Response to short comment #1, by Dr. Jens Müller

We thank Dr. Müller for the helpful comments and suggestions, each one is adressed below (comment in black, response in red).

Short summary
The authors present an update of the GLODAPv2.2019 data product, by adding new data from 106 cruises. Before addition, observations of 12 core variables have undergone a primary (f flag) and secondary (qc flag) quality control. The secondary quality control is based on the comparison of new data with those contained within GLODAPv2.2019. Adjustments were - if necessary - applied to the new data, in order to correct for biases between measurements from different cruises, but preserve temporal trends in the variables. The merged data product includes observations from 946 cruises and extends until 2019.

General comments
The overall quality of this data product and its description in the companion manuscript appear very high. I have no general comments which would require a revision of fundamental aspects of the data set as a whole. The updated product GLODAPv2.2020 is an invaluable contribution for the scientific community and an essential prerequisite to reach the stated goal of documenting "the state and the evolving changes in physical and chemical ocean properties, e.g., the inventory of the excess CO2 in the ocean".
This review is written from the perspective of a new user of the product.

Specific comments
Following specific issues were identified and might (if taken into account) require a revision of some aspects of the data product:

-l.412: "Neutral density was calculated using Sérazin (2011)." It should be noted that the reference given here refers to a master thesis and that the proposed polynomial approximation of neutral density in this thesis has not undergone peer review. Furthermore, polynomials were fitted to a preliminary neutral density data set with known issues (pers. comm. P. Barker and G. Sérazin). To take those limitations into account, the computed density variable gamma could either be revised, removed or labelled as preliminary in the main text.
Thank you for alerting us on this issue.

- Changes made: We have replaced the neutral density values in the merged product files with values calculated according to Jackett and McDougall (1997). This is described in Sects. 3.3.1 and 3.3.2.

-It might be helpful for some users if the f flag value would distinguish between interpolated and calculated values.
A WOCE flag value of 0 does indeed indicate values that either have been interpolated or calculated. Interpolation is only carried out for salinity, oxygen and nutrients while calculations are only carried out for seawater CO_2 chemistry variables. As such, interpretation of the 0 flag is unique. This is now clearly stated in the manuscript, in section 3.1, 3.3.2 and 6. Whether to change this and introduce a new flag, is a topic that will be considered for future updates.

-l.190: It is stated that "not all offsets larger than the initial minimum limits have been

adjusted for…. Conversely, in some cases where data and offsets were very precise and the cruise had been conducted in a region where variability is expected to be small, adjustments lower than the minimum limits were applied." I was wondering whether at all an initial minimum adjustment limit needs to be defined and what the added value of this definition is. Would it be possible to define an offset-to-precision ratio that could rigorously be applied to all decisions?
This is true, a limit based on the criteria mentioned (offset-to-precision ratio) seems more meaningful, and we will explore ways to implement this for future versions of this data product.

-l.249: An adjustment of -3 μmol kg^{-1} was applied, although an offset of 3.68 ± 0.83 μmol kg^{-1} was found. Is this difference intentional? What is the general rule on how the adjustment values are set?
Adjustments are typically round numbers relative to the precision of the variable considered. There are no particular rules about rounding down or up; we look for example, on whether there is a difference in the offset in recent vs older crossovers. We also consider additional evidence from the other methods. Here, we settled for -3 μmol/kg, as the CANYON-B and CONTENT analyses suggested a bias of 3.4 and 2.7 μmol kg^{-1}, respectively. This also helps to make the adjustment as small as meaningfully possible, in case there actually is an increasing trend in TCO_2 from uptake of anthropogenic carbon.

- Changes made: The sentence in question has been revised to : "In this case -3 μmol kg^{-1} was applied. This is somewhat less than indicated by the crossover analysis, but such a small adjustment is supported by the CANYON-B and CONTENT results (Sect. 3.2.3)."

Technical corrections
Following comments address the presentation of the data product, and cover also aspects that are not purely technically:

-The presentation of the flagging scheme could be improved, aiming at clarity from a user perspective. Taking table 2 as an example, it confused me that labels 0-9 are presented, whereas the data product only uses f flag values 0, 2, and 9. Readers currently need to refer to footnotes in column "Merged product files" to find out that WOCE flags 6 and 7 were set to 2, whereas 3, 4, 5, and 8 were set to 9. Furthermore, the term "Not used" might add to the confusion, as it can easily be misinterpreted as "observations were not used" rather than the intended "the flag value was not used". Starting table 2 with the first column indicating f flag values that are actually used in the data product would greatly improve clarity and avoid potential misinterpretation of the flagging scheme.
We agree that this should be better described and have made changes in the text and in table 2, which hopefully convey differences in flagging schemes between the original exchange formatted data files and the merged product files.

- Changes made: The underlined text has been added to the paragraph where the WOCE flags are first mentioned in Sect. 3.1: "Each data column (except temperature and pressure, which are assumed "good" if they exist) has an associated column of data flags. For the original data exchange files, these flags conform to the WOCE definitions for water samples and are listed in Table 2. For the merged and adjusted product files these flags are simplified: questionable (WOCE flag 3) and bad (WOCE flag 4) data are removed and their flags are set to 9. The same procedure is applied to data flagged 8 (very few such data exist). WOCE flags 1 (Data not received) and 5 (Data not reported) are also set to 9, while 6 (Mean of replicate measurement) and 7 (Manual chromatographic peak measurement) are set to 2, if the data appear

good. Also, in the merged product file a flag of 0 is used to indicate a value that could be measured but is somehow approximated: for salinity, oxygen, phosphate, nitrate, and silicate, the approximation is conducted using vertical interpolation; for seawater CO_2 chemistry variables (TCO_2, TAlk, pH, and fCO_2), the approximation is conducted using calculation from two measured CO_2 chemistry variables (Sect 3.2.2). Importantly, interpolation of CO_2 chemistry variables is never preformed, and thus a flag value of 0 has unique interpretation."

