Response to review by referee #2, Dr. Nicolas Metzl

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

General comments:
Since 15 years GLODAP data-bases (from 2004 to 2019, including CARINA, PACIFICA) are widely used in the community, not only to evaluate the change of CO2 in the ocean or acidification (e.g. Gruber et al 2019; Jiang et al 2019), but also to compare and validate ocean and climate models (e.g. CMIP5, Bronselaer and Zanna, 2020 for a recent publication). The GLODAP data-set is also an important synthesis for GOA-ON activities and to construct climatology (e.g. Broullón et al, 2020).

Here, authors present an updated version of the GLODAP effort. This includes 106 new cruises quality controlled (QC), inclusion of new fCO2 observations (not QCed) and comparison of secondary QC with reconstructed properties using neural network methods (named CANYON-B and CONTENT).

The effort consists mainly in (i) format and check the data received from PI or available in different locations (NCEI/OCADS, PANGAEA, CCHDO), (ii) performed a secondary QC to identify data biases (if any) and separate from real temporal changes of the properties that could be low relative to the mean concentrations and (ii) construct final formatted products with adjusted data and associated flags for easy use at global or regional scales.

The paper is basically structured from the previous manuscript (Olsen et al 2019) and I therefore have only few comments regarding this new version (v2020). Most suggestions are for clarity, here thinking to readers that would discover only now the GLODAP project (e.g. new students in the field).

As fCO2 data are now included, GLODAP is in a way a companion data-base to SOCAT dedicated to surface fCO2 data (Bakker et al 2016) also annually updated (Bakker et al 2020). Both products were already used together for specific analysis (e.g. comparing pH fields from GLODAP and SOCAT, Jiang et al 2019). It might be useful for future to attempt incorporate fCO2 data that are in GLODAP but not yet in SOCAT. In this context few words might be added at the end in the conclusions/perspectives.
This is an interesting suggestion, thanks. There are indeed many sources of fCO_2 data, and there are also potentially many issues related to the various measurement techniques and different levels of, and approaches to, their QC. For GLODAPv2.2020, fCO_2 was not quality controlled. A unified look at ocean fCO_2 data seems worthwhile but would be very demanding, in particular related to differences in sampling strategies.

- Changes made: The following sentence was added to the second paragraph of Sect. 6 Summary, to make it clear that the fCO_2 data in GLODAP have not been subjected to quality control: "The number of measured fCO_2 data are 33 924; note that these data were not subjected to quality control."

  The following sentence has been added at the very end of Sect. 6 Summary, to make it clear that QC of fCO_2 data is needed, although at this stage we are not in a position to suggest any particular procedure: "As mentioned above, the included fCO_2 data have not been subjected to quality control, therefore no uncertainty estimate is given for this variable. This should be conducted in future efforts."

In this version, authors used CANYON-B and CONTENT methods (I think this was not systematically performed in v2019). This is a new and an elegant way to check and compare secondary control (and bias if any). This is a new step in GLODAP that might be recalled in the abstract for this version.

Thank you. We now mention this in the abstract.

- Changes made: We have added the following sentence to the abstract: ”Comparisons to empirical algorithm estimates provided additional context for adjustment decisions, this is new to this version.”

Something not very clear concerns the QC for historical cruises. With the new cruises in hand, I was not sure at the start if the QC of previous cruises in the same regions has been checked again and would lead to new corrections for cruises already in GLODAPv1, CARINA or v2019. However, as specify in the manuscript (line 145) I understand that a complete revision of QC would be performed in 2023 (after 3d GO-SHIP).

We realise that this is mentioned rather late in the manuscript, but hope that the paragraph on the different types of GLODAP updates now included in the introduction in response to the comment from Matthew Humphreys, clarifies this early on.

