

Referee #1

We would like to thank the reviewer for the constructive and insightful comments. They have helped us to further clarify the motivation of this study and, importantly, to improve the discussion of the limitations of the method, contributing to the overall improvement of the manuscript. We have carefully considered the reviewer's suggestions and have revised the manuscript accordingly.

Major comments

This manuscript creates a monthly dataset of anthropogenic carbon based on Argo float data using a method based on a previous paper (Asselot et al. 2024). While I think a monthly analysis and extension to full depth by utilising Deep Argo in the Irminger Sea is very useful and the C_{ant} uptake analysis of the time range studied is sound, overall, I think this is a missed opportunity to further scrutinise and improve the Asselot et al. 2024 results (which were authored by almost the same team of researchers). Instead, it seems like the same algorithm was applied to the Irminger Sea. Some passages in the discussion, e.g. when the benefits of Argo float data are praised, seem like they belong more to the original paper first outlining the method, not in the re-application of the method. I think it would be helpful to discuss limitations and improvements of the method more. After all, even though uncertainty values are given as just being slightly higher than ship-based estimates (7.5 $\mu\text{mol/kg}$ vs. 5.2 $\mu\text{mol/kg}$), C_{ant} estimates are still being calculated from other estimates (i.e. from ESPER_NN and CONTENT), which would make a case for higher scrutiny. There are only very few statements regarding limitations, e.g. L343-344.

We agree that the manuscript would benefit from a more critical discussion of the limitations of the method and from a clearer consideration of how the present analysis relates to, and extends upon, the results of Asselot et al. (2024). To address this latter point, and following your minor comment below (L423–424: “These aims would’ve been good to state at the beginning”), we have revised the last paragraph of the Introduction to provide a more detailed description of the study objectives, including the temporal changes investigated, as shown in the response to the indicated minor comment.

Regarding the more critical discussion of the method limitations, we have included in the revised manuscript a dedicated section 2.4 *Evaluation of neural network predictions against observations* where we assess in more detail neural network performance and expand the discussion of the methodological limitations and uncertainties associated with neural networks; and added a paragraph in the discussion section on the limitations of the neural networks.

In new section 2.4, we evaluate the neural network (NN) performance against the GLODAPv3 dataset (Lange et al., 2026, preprint) for the Irminger Sea domain over the period of study, 2011–2021. More specifically, the T, S, O_2 , and P measurements from GLODAPv3 are used as input predictor variables for the NN, and the NN's outputs (e. g. DIC or A_T) are compared with the corresponding measurements available in GLODAPv3, the most recent data for the Irminger Sea dating from 2019.

For CANYON-B/CONTENT, used to derive DIC and A_T , we performed the analysis separately for both the NN training period (2011-2013) and the NN post-training period (2014-2019). The performance of ESPER_NN (used to derive nutrient concentrations) is limited to the period 2011-2019 within which ESPER_NN was trained, i.e., we could not perform a post-training assessment. Results (units: $\mu\text{mol kg}^{-1}$) are summarized in the table below (Table 1A). Note that for CANYON-B/CONTENT DIC and A_T , values are presented for both the training / post-training periods :

Table 1A.

		DIC	A_T	Silicate	Nitrate	Phosphate
CANYONB/CONTENT	RMSE	4.84 / 5.64	2.93 / 3.42	-	-	-

vs obs.	bias	-2.28 /-2.74	0.25 /-1.22	-	-	-
ESPER	RMSE	-	-	0.66	0.55	0.043
vs obs.	bias	-	-	0.11	0.14	0.007
Sample size	Num. data	913 / 849	1253 / 1255	1512	1510	1235

For DIC and AT, the RMSE increases from 4.84 to 5.64 and from 2.93 to 3.42, respectively, for the post-training period (2014–2021) compared to the training period (2011–2013). Despite this moderate increase in the error, the mean DIC bias remains of similar magnitude for both periods (–2.28 to –2.74). For AT, however, the bias shifts from a small positive value (0.25) to –1.22. Negative biases in DIC and AT are consistent with those reported by Bittig et al. (2018) at the global scale for their validation dataset, which included 20% of the GLODAPv2 data not used for training, as well as several GO-SHIP repeat hydrography cruises not included in GLODAPv2, comprising a total of 19 cruises conducted between 2012 and 2017 across all ocean basins. Despite the moderate increase in the RMSE in the post-training period for DIC and A_T, the presented statistics and the residual distributions shown in Figure 2 (and Appendix C) indicate that CANYON-B/CONTENT largely maintain comparable predictive performance for the post-training period.

