicated shape) is a very simple task in my opinion, and it doesn’t deserve that much of attention. The most interesting subject of the paper is the filtering procedure that could gain more attention and more space for its description on the paper.

Suggestion accepted. Figure 3 was removed along with its description (lines 146-155).

189-191: The authors talk a lot about the example in the ETP off Mexico, but, at which degree is their method applicable to larger or different polygons?

We add the results of the method applied to five extra study areas with different characteristics, in lines 216-229, Table 2 and in the "Supplementary Material #1" at the end of this document.

216-219: "The current platforms already provide graphics and data from the profilers, as well as filters to display or download the data, however, the geographical filter they use is by maximum and minimum coordinates, so it is only possible to filter by polygons in rectangle or square shape without rotation" I see now the interest on showing that with the tool provided by the authors users can choose irregular polygons. This advantage in comparison with other platforms is great, and it should be mentioned earlier in the text. However, I still thing that it can be said in one or two sentences and that too much detail on this is included in the text before (in the discussion is fine).

Now, this is mentioned at the end of the introduction (lines 45-47).

219: define JCOMMOPS (and change analyzes for analyses)

Suggestion and correction accepted (lines 284 and 285).

230: This sentence is not a conclusion of this study, it should be removed. This is something between Argo and WOA.

Suggestion accepted (lines 295 and 296).

233: 80% regarding what? Earlier in the text, the authors said that the data recovered in comparison with the DMQC of Argo were 30% and 10% respectively for the first and second filter. I recommend to define the criterion for the recovering percentage, either regarding the total amount of data or regarding the data that are discarded by the Argo DMQC.

The sentence was rewritten to be clearer (lines 298 and 2999).

Technical comments:

- data is plural, please correct the concordance with the verbal tenses throughout the Manuscript

It was revised and corrected throughout the Manuscript.

- 113: New sentence after "worked correctly"

This paragraph was shortened according to the reviewers' suggestions (lines 146-155).

113-115: “in addition to establishing the range of maximums and minimums of the latitude and longitude of the polygon to discard the profiles measured outside it, allowed the PIP algorithm to filter only the profiles made near or inside the polygon”. This sentence is oddly written and seems kind of obvious.

This paragraph was shortened according to the reviewers' suggestions (lines 146-155).

- 194-195: “since these processes are automatic and search for data that is impossible or outside the global and regional ranges” Please rewrite

Suggestion accepted (lines 255 and 256).

Authors thank the reviewers for their comments.

Hereafter, the reviewer's comments are in black and the authors' answers in blue.

General comment

I find it a bit difficult to understand the aim of study and that is already evident in the abstract, which principally should give the reader a clear understanding of the research questions addressed and results obtain. Major parts of section 2 and 3 focus on the selection of data for the cluster analysis and present an algorithm to obtain these through spatial polygons. The authors seem to imply that their PIP needs introduction and is something new. However, there are build-in functions in libraries such as MATLAB which provide users with exactly that functionality. It would be much more important to address the cluster analysis that is performed on the selected data and what needs to be done so it can optimally work. Their rather ad-hoc choices of polygon in figure 2 needs better interpretation. This polygon for the EEZ of Mexico covers the Gulf of Mexico with Atlantic waters and part of the Pacific. Why would one want to perform a cluster analysis that is supposed to ensure that real-time Argo data show same hydrographic relations as delayed-mode Argo in such a polygon? The data collected in the polygon present the cluster algorithm with two major hydrographically different areas and the hydrographic differences between Atlantic and Pacific are so much greater than the salty drift in the Argo CTD cells.

That is true, PIP algorithms are not new, however these tools are not currently used by Argo data access platforms, which would be beneficial to all users. In accordance with the reviewers' suggestions, we reduced the description of the PIP algorithm (lines 50-65 and 152-164), however we believe that it is a good proposal for current platforms to provide even more information than they already provide. The use of the EEZ of Mexico is a way of exemplifying how this same PIP algorithm can be used so that users of the data access platforms obtain statistical data of the study area of interest, at no time is it mentioned that the data is mixed or used to run the proposed algorithm. On the other hand, the manuscript mentions that the proposed algorithm was integrated into the web application (lines 148-150) and that the user can select a study area within the EEZ of Mexico to discard data with salinity drifts. If the users who are using the proposed algorithm decide to use it in very large areas with different hydrographic characteristics, they are not making good use of the algorithm.