- Changes made: The following paragraphs have been added at the end of the introduction:
  “Within this there are two types of GLODAP updates: full and intermediate. Full updates involve a reanalysis, notably crossover and inversion, of the entire dataset (both historical and new cruises) and all adjustments are subject to change. This was carried out for GLODAPv2. For intermediate updates, recently-available data are added following quality control procedures to ensure their consistency with the cruises included in the latest GLODAP release. Except for obvious outliers and similar types of errors (Sect. 3.3.1), the data included in previous releases are not changed during intermediate updates. Additionally, the GLODAP mapped climatologies (Lauvset et al., 2016) are not updated for these intermediate products. A naming convention has been introduced to distinguish intermediate from full product updates. For the latter the version number will change, while for the former the year of release is appended. The exact version number and release year (if appended) of the product used should always be reported in studies, rather than making a generic reference to GLODAP.
  Creating and interpreting inversions, and other checks of the full data set needed for full updates are too demanding in terms of time and resources to be preformed every year or two-years. The aim is to conduct a full analysis (i.e., including an inversion) again after the third GO-SHIP survey has been completed. This completion is currently scheduled for 2023, and we anticipate that GLODAPv3 will become available a few years thereafter. In the intermin, presented here is is the second intermediate update, which adds data from 106 new cruises to the last update, GLODAPv2.2019 (Olsen et al., 2019). “

Also, many colleagues used the GLODAP gridded products that were constructed from GLODAP-v2 (Lauvset et al 2016). Will you also revisiting this gridded product now or wait for the 2023 version ? This might be specified in the manuscript.

The gridded product will not be updated now. The changes would likely be rather small, as the main source of uncertainty in the gridded product is lack of observations in certain regions. The data added in GLODAPv2.2019 and GLODAPv2.2020 are mostly repeat observations, extending the coverage in time and not in space. We cannot commit, now, to making new climatologies for v3. This depends on funding. Therefore, we simply add a statement that the intermediate products are not accompanied by a gridded product

update.

- Changes made: The sentence “Additionally, the GLODAP mapped climatologies (Lauvset et al., 2016) are not updated for these intermediate products.” has been included in the second final paragraph of the introduction.

Another remark concerns the new cruises to be added in GLODAP. I understand that new cruises (106) were recently obtained from NCEI or PANGAEA or from PIs. However, I suspect there are many other cruises in the community (published) and it would be useful to find the best way to get more cruises in the future and invite new PIs to contribute.
Yes, there is certainly room for improvement. Right now, apart from close interaction with GO-SHIP and CCHDO, the level of formalization for addition of data is very low. While no changes were made to this end in the manuscript, we will explore ways to obtain more publicly available datasets.

Overall, I recommend publication after few minor revisions.

Below I list specific and minor comments (mostly details for clarity for a reader who discover Glodap for the first time). At the end of the review few technical questions regarding the files on-line.
;;;;;;;;;;;;;;;;;;;;;;

Specific comments:
C-01: Title: The title includes only acronyms of the project (GLODAP). Would it be useful to recall that this concerns Ocean biogeochemical observations in the water column? A Suggestion for a title: “An updated version of global interior ocean biogeochemical observations, GLODAPv2-2020”.
That is a good suggestion

- Changes made: Title has been changed to “An updated version of the global interior ocean biogeochemical data product, GLODAPv2.2020”

C-02: Page 2, line 44: “the inclusion of available discrete fugacity of CO2 (fCO2) values in the merged product files”. Does this new inclusion concerns only the new cruises added in v2020 or did you also add this parameter for historical cruises ? (this is specify later, Line 369).

- Changes made, added “(also for historical cruises)” to sentence in question.

C-03: Page 4, Line 121: “The data collected across the Davis Strait”. Maybe specify where is the Davis Strait for those not familiar with the Indian Ocean….(….Atlantic of course)
☺

- Changes made, added “between Canada and Greenland” after “Davis Strait”

C-04: Page 5, Line 175: For new users: Not sure to clearly understand all Flag definitions listed in Table 2.
Indeed, this Table is a bit brief, and may lead to misunderstandings, as also pointed out by Jens Muller in his short comment.