The same evaluation was performed for the target variable, C_{ant}. In this case, our reference value for C_{ant} is obtained via the ϕ_{CT}^O method applied to GLODAPv3 T, S, O₂, A_T, DIC and nutrient (silicate, nitrate, and phosphate) bottle data. When using GLODAP data to calculate the difference between C_{ant} derived from the observed values of nutrients, O₂, A_T and DIC and C_{ant} estimated with the neural network derived biogeochemical variables (A_T, DIC and nutrient), the resulting error reflects the uncertainty associated solely with the NN. This error can be decomposed into a bias and a Root Mean Squared Deviation or Root Mean Squared Error (RMSE). The results (units: $\mu\text{mol kg}^{-1}$), covering the period from 2011 to 2019, are summarized in the table below (Table 1B):

Table 1B.

		DIC*	A _T	Silicate	Nitrate	Phosphate	C _{ant}
CANYON-B/CONTENT	RMSE	5.49	4.24	-	-	-	-
vs obs.	bias	-2.79	-3.2	-	-	-	-
ESPER	RMSE	-	-	0.59	0.59	0.045	-
vs obs.	bias	-	-	0.08	0.05	0.007	-
NN-based C _{ant}	RMSE	-	-	-	-	-	3.34
vs ref. C _{ant}	bias	-	-	-	-	-	-0.48
Sample size	Num. data	501	501	501	501	501	501

* DIC GLODAPv3 data include both measured DIC values and calculated DIC values from measured pH and A_T via CO2SYSv3 (Sharp et al., 2020) when measured DIC is not available.

For C_{ant}, we obtained a negative bias of -0.48 $\mu\text{mol kg}^{-1}$, which represents less than 20% the magnitude of the negative bias obtained for DIC. In a very simplified way, in the ϕ_{CT}^O method, C_{ant} is estimated as:

$$C_{ant} = \text{DIC} - C_{bio} - \text{DIC}^{\text{preindustrial saturation}} - \Delta\text{DIC}^{\text{preindustrial disequilibrium}}$$

Where the term C_{bio} relies on A_T and **nutrient** data for its estimate, and the term $\Delta\text{DIC}^{\text{preindustrial disequilibrium}}$ relies on **nutrient** data for its estimate. Hence, the apparently better prediction of C_{ant} compared to DIC may actually result from partial compensation among the biases in the different input variables — that is, the negative biases in DIC and A_T corresponding to approximately 0.1% of their respective mean values, and the positive biases in nutrients corresponding to 0.3–0.8% of their respective mean values — across the ϕ_{CT}^O method calculations, such that their combined effects lead

to a smaller resulting bias in C_{ant} . The obtained RMSE C_{ant} value of $3.3 \mu\text{mol kg}^{-1}$ provides a measure of the magnitude of the uncertainty associated solely with the NNs and can therefore be used as a reference for assessing whether differences between two NN-based C_{ant} estimates are sufficiently large to be distinguished from the NN error.

We kindly refer the reviewer to the newly added Section 2.4 of the revised manuscript to see how the results described above have been incorporated.

The introduction of the main text sounds too much like the previous Asselot et al. 2024 paper, with a few updated numbers and similar wording (e.g. “Since the beginning of the industrial revolution, human activities have emitted large amounts of carbon dioxide (CO_2) in the atmosphere via fossil fuel burning” vs. “Since the beginning of the industrial revolution, human activities such as fossil fuel burning and changes in land-use have led to the emission of large amounts of carbon dioxide (CO_2) into the atmosphere” in Asselot et al. 2024). Later in the introduction there are more phrases that closely resemble Asselot et al. 2024, for example “finest temporal resolution to date” in L127 (“finest spatio-temporal resolution to date” in Asselot et al. 2024). It would be good to rewrite it and check it carefully. Better yet, include some explanation what you are doing differently. This is only done much later, in L423. For the reader it would be better to make this clear early.

