Response to review by referee #3.

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

This is a "living data" update document that discussed the addition of 106 cruises to the GLODAPv2.2019 data set. These data have been extremely valuable to the community and represent an important asset to maintain and update. The manuscript is well written and informative. I only have a few minor comments below.

Line 92-93: The authors don't distinguish between discrete and in situ sensor measurements here. I assume they are referring to CTD calibration problems with respect to the sensor measurements of salinity and oxygen, not the measurements of collected samples. Please clarify, particularly in light of the merging discussed in section 3.2.1.
Yes, indeed, we are referring to lacking calibration of the data from CTD mounted sensors.

- Changes made: Sentence revised to "For salinity and oxygen, lack of calibration of the data from the conductivity-temperature-depth (CTD) profiler mounted sensors is an additional and widespread problem, particularly for oxygen (Olsen et al., 2016)."

Lines 95-99: The manuscript uses some rather subjective terms without defining their meaning in this context. For example, "poor precision can render a set of data unusable" or "to minimize severe cases of bias". What is the definition of poor precision or severe bias?
We now provide more concrete information on what is meant with these terms, without going overboard with numbers and definitions as this is a general introduction, and as such we are reluctant to discuss details about each and every variable considered. Besides, the data are evaluated on a case-by-case basis, depending on region and availability of already existing data, for instance; we do not have a strictly enforced global set of limits.

- Changes made:
  The sentence "In rare cases poor precision can render a set of data unusable"
  has been revised to:
  "In rare cases poor precision - many multiples worse than that expected with current measurement techniques - can render a set of data of limited use."

  The sentence: Adjustments are applied on the data to minimize severe cases of bias"
  has been revised to:
  "Adjustments are applied to the data to minimize cases of bias that could be confidently established relative to the measurement precision for the variables and cruises considered. "

Lines 98, 108: There are a notable number of grammatical errors in the text that should be fixed. A couple of examples are, "Adjustments are applied on the data"(should be 'to the data') or "A particular important source" (should be 'A particularly important source'). Please review the entire document for these grammatical errors.
Thank you for pointing out these errors, which have been corrected. The text has been carefully read and corrected by all authors, many of whom are native English speakers. We hope the number of grammatical errors has been minimized.

Line 123-124: The authors decided to include cruises on the Merian, Meteor, and the Garcia del Cid that did not have any nutrient or carbon data. I thought nutrients and carbon were

the primary parameters for this data set. Why did the authors decide to include these data and not the thousands of other cruises that also do not have carbon data. This seem inconsistent with the goal of this project.

The emphasis for GLODAP is seawater inorganic carbon chemistry, as well as other carbon-relevant and related variables. This includes the transient tracers CFC-11, CFC-12, CFC-113 and SF_6, as these are frequently used to determine ocean inventories of anthropogenic carbon (e.g., Waugh et al., 2006). Rarely measured stable carbon isotopes are also relevant, as these are often used for the same purpose (e.g., Quay et al., 2017), and while we do not quality control such data, they are included to ensure their wider availability. There are not thousands of other cruises with such data. We have now included some text on these deliberations:

- Changes made: The following sentences have been included at the start of Section 2: "Not all cruises have data for all of the above-mentioned 12 core variables; for example, cruises with only seawater CO_2 chemistry or transient tracer data are still included even without accompanying nutrient data due to their value towards computation of, for example, carbon inventories. In some other cases, cruises without any of these properties measured were included – this was because they did contain data for other carbon related tracers such as carbon isotopes, with the main intention of ensuring their wider availability."

Line 150: define data center acronyms the first time they are used, or at least provide links to the data centers.

- Changes made, links to the data centers are now provided

Line 193-195: Were the original data generators consulted before adjustments were made to the data? I believe in the past there was a step that involved checking with the people that originally made the measurement to get their perspective on possible offsets.

Indeed, during preparation of the first version of GLODAP (Key et al., 2004), data originators were contacted for consultation on possible offsets. This practice was abandoned for GLODAPv2, with more than 700 cruises and over 1200 adjustments made, this became impractical. GLODAP is presently a volunteer effort and there is no capacity for routinely approaching principal investigators for every adjustment considered. However, members of the GLODAP Reference Group (i.e., the authors of this contribution) frequently possess first hand experience with the data, or are even the cruise PIs. In exceptional cases, for example where no primary QC seems to have been applied, we do reach out to the PIs.