- Changes made:
  - We have added a citation to Swift (2010) in the Table header, which provides full details on the flags used in the exchange format original data files
  - We have expanded the table caption to make it clear that the flagging scheme in the merged product files is simplified (added text is underlined):

"Table 2. WOCE flags in GLODAPv2.2020 exchange format original data files (briefly; for full details see Swift, 2010) and the simplified scheme used in the merged product files"

- We have added the underlined text in the paragraph of section 3.1 where Table 2 is first mentioned: "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 flag 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."
- For the 'Merged product files' column in Table 2 we have changed "Not used" to "Flag not used"

C-05: Table 2: for clarity, it might be useful to assign different flag for interpolated and calculated values (both flag 0). Maybe for the next version.

- Changes made. To be clear about the unique interpretation of the 0 flag for different variables, we have added the following sentences in Sect. 3.1: "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."

C-06: In table 2, you list "b" "Data are not included in the GLODAPv2.2020 product files and their flags set to 9. " Does that mean that original flag 3 (Questionable but sometimes maybe real signal) are not included in the files ? However this is explained later, line 395
Yes, these are removed from the product file. We now explain this in the paragraph that introduces the table (see response to C-04).

C-07: In table 2, you list "c" for replicate: "Data are included, but flag set to 2 ". This suggests that all replicate are acceptable (or some were also identify as outliers and thus moved to flag 9 or deleted ?).
We now clearly state in the paragraph that introduces this table, that replicates are only kept if the value appears valid, please see response to C-04.

C-08: Page 6, Line 197: "comparison of deep-water averages". Specify the layers here ? How this is selected in the high latitude (e.g. bottom water formations, where anthropogenic CO2

is found to be relatively high in water column ?).
This is the introductory paragraph for Section 3, stating what is to be presented in the subsections to come, among them Sect. 3.2.2, where the full details of the comparisons and what depth layers are used are provided. To avoid repetition, we do not go into these details here.

C-09: Page 6, Line 200: Add reference to CANYON-B and CONTENT (first time listed here) ?

- Changes made: Reference to Bittig et al., 2018, added.

C-10: Page 6, Line 226: "In areas where a strong trend in salinity was present". Any example for this version ?
This is a leftover from the earlier versions of this paper; no strong salinity trends were present in the crossovers evaluated.

- Changes made: sentence deleted.

C-11: Page 7, Line 235: "convection occurs (such as the Nordic, Labrador, and Irminger seas)". How do you select the layer in region of bottom water formation (e.g. SR03 for this version) ? Might be interesting for new readers to show another QC example (as presented in Figure 3 for the North Pacific).
Whether to use 1500 or 2000 dbar is determined on a case by case basis, by looking at the crossover comparisons for the two options, with respect to the accuracy of the information provided on the comparability between the data and whether changes in some layers seems related to actual change. In regions of bottom water formation change is expected, and results are scrutinized in light of this. We have revised the text to make it clear that subjective choices are involved, and that we always evaluate the results for presence of actual change, in order to not adjust this away.

- Changes made: The text on depth limits for crossover analysis has been extended and revised:
  "Either the 1500 or 2000 dbar depth surface was used as upper bound, depending on the number of available data, their variation at different depths, and the region in question. This was evaluated on a case-by-case basis by comparing crossovers with both depth limits and using the one that provided the most clear and robust information. In regions where deep mixing or convection occurs, such as the Nordic, Irminger and Labrador seas, the upper bound was always placed at 2000 dbar; while winter mixing in the first two regions is normally not deeper than this (Brakstad et al., 2019; Fröb et al., 2016), convection beyond this limit has occasionally been observed in the Labrador Sea (Yashayaev and Loder, 2016). However, using an upper depth limit deeper than 2000 dbar will quickly give too few data for robust analysis. In addition, even below the deepest winter mixed layers properties do change over the time periods considered (e.g., Falck and Olsen, 2010), so this limit does not guarantee steady conditions. In the Southern Ocean deep convection beyond 2000 dbar seldom occurs, an exception being the processes accompanying the formation of the Weddell Polynya in the 1970s (Gordon, 1978). Deep and bottom water formation usually occurs along the Antarctic coasts, where relatively thin nascent dense water plumes flow down the continental slope. We cautiously avoid such cases, which are easily recognizable. In order to avoid removing persistent temporal trends, all crossover results are also evaluated as a function of time (see below)."