The other area in which the paper needs major revisions is the description of the Argo data. The terminology used here is often too vague (sometimes also wrong). Please take more care to explain to the reader the structure of the Argo data set, the doubled data structures in the files (ADJUSTED versus original data). I am also not sure what the authors have selected as RTQC and DMQC data. Is it R-files versus D-files? Which quality flags were selected for both.

And what DATA_MODE has been selected? In case DATA_MODE is A, did they select the raw data or the *_ADJUSTED?

A more detailed explanation of the quality controls was added in lines 27-40, 91-96 and 261-262. The structure of the Argo dataset (ADJUSTED versus original data) is now explained on lines 31-35). And the data that is used as input for the proposed algorithm is now specified in lines 114-115.

The Argo data management invests a huge amount of effort in the data quality control and since this is time consuming any advances in more automated drift detection would be welcome. But these methods have to well described and tested. Considering the variability in the ocean and the small drift signals from deterioration of the conductivity cells, these are hard to distinguish from the background noise. The examples shown here deal with really huge offsets/jumps in salinity and are easy to detect. I would have assumed the real-time quality tests would have flagged these data already as bad and am wondering if the authors have considered the quality flags for the real time data properly.

We do not detract from Argo or its quality controls and we also believe that they do a great job, however, we are proposing an algorithm that can be used by users who wish to use RTQC data that do not have salinity drift problems and that at the same time have similar patterns to the DMQC data. We add the results of the method applied to five extra study areas with different hydrographic characteristics and differences in extension and location, in lines 219-235, Table 2 and at the end of this document in " Supplementary Material #1". Finally, it is true that real-time quality control tests can mark salinity drifts as bad, but it is also possible that they do not, as shown in "Supplementary Material #2" at the end of this document.

It seems to me as if the manuscript is in an too early stage and thus the scientific results and conclusions are not yet presented in a clear, concise, and well-structured way. The use of English language could be improved.

The manuscript has been substantially improved and is now in a better version. English was also revised and improved.

Specific comments are given directly in the pdf version of the manuscript.

PDF version comments

1-10: The abstract is confusing. needs a better definition of the aim of the study, explain the analyis conducted in a more precise way.

In this version of the manuscript the abstract was modified and improved. Now the aim of the work is explicitly mentioned (line 5) and a reference to the tests of the methodology in other areas of study has been added (lines 11 and 12).

13: Never heard that abbreviation before.

It is an abbreviation formed by the words Hydrographic Autonomous Profilers (HAP), to refer to them in the rest of the writing.

14-16: I am not sure what is the purpose of this sentence.

The sentence above provides a brief introduction to profilers (lines 14-16). This sentence links these profilers to the Argo program and further indicates to the reader that the data always undergoes quality control before being published (lines 17 and 18).

18: This is a bit confusing, please make clearer that this is the number of floats deployed since the start of the program.

Suggestion accepted (lines 19 and 20).

20-21: That is a strange assumption. The ratio of good to bad data is not necessarily dependent on the amount of profiles in an area.

We rewrite this sentence to be clearer (lines 23-26).

23: What is the difference between validated and verified.

To validate is to give strength or firmness to something and to verify is to examine the truth of something. It is true that the concepts are similar, we remove "verify" to avoid confusion (line 27).

24: This means the argo quality control system.

Suggestion accepted (line 28).

25: No the goal of the RTQC is to esure that data in the real time stream do contain quality flags indicating the quality of the measured data.

Yes, we changed "goal" to "requirement" and rewrite the sentence to be clearer (lines 29-31).

26: Please explain what you consider serious errors.

This is not about what we considered, it is about what is specified in the manuals. We rewrote the sentence to be clearer and more specific (lines 31-33).

27: consistency with the hydrography of the area is not really a check performed. There are broad global and local range checks that are available.

Correction accepted (line 33).

28: No that is absolutely not the case. Argo does not replace the RT data. In the delayed mode process the real time data are preserved and any correction is copied into an additional variable.

We changed some words to be more specific (lines 35 and 37).

