Review for Chen et al manuscript *Episodic subduction patches in the western North Pacific identified from BGC-Argo float Data*

In this manuscript, Chen and co-authors analyse the occurrence of subduction events using a dataset of 43 BGC-Argo floats in the Kuroshio Extension (western North Pacific). As demonstrated in Llort et al, (2018), BGC-Argo float profiles are a costly-efficient way to observe events of small-scale water subduction, also known as eddy-pump or eddy-subduction pump (ESP, Boyd et al, 2019). Recent studies have shown that this mechanism can contribute to the biological carbon pump but there are still large uncertainties on how important this contribution is compared to other pathways of carbon export (Boyd et al, 2019; Resplandy et al, 2019).

Chen et al contribute to this knowledge gap by thoroughly analyzing new data in an important region where the ESP has not yet been quantified. Besides, authors revise the only detection method published to date (Llort et al, 2018) and provide a new version. The paper is very well written, with great figures and appropriate citations. Although these original elements justify the publication of Chen et al at Biogeosciences journal, there are two major issues (described in paragraphs below) in the current version of the manuscript that needs to be addressed before acceptance. For this reason, I would recommend major revisions.

The two issues are related to the estimation of the Carbon and Oxygen inventories and fluxes associated with the episodic events.

1) The first and most important issue is that the dataset used in this study contain no measurement of optical backscattering, the data generally used to estimate particulate organic carbon (POC) with BGC-Argo floats. Instead, the only biogeochemical variable sampled by the floats used here is oxygen. That's unfortunate and strongly impacts the estimates and conclusions on the role played by subduction events to export carbon into the deep ocean. Authors try to circumvent this handicap by applying a C:O ratio that is not up-to-date nor adapted to this purpose. Authors justify this ratio (C:O, 117:170) citing two publications (Anderson and Sarmiento, 1994 and Feely et al, 2004). None of the two publications are referenced in the manuscript, besides there's no reference to C:O ratio in Feely et al, 2004 paper. I did find the cited ratio in Anderson and Sarmiento, 1994 (A&S94 hereinafter), who concluded that at large scale there's no significant change on the C:O ration between 400 and 4000m depth. A&S94 include however a caveat about the use of this ratio that Chen et al authors ignored or neglected. The very last paragraph of A&S94 states:

> *As these are long-term, basin-wide, net-ecosystem utilitzation ratios, **they might not be applicable on short timescales or length scales** (...). Also, **these ratios may not be applicable** to high-latitude regions or **the ocean above 400m.***

These two sentences, which are not addressed by Chen et al, suggest that the use of C:O for subduction events is inadequate. Authors should address this major issue, either by removing all the analysis of the carbon export, looking for other methods to estimate carbon from the data available, or incorporating additional analysis to propose C:O ratio that can be used in the context of episodic subduction in this region. In the latter case, an analysis of uncertainties will also be necessary.

The analysis on Oxygen injections is on the contrary valid as it is based on measurements from the floats. Besides it contains interesting thoughts on how these injections might ventilate low oxygen subsurface waters in low-to-mid latitude oceans. So, an option would be to focus the paper only on Oxygen injections and include one paragraph on why carbon export fluxes could not be estimated.

2) The second issue impacts the estimates for both Carbon and Oxygen fluxes. While authors took great care on the detection and analysis of anomalies and its associate inventories (Eq 4) the assumption used for transforming these inventories to export fluxes is not convincing. In lines 241-244 authors assume that the average lifetime of subducted water patch is 1 year. Authors argue that they apply this assumption to avoid choosing an arbitrary vertical velocity, but this average lifetime seems arbitrary to me too. To my knowledge, there's no estimates of how much time these water masses maintain differentiated properties in the mesopelagic zone and we can imagine numerous physical and biogeochemical processes influencing them. These processes cannot be considered with the current dataset but authors could do a detailed analysis of the mixed layer depth variability. How deep is mixing penetrating? How often do storms reset vertical distributions? Which is the regional variability of the maximal MLD and its variance over the region of interest? This analysis would provide some insights on which water masses will remain below the permanent pycnocline (in the current version of the manuscript authors used the value 450db but this is not justified), and on the average lifetime of anomalies above the permanent pycnocline.

More specific comments:
L202 I suggest removing the last sentence of the paragraph. The idea that the improved detection method would detect more events in other datasets (I understand that authors are referring to Llort et al, 2018 dataset) appears several times in the manuscript. I don't see the interest of this statement without providing any data to back it up. It would be interesting to test the new detection method on the same dataset used in Llort et al, 2018 to quantify the improvement in detection. Without this quantification it doesn't make much sense to compare the two studies as Llort et al, 2018 dataset covered the whole Southern Ocean with lots of profiles in low EKE regions, while Chen et al focuses in a much smaller area and with floats more localised in a region of mid-to-high EKE.

L293-4 Again, a speculative comparison in "more signals of subduction (...) that had been previously recognized." Previously by who? If authors refer to Llort et al, 2018 I think the comparison is not valid for the arguments exposed above.

L297 I don't understand this paragraph. In particular, I don't understand why authors argue that "The ephemeral nature" of the anomalies suggest that "they stemmed from distinct subduction events". Some paragraphs above authors assumed that these anomalies could last up to 1 year...

L305 I feel that the use of "modified" here is not clear. Modified respect to what? I assume authors refer to Llort et al, 2018 but I don't see the necessity to compare both methods, except if authors decide to properly quantify the performance of one against the other. This comes back to the comments for L2XX. It should be better to talk about "our method" or create some acronym/code to be sure the reader understands which method is being used or referred to.

Fig 6 The peak in March is surprising and it's not clear to me how can be explained by "large-scale subduction" (L339). Large-scale subduction should not be detected by your method, why it is then affecting to the number of event detected?
Also, could you plot the average MLD dynamics of all floats in the dataset to show the shoaling during this time of the year.

L483 Remove word "currently" as the sentence starts with "Current global-scale...".

L485 Replace "added" by "additional"

As recommendation I would suggest making the detection method public and available by uploading the scripts in GitHub or similar. I didn't do that when I published my paper and I strongly regret it.

Thanks for your work!

Joan Llort