C-12: Page 7, Line 238: Maybe recall that 49UP20160109 is new while 49UP20160703 was QCed in v2019.

- Changes made: The underlined text has been added to this sentence: "As an

example of crossover analysis, the crossover for TCO_2 measured on the two cruises 49UP20160109, which is new to this version, and 49UP20160703, which was included in GLODAPv2.2019, is shown in Fig. 3."

C-13: The example in Figure 3 shows in 3a blue dots on the map, but I suspect these stations (far east) were not used to evaluate the QC.
This is correct indeed. Thank you for pointing this out, it is certainly worthwhile to mention that only stations shown in panel b are used for the crossover analysis.

- Changes made. The following clarification has been made in the caption of Figure 3: "Panel (a) show all station positions for the two cruises and (b) show the specific stations used for the crossover analysis."

C-14: Page 7, Lines 245-250: For 49UP20160109, maybe specify that no temporal changes was observed for salinity (i.e. you used TCO2 here, not normalized TCO2 as suggested in Line 227 for some cruises).
As mentioned under C-10, salinity normalization was not needed for any crossover, and therefore not mentioned in this manuscript anymore. Thus, we did not mention here that the data were not salinity normalized.

C-15: Page 7, Line 245: Figure 4 shows the TCO2 cross-over for 49UP20160109 versus GLODAPv2-v2019. The cruise 49UP20160703 is also plotted and thus was in GLODAPv2-v2019, although conducted after 49UP20160109 (just to clarify for a new user).
In response to comment C-12, we now mention that 49UP20160703 was included in GLODAPv2.2019.

C-16: Page 7, Line 256: "they are included in the product but with a secondary QC flag of 0 (Sect.6)". Sect 6 (?)
The statement on the lack of full QC on the Davis Strait cruises and the consequential assignment, and interpretation of, secondary QC flag 0 has been moved to Section 4.2 Adjustment Summary.

- Changes made: The following paragraph has been added at 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."

C-17: Page 7, Line 259: "A few new cruises had no or very few valid crossovers with GLODAPv2 data." Which cruises ? Would it be relevant to add a column in Table- Annexe 1 with a remark specifying what kind of secondary QC has been performed for each cruise (e.g. Standard QC, MLR, no QC) ?
For the 106 new cruises, MLR and deep water averages were used in a complimentary fashion, i.e., none of secondary QC were only based on these types of analyses. We have revised Sect. 3.2.3 to convey this.
The type of secondary QC varies not only per cruise, but also per variable. Different types of QC (e.g. Standard QC, MLR, no QC) can be applied for different variables on certain cruises. The various QC types can also be applied in combination. It is not practically possible to include this information in Table – Annexe 1. The most important information regardless appears in the online adjustment table.

- Changes made: The first sentences of Section 3.2.3 have been revised to: "MLR analyses and deep water averages, broadly following Jutterström et al. (2010), were also used for the secondary QC of salinity, oxygen, nutrients, TCO_2, and TAlk data. These approaches are particularly valuable when a cruise has either very few or no valid crossovers with GLODAPv2, but are used more generally to provide more insight on the consistency of the data. The latter was the case for the 106 new cruises; i.e., no adjustments were reached on the basis of MLR and deep water average analyses alone. "

C-18: Page 8, Section 3.2.3: I understand the description but what are the results and which cruise ? Would be interesting to show an example for a cruise that is QCed using MLR.
As no cruise was fully QC'd using MLR, we have not included such an example, but will consider this for the next version of GLODAP.