Line 256: This is the first time that a -888 label is discussed in the text. What does this mean? The same comes in later with -777 and -666 labels.

Thank you for pointing this out. These labels hadn't really been properly explained in this manuscript, only in the GLODAPv2 article (Olsen et al., 2016). In addition, the text in the passage in question (i.e., line 256) better belong in Section 4.2, Adjustment summary as it mostly pertain results.

Changes made: The labels are now explained at the very start of Section 4.2, and presented in a new Table (Table 5). The text on the Davis Strait cruises, pointed out by the reviewer, has been moved from Sect 3.2.3, and used as an example of cruises not fully QCd. The first paragraph in Sect. 4.2 is now: "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. "

Lines 280-282: Why did the authors use the full GLODAPv2 data to estimate TAlk from Salinity. Wouldn't it make more sense to calculate an average ratio for the data from that cruise rather than use a global ratio that includes data from other oceans? Also, doesn't the ratio change with depth

TAlk is estimated here, with the purpose of converting pH measurement scale and/or reporting temperature/ pressure. The uncertainties introduced by a using global ratio instead of actually measured TAlk are very small. For the scale conversions the uncertainties are on the order of 10^{-7} pH units, which is fully negligible. For the temperature and pressure conversions the uncertainties are 0.001 pH units (evaluated using 2 standard deviations around the 67 ratio, i.e. TAlk/S = 67 ± 4.1 μmol/kg/permil). This is an order of magnitude smaller than the stated uncertainty for the pH in the merged product, 0.01-0.02 units.

Calculating the TAlk vs. Salinity ratio for the cruise in question is usually not possible since TAlk often has not been measured at all at these cruises (or very few measurements exist).

We do agree, though, that more sophisticated approaches exist for estimating alkalinity (Bittig et al., 2018; Broullon et al., 2019), and since Bittig et al. (2018) is already used to estimate missing PO4 and Si, it will be considered for missing TAlk data in future GLODAP updates.

- Changes made: We provide more quantitative information on the uncertainties introduced by the approximation, in section 3.2.4.

References:

Bittig, H. C., Steinhoff, T., Claustre, H., Fiedler, B., Williams, N. L., Sauzède, R., Körtzinger, A., and Gattuso, J.-P.: An alternative to static climatologies: Robust estimation of open ocean CO_2 variables and nutrient concentrations from T, S, and O_2 data using Bayesian Neural Networks, Frontiers in Marine Science, 5, 2018.

Broullon, D., Perez, F. F., Velo, A., Hoppema, M., Olsen, A., Takahashi, T., Key, R. M., Tanhua, T., Gonzalez-Davila, M., Jeansson, E., Kozyr, A., and van Heuven, S.: A global monthly climatology of total alkalinity: a neural network approach, Earth Syst Sci Data, 11, 1109-1127, 2019.

Olsen, A., Key, R. M., van Heuven, S., Lauvset, S. K., Velo, A., Lin, X. H., Schirnick, C., Kozyr, A.,

Tanhua, T., Hoppema, M., Jutterstrom, S., Steinfeldt, R., Jeansson, E., Ishii, M., Perez, F. F., and Suzuki, T.: The Global Ocean Data Analysis Project version 2 (GLODAPv2) - an internally consistent data product for the world ocean, Earth Syst Sci Data, 8, 297-323, 2016.

Quay, P., Sonnerup, R., Munro, D., and Sweeney, C.: Anthropogenic CO2 accumulation and uptake rates in the Pacific Ocean based on changes in the C-13/C-12 of dissolved inorganic carbon, Global Biogeochem Cy, 31, 59-80, 2017.

Waugh, D. W., Hall, T. M., McNeil, B. I., Key, R., and Matear, R. J.: Anthropogenic CO2 in the oceans estimated using transit time distributions, Tellus B, 58, 376-389, 2006.