29-31: Please rewrite this sentence it is not clear. I assume you wanted to say, that Argo data users are adviced only to use delayed mode data for scientific analysis or check data quality of real time data individually.

Yes, suggestion accepted (lines 37-40).

34-35: This sentence is unclear.

The sentence was rewritten (lines 44 and 45).

35: What do you mean by point in polygon algorithm? I assume you want to detemine if a certain profile position falls inside a specified polygon or not.

Yes, a point in polygon problem asks if a given point is inside or outside of a polygon.

37: Please rephrase that does not sound right.

The sentence was rewritten (lines 46-48).

39-41: The sentence is confusing, I am not sure what you want to say and specifically what you mean by 'filter the data'.

This paragraph was shortened to make it easier to understand (lines 50-65).

47-48: This sentence is unclear. I am also not sure that the 'problem' of determining if data fall into a certain area (polygon) need any detailed description. Programs such as MATLAB offer build-in-functions for this purpose.

This paragraph was shortened according to the reviewers' suggestions (lines 50-65).

49-50: I have no idea what you mean by this.

This comment is not clear but, this paragraph was shortened according to the reviewers' suggestions (lines 50-65).

53-54: Please be more specific. which currents are you talking about? And what does high complexity mean?

This paragraph was modified to be more specific (lines 66-74).

58: What does TPCM stand for? I am not sure from the sentence above and is it really needed.

This sentence was rewritten to be clearer (lines 75 and 76).

66-67: How do you disregard the Argo profiles which are also part of the NCEI database.

They were not ignored, WOA18 also has quality controlled data from other instruments that are objectively analyzed, if the measurements from the other instruments were significant they would affect the statistical mean and would not be similar to the Argo database.

68: This is unclear.

The sentence was rewritten (line 89).

69: No this is not right, all data of Argo are part of the RTQC. I assume you wanted to say that 30% of the data have not yet undergone DMQC.

Yes, we assumed that the reader would find it easier to differentiate between the data that have passed through the RTQC and those that have not. The sentence was rewritten to avoid confusion (line 90), here it is important to relate 30% to the RTQC.

70: No, that also needs to be rewritten. The real time quality controls create flags for the real time data and the delayed mode quality control creates flags for the adjusted data. The

meaning of flags is the same and there is no distinction from best to worst. The meaning of 1 is good data, 2 is possibly good data, 3 is possibly bad data and 4 is bad data.

In the real-time quality control, there are also data adjustment and the adjusted data flags are replaced in the delayed-mode quality control. We extended the paragraph to be clearer (lines 91-96).

72-73: I have no idea what you mean and it sounds wrong.

This comment is unclear. The paragraph was rewritten to try to be clearer (lines 96-99) and the "Supplementary Material #2" was added at the end of this document, to show that it is not wrong.

97-101: This polygon for testing does not make any sense to me. Why would you mix data from the Pacific with those from the Caribbean. They are totally different in hydrographic properties.

We are not mixing data, here we are talking about the study area used for the web application. We are not saying that we will apply the proposed algorithm to all data within the EEZ of Mexico, we are describing the methodology (data download, processing, etc.), also, it is explicitly said that the proposed algorithm was implemented to the web application (lines 148-150) and that using polygon within this zone can be applied. If the users who are using the proposed algorithm decide to use it in very large areas and with different hydrographic characteristics, they are not making good use of the algorithm.

120-121: At the scale of the TS diagrams not much can be seen about the data quality of the DMQC Argo data versus other data bases. If you want to draw any conclusion about the quality of the argo delayed mode process you need to focus on the deep waters or device other presentations.

We understand your comment and appreciate it, but these two paragraphs were eliminated as a suggestion by the reviewers to avoid redundancy with what was already explained in the methodology (lines 152-164).

125: No, that is not right. Real time data with qc 1 do not per se contain salinity drift. Not all floats receive a salinity correction in delayed mode, although some floats might.

See the "Supplementary Material #2" attached at the end of this document.

128-129: Could you please clarify which qc flags you have used in figure 5 for the real time data? Is it only qc=1 ? And are you only plotting those profiles as real time which have not yet been through delayed mode or also those that been have delayed mode data in which either a correction was applied to bring float data into consistency with climatology or flag data as bad (QC=3,4)? I would be surprised to have so many bad real time data goto into the data stream in which gross errors are not detected or picked-up by priority list for dmqc.