C-19: Page 8, Line 277: "Altogether 82 of the 106 new cruises included pH data." Here specify this is measured pH, not calculated (so there is no confusion with pH calculated for other cruises).

- Changes made: Sentence revised to (new word underlined) "Altogether 82 of the 106 new cruises included <u>measured</u> pH data."

C-20: Page 8, Line 291: "The pH data of 840 of the 936 cruises in GLODAPv2.2020". Again, specify if pH data here were measured or calculated or both.
We agree that this is not clear, and not all of the 840 cruises included measured pH. This paragraph has been extensively expanded following comments from Dr. Williams, and the specific sentence has been altered to: "In contrast to past GLODAP pH QC, evaluation of the internal consistency of CO_2 system variables was not used for the secondary quality control of the pH data of the 106 new cruises."

C-21: Page 8, Line 305: Maybe recall the mean uncertainty associated to CANYON-B and CONTENT (see table 1 in Bittig et al 2018, i.e. about twice the adjustment limits fixed for GLODAP listed in Table 3).
We are reluctant to mention specific uncertainties for CANYON-B and CONTENT. These vary with depth and with location, and specifically for nutrients, are stated in absolute terms (concentration) in Bittig et al. (2018), rather than relative as used for the adjustment limits,

so the comparability and transferability of directly stating these values is small. We do recognize the need for more clearly relaying that we did in fact explicitly consider these uncertainties in our assessment, however. Therefore we have revised and expanded the sentence in question.

- Changes made: The sentences:
  "Of course, we kept in mind that this relies on the accuracies of the T, S, and O_2 data and of CANYON-B and CONTENT in themselves. Used in the correct way and with caution this tool is a powerful supplement to the traditional crossover analyses. "

  has been replaced with the following:

  "Used in the correct way and with caution this tool is a powerful supplement to the traditional crossover analyses. Specifically, we gave no weight to comparisons were the crossover analyses had suggested that the S and/or O_2 data were biased as this would lead to error in the predicted values. We also considered the uncertainties of the CANYON-B and CONTENT estimates. These uncertainties are determined for each predicted value, and for each comparison the ratio of the difference (between measured and predicted values) to the local uncertainty was used to gauge the comparability."

C-22: Page 8, Line 305: As it is new results presented here (and probably also used in the next version), I think some more information is needed. For CANYON-B and CONTENT are you using results based on GLODAP-v2 data (Bittig et al 2018) or an updated version using GLODAPv2-2019. Is the comparison presented here (Figure 5) validate the QC for the new cruises or validate CANYON-B and CONTENT reconstructed fields? It is reassuring to get about the same results as CANYON-B and CONTENT were trained with GLODAP.
We already state that "These approaches were developed using the data included in the GLODAPv2 product" (line 299-300 in discussion paper). Moreover, from the text and context it is apparent that we validate the new cruises. Finally, we agree that the agreement is reassuring.

C-23: Page 8, Line 308: Figure 5: not easy to see the black dots (measured values).
This is true, and in large part a consequence of the overlap between the predicted and measured values. We prefer not editing the figure. One can see the black dots zooming in. We will add a sentence in the caption to explain that the black dots are in large part hidden by the red/blue dots.

- Changes made: The sentence in the caption explaining the color scheme, has been revised, new text underlined: "Black dots <u>(which to a large extent hidden are by the predicted estimates)</u> are the measured data, blue dots are CANYON-B estimates and red dots are the CONTENT estimates."

C-24: Figure 5: there is no units (to be added in captions ?).

- Changes made: Units have been stated in the caption.

