Ocean Science Discussions

Ocean Sci. Discuss., referee comment RC2
https://doi.org/10.5194/os-2021-22-RC2, 2021

Comment on os-2021-22

Anonymous Referee #2

Referee comment on "Filtering method based on cluster analysis to avoid salinity drifts and recover Argo data in less time" by Emmanuel Romero et al., Ocean Sci. Discuss., https://doi.org/10.5194/os-2021-22-RC2, 2021

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.

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?

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.

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.

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

Please also note the supplement to this comment:
https://os.copernicus.org/preprints/os-2021-22/os-2021-22-RC2-supplement.pdf