- Changes have also been made in Table 2, specifically, we have replaced 'Not used', in the third colum, with 'Flag not used', to make it more clear that it is the flags that are not used. We prefer to leave the column order unchanged, as having the scheme for original files first and product files last, aligns with the extent to which files are modified with our procedures.

Likewise, in table 5 rownames (first column) are not intuitive. I was wondering what -888 does stand for. Does this label occur in the data set?
Reviewer 3 also pointed out lacking explanation of the -888, and similar, codes, which are used in the online Adjustment Table, and as shorthand for various actions in the manuscript. We agree these needs explanation.

- Changes made: the meaning of -888 and the other codes are now explained in a new paragraph added to the start of Section 4.2: "The secondary QC has 5 different outcomes, provided there are data. These are summarized in Table 5, along with the corresponding codes that appear in the online Adjustment Table and that are also occasionally used as shorthand for decisions in the coming text. The level of secondary QC varies among the cruises. Specifically, in some cases data were too shallow or geographically too isolated for full and conclusive consistency analyses. A secondary QC flag has been included in the merged product files to enable their identification, with "0" used for variables and cruises not subjected to full secondary QC (corresponding to code -888 in Table 5) and "1" for variables and cruises that were subjected to full secondary QC. The secondary QC flags are assigned per cruise and variable, not for individual data points and are independent of—and included in addition to—the primary (WOCE) QC flag. For example, interpolated (salinity, oxygen, nutrients) or calculated (TCO_2, TAlk, pH) values, which have a primary QC flag 0, may have a secondary QC flag of 1 if the measured data these values are based on have been subjected to full secondary QC. Conversely, individual data points may have a secondary QC flag of 0, even if their primary QC flag is 2 (good data). A 0 flag means that data were too shallow or geographically too isolated for consistency analyses or that these analyses were inconclusive, but that we have no reasons to believe that the data in question are of poor quality. Prominent examples of this for this version are the 10 new Davis Strait cruises: no data were available in this region in GLODAPv2.2019, which, combined with complex hydrography and differences in sampling locations, rendered conclusive secondary QC impossible. As a consequence, most, but not all, of these data (some being excluded because of poor precision after consultation with the PI) are included with a secondary QC flag of 0. "
- A new table 5 has been added:

Table 5: Possible outcomes of the secondary QC and their codes in the online Adjustment Table

Secondary QC result	Code
The data are of good quality, consistent with the rest of the dataset and should not be adjusted.	0/1[a]
The data are of good quality but are biased: adjust by adding (for salinity, TCO_2, TAlk, pH) or by multiplying (for oxygen, nutrients, CFCs) the adjustment value	Adjustment value
The data have not been QC'd, are of uncertain quality, and suspended until full secondary QC has been carried out	-666
The data are of poor quality and excluded from the data product.	-777
The data appear of good quality but their nature, being from shallow depths, coastal regions, without crossovers or similar, prohibits full secondary QC	-888
No data exist for this variable for the cruise in question	-999

[a]The value of 0 is used for variables with additive adjustments (salinity, TCO_2, TAlk, pH) and 1 for variables with multiplicative adjustments (for oxygen, nutrients, CFCs). This is mathematically equivalent to 'no adjustment' in each case

Finally, several important information about flags are given in section 3.3.2 (Merging), but might be better placed in 3.1 (Data assembly and primary quality control) and 3.2 (Secondary quality control).

- Changes made. The information about WOCE flags has been added to Sect 3.1, as explained in response to Technical Correction #1, and information about the secondary QC flags has been added to paragraph 4.2, in response to Technical Correction #2

-l.45: The entire data product contains “measurements from more than 1.2 million water samples”. However, this number decreases significantly when the number of available core variables is considered. As an example, I found in the merged master file <0.5 million dissolved inorganic carbon (tco2) observations and <10.000 observations with all core variables being available (in both cases ignoring f and qc flags). To this end, readers might benefit from a more detailed description of the data set. Giving expected row numbers for a few exemplary combinations of subsetting conditions would be valuable

- Changes made We have included some illustrative examples in a new paragraph in Section 6: “The total number of data records are 1 275 558. Records with measurements for all 12 core variables, salinity, oxygen, nitrate, silicate, phosphate, TCO_2, TAlk, pH, CFC-11, CFC-12, CFC-113, and CCl_4 are very rare; only 2026 records have measured data for all 12 in the merged product file (interpolated and calculated data excluded). Requiring only two measured seawater CO_2 chemistry variables in addition to all the other core variables brings the number of available records up to 9 230, so this is also very rare. A major limiting factor is simultaneous availability of data for all four freon species, only 26 277 records have measurements of CFC-11, CFC-12, CFC-113, and CCl_4 while 400 587 have data for at least one of these (not considering availability of other core variables). A total of 398 757 records have measured data for two out of the three CO_2 chemistry core variables. The number of measured fCO_2 data are 33 924; note that these data were not subjected to quality control. The number of records with measured data for salinity, oxygen, and nutrients are 798 703, while the number of records with salinity and oxygen data are 1 077 859. All of these numbers are for measured data, not interpolated or calculated values. ”

References

Jackett, D. R. and McDougall, T. J.: A neutral density variable for the world's oceans, J Phys Oceanogr, 27, 237-263, 1997.