C-25: Figure 5: Like for Figure 3 and 4, it would be nice to show another example, e.g. SR3 or Davis Strait ? Or an example where the comparison between QC from GLODAP and CANYON-B/CONTENT does not work (if any). This is a suggestion not absolutely needed.
Based on the current large numbers of figures in this manuscript, we chosen to not follow this suggestion.

C-26: Page 9, line 320: "Another advantage of CANYON-B and CONTENT is that by

considering the each data point in it self, primary QC issues has been revealed and corrected for some of the cruises." Which cruises ? Give some examples ?
We have revised the sentence and added an example.

- Changes made: The sentence in question has been revised to: "Another advantage of CANYON-B and CONTENT is that these procedures provide estimates at the level of individual data points, e.g., pH values are determined for every sampling location and depth where T, S, and O_2 data are available. Cases of strong differences between measured and estimated values are always examined. This has helped to identify primary QC issues for some variables and cruises, for example a case of an inverted pH profile at cruise 32PO20130829, which has been amended."

C-27: Page 9-10: Section 3.3.1. Lines 332-358: This is a list of revisions and would be better to move this section in an Annex but keep in Section 3.3.1 the fCO2 information (lines 359-375) as it is new data added in v2020.
While we agree that the list is tedious, we prefer to keep it the main text as this is very much what the intention of the manuscript is, documenting significant additions and *changes* to the dataset.

C-28: Page 10: Concerning fCO2, in the GLODAP files there are now both fCO2 measured and calculated in the same column. Authors indicate that all values were converted to 20_C. However, in the data-files, there are fCO2 values with fCO2temp fixed at -9999. I missed something here and not sure if all fCO2 values in the files are at the same temperature, pressure or at local temperature etc: : : Also, there are fCO2 values with flag 0 or 2. What was the criteria for fCO2 with flag 2 ? How users can easily separate the fCO2 measured and calculated in the files ? This is important to clarify if one uses both GLODAP (in surface) and SOCAT to merge both products.
We thank you for checking the product files carefully. Indeed, fCO_2 data without accompanying temperatures occurred. This is an error. The product files have been corrected now. fCO_2 data flagged 2 are measured, while fCO_2 values with flag 0 are calculated, as is the case for all seawater CO_2 data.

C-29: Page 10, line 364: "These calculated TAlk values were, however, not included in v2.2019." Does that mean that all TALK values with flag 0 in the files are only interpolated values (i.e. not calculated as an option suggested in table 2).
With the more extensive explanations of the flags added in Section 3.1 (see response to C-05) we hope that it has become clear that seawater CO_2 chemistry variables, such as TAlk, flagged 0, are not interpolated, only calculated.
Moreover, the sentence in question relates to the previous version of this product, v2.2019. We realize now, that this sentence might cause confusion and is unnecessary.

- Changes made: The sentence has been removed.

C-30; Page 11, Lines 397-398: For flags 6 and 7 now set to flag 2, recall that this only applied for valid data (i.e. obvious outliers deleted also for these replicates ?).

- Changes made: The underlined text has been added to the sentence: "All flags 6 (replicate measurement) and 7 (manual chromatographic peak measurement) were set to 2, provided the data appeared good."

C-31: Page 11, Line 399: "Missing sampling pressures or depths were calculated following UNESCO (1981)." This is obvious but maybe rewrite following: "Missing sampling pressures (resp. depths) were calculated from depths (reps. pressures) following UNESCO (1981)."

- Changes made: Revised according to suggestion.

C-32: Page 11-12, Lines 405 and 432: Flag 0 is used for both interpolated and calculated values. Why not using different flag ? (for next version)
As explained in response to comment C-05, interpretation of WOCE flag 0 is unique, and this is now clearly stated in Section 3.1. Nevertheless, we now also reiterate these principles in this section.