We believe that it is not necessary, in the previous paragraph (lines 170-172) it is explicitly stated that "… The RTQC and DMQC data were plotted in the TS diagrams together…". The data from both controls are being plotted and there is no mention of any kind of distinction using the flags, so all the data is being used.

134-135: I have no idea what you mean.

This comment is unclear. In this sentence it is said because we applied the cluster analysis with data greater than 1500 m. Another way of saying that the variations in salinity and temperature are imperceptible at these depths, that is to say that the salinity data are less dispersed than at lower depths, as mentioned in the previous lines (lines 177 and 178).

140-143: In the dmqc process at Argo the limit for correcatability is drawn at at threshold of 0.05. Beyond that data are flagged as bad . And it seems unlikely to me to learn anything from data that are indicated in the figure an need to be corrected by these huge offsets.

And that is precisely why the proposed algorithm makes a separation between the data with salinity drifts and those that follow the patterns of the data in DMQC. This is a filter, not an adjustment and it is worth it because even the flags indicated in 1 show salinity drifts, as shown in "Supplementary Material #2" at the end of this document.

144: Figure says it is december.

Yes, it was a mistake, it must say December (line 189).

145-146:

This markup has no attached text.

150-151: I have no idea what you mean.

This comment is unclear. The phrase highlights how after the first filter there are still salinity drifts in some months (lines 196-198).

Caption of Figure 7: The headers of the three subpaneles are all the same, if these are iterations steps these shoudl be indicated.

Suggestion accepted (now Figure 6).

163-164: I have no idea what you want to say.

This comment is unclear. This complete paragraph explains how filters can be used (lines 210-214) and they are not limited to using just one filter or another, but also a combination of them.

171-72: Why do you think it is worthwhile showing the the selection of profiles in the polygon works? Isn't that the mandatory for the algorithm and just a simple prequisite for your analysis? And as indicated above I don't think hydrographically the data in the golf of Mexico should be mixed with the Pacific.

Because right now we are talking about the web application, not just the algorithm. The data will not be mixed unless the user who uses the algorithm so decides, as explained above.

173: What do you mean by that?

That there are more than 350 points that limit the polygon, is just a sentence to describe the study area that is being talked about at the moment.

182: Why are you showing here just one float? What is the connection to your analysis?

It is an example of what the web application does. In these paragraphs of the manuscript we are not talking about the analysis.

196: As said above that is not the purpose of the real time flags. They are meant to help the operational users avoid bad data.

This is correct and for this reason we propose a filter that is meant to discard data with salinity drifts.

196:

This markup has no attached text.

198: Is that a typo?

Yes, it is a mistake (line 266).

198: That is the recommendation from the Argo programme to use onyl delayed mode data for scientific analyis and wait for dmqc to be perfomed in due time.

We rewrote the paragraph to be more specific (lines 264-266).

199-200: That should hopefully be the case otherwise the whole dmqc process is not working ok.

This is precisely why we highlighted it in the discussion, so that the reader does not lose sight of it.

200: The drift in the real time data you have shown here are too big to really make good use of the corrected data. It would be more interesting to see if you could detect smaller misfits.

We will take it into consideration.

208: Typo?

Yes, it is a mistake (line 276).

210: Typo?

Yes, it is a mistake (line 278).

215-216: I have no idea what you mean?

This comment is unclear. This sentence lists some of the platforms available to access Argo data, so that the reader knows which platforms we are talking about in the following lines.

219-224: This paragraph is hard to understand please revise it. And JCOMMOPS has been renamed to OCEAN-OPS. I also don't see the connection to your work. At ocean ops you can

certainly select data from different predefined ocean areas and download them , but there is no data analyis invovled.

At the time this work was written, its name was JCOMMOPS, but we already updated it (lines 290 and 291). The connection with the work is to propose that this and the other platforms could include a PIP algorithm to improve and increase the amount of statistics they provide to their users.

245: This is absolutely not sufficient as data acknowledgement. Argo is a living data set and you need to indicate when you have downloaded the data. there are dois provided for monthly snapsjhots. Please refer to the Argo data management page. Also you need to give information on the WODB data set.

Suggestion accepted (lines 317-321).