- Changes made: The underlines text has been added to the sentences in question:

  (Line 405) "Missing salinity, oxygen, nitrate, silicate, and phosphate values were vertically interpolated whenever practical, using a quasi-Hermetian piecewise polynomial. "Whenever practical" means that interpolation was limited to the vertical data separation distances given in Table 4 in Key et al. (2010). Interpolated salinity, oxygen, and nutrient values have been assigned a WOCE quality flag 0."

  (Line 432)"Calculated seawater CO_2 chemistry values have been assigned WOCE flag 0. Seawater CO_2 chemistry values have not been interpolated, so the interpretation of the 0 flag is unique."

C-33: Page 11, Line 416. Concerning the "Missing seawater CO2 chemistry variables". Are the calculated properties used only measured data (i.e. TALK and TCO2) or also interpolated values ? In other words, are the fCO2 and pH interpolated values based on calculated fCO2 and pH or recalculated with interpolated TALK/TCO2 ?
We hope that it is clear now, and also in the manuscript, that no seawater CO_2 chemistry variables were interpolated.

C-34: Page 13, Line 486: "For example, Arctic Ocean phosphate, Indian Ocean silicate and TCO2, and Pacific Ocean pH data all show considerable improvements." For Indian, in Table 6 improvement is for TALK, not TCO2 ?
Indeed, this is correct and has been amended.

- Changes made: TCO2 has been replaced with TAlk in the sentence in question.

C-35: Page 15, Line 544: Weatherall et al., (2015): not in references.
Thank you for pointing this out.

- Changes made: Weatherall et al., (2015) has been added to the reference list.

C-36: Now concerning the files, for curiosity I had a look at the Indian.cvs file and have few questions that could be also valid for other basin. The questions below are obvious for someone familiar with Glodap, but mainly addressed here to help new users.

C-36a: Why the QC flags for S or O2 are 0 for several cruises although flag WOCE are2 ? Is it because the secondary QC is not available for these cruises ?
This is correct. We have added text in Sect. 4.2 to explain this (see response to C-16).

C-36c: There are data withWOCE flag=0 for O2, Nitrate, Silicates, Phosphates, TCO2, TALK, pH, and associated to QC flag = 1. Is it because these are interpolated values for a cruise/station for which a secondary QC was performed ? If QC has been performed (QCF=1) one would expect a WOCE flag different from 0 ? I thought the QC is based on original data (not interpolated or calculated). Could that be clarified ?
This is correct. We have added text in Sect. 4.2 to explain this (see response to C-16).

C-36d: There are data with flag 9 associated to QC flag=1. Again, is it because QC flag (0,1)

are assigned for a cruise/station not for each data?
This is correct. We have added text in Sect. 4.2 to explain this (see response to C-16).

C-37: In the data files on-line (e.g. GLODAPv2.2020_Indian_Ocean.cvs) I would suggest to add units for each column.
Yes, and this has been discussed in the GLODAP group as well, and will likely be done for the next update.

C-38: And for next versions, I think for clarity a different flag should be assign for calculated (e.g. fCO2, pH) and interpolated values. This might help some users to select only measured+interpolated values. In references:
As stated earlier, the interpretation of WOCE flag 0 is unique for the different variables. As such there is no need to having a different flag for interpolated (salinity, oxygen, nutrients) vs calculated values (TCO_2, TAlk, pH, fCO_2). We hope this, now, is clear in the manuscript as well.

I think each reference should now have a DOI
Line 663: "Hood, E. M., Sabine, C. L., and Sloyan, B. M.: The GO-SHIP hydrography manual: A collection of expert reports and guidelines, 2010." Specify the publisher ?
DOI ?

- Changes made: publication information has been completed to:
  Hood, E. M., Sabine, C. L., and Sloyan, B. M. (Eds).: The GO-SHIP hydrography manual: A collection of expert reports and guidelines, IOCCP Report Number 14, ICPO Publication Series Number 134, available at http://www.go-ship.org/HydroMan.html (last access: 16 October 2020), 2010.