We have rephrased several sentences in the introduction. We refer the reviewer to our response to the specific minor comments for details on how we rephrased the introduction.

Section 2.1 lacks detail on how data was chosen and why those choices have been made. Why was “probably good” data included, given that ARGO floats already suffer from reduced data quality compared to ship-based data?

The use of probably good Argo data is what we did in Asselot et al. (2024) and is a common practice among Argo users. See for instance : Sharp et al. (2023); Kolodziejczyk et al. (2024), Johnson and Purkey (2024).

We clarified the text as follow: “We retained data identified as “good” in the Argo data stream (quality control (QC) flag =1 or 2) and excluded data that had not undergone quality control (QC flag =0), data flagged as “bad” (QC flag=3 or 4), or those that were missing, changed or interpolated (QC flag=5, 8 and 9) (Wong et al., 2025).”

Sharp, J. D., Fassbender, A. J., Carter, B. R., Johnson, G. C., Schultz, C., and Dunne, J. P.: GOBAI- O_2 : temporally and spatially resolved fields of ocean interior dissolved oxygen over nearly 2 decades, *Earth Syst. Sci. Data*, 15, 4481–4518, <https://doi.org/10.5194/essd-15-4481-2023>, 2023.

Kolodziejczyk Nicolas, Portela Rodriguez Esther, Thierry Virginie, Prigent Annaig (2024). ISASO2 : Recent trends and regional patterns of Ocean Dissolved Oxygen change . *Earth System Science Data* , 16(11), 5191-5206 . Publisher's official version : <https://doi.org/10.5194/essd-16-5191-2024> ,

Johnson, G. C., & Purkey, S. G. (2024). Refined estimates of global ocean deep and abyssal decadal warming trends. *Geophysical Research Letters*, 51, e2024GL111229. <https://doi.org/10.1029/2024GL111229>

Why was the data last accessed in 2021?

Data was last accessed in December 2023 (there was a typo in the text).

Overall, this seems to be an update of Asselot et al. 2024 to include more data points, but if that's the case it would be good to include more recent data – or at least explain why this wasn't done.

When we accessed the data in 2023, the most recent year for which complete annual coverage (all months) of the delayed-mode profiles was available was 2021. Furthermore, 2011 is considered the first year with sufficient spatial coverage of Argo-O₂ floats. Our dataset thus covers a ten-year cycle. We acknowledge that extending the time series would at this time be a feasible option, given the more recent availability of DM Argo and Deep Argo data. However, in full transparency, the postdoctoral project that motivated this work has now ended, which limits our capability to do it. We believe that this one-decade record contains a robust and scientifically valuable signal that warrants publication in its own right, and that presenting the dataset in its entirety provides a coherent assessment of the basin scale C_{ant} inventory variability over this period.

Why was the old GLODAPv2.2020 used and not a newer version (GLODAPv2.2021, GLODAPv2.2023 or more recently GLODAPv3), where more cruises are included and more corrections applied? Wouldn't it be better to retrain an updated dataset? This is also noted in Asselot et al. 2024, where they state, "Therefore they must be trained with new data as the ocean changes".

GLODAPv2.2020 was used to train the ESPER_NN neural networks; however, we did not perform this training phase, it was conducted by the authors who developed these neural networks. Retraining the neural networks with a more recent version of GLODAP would constitute a completely different study, resulting in a new version of the neural networks that would need to be evaluated, reviewed and published. Furthermore, our Argo data predominantly cover a period included in the ESPER_NN training period, and for CANYONB/CONTENT, whose training period extended only through 2013, we found no evidence of degraded neural-network performance in post-training predictions (see Section 2.4 in the revised manuscript).

Also, data accuracies are listed, but there is no discussion about whether this is acceptable for the task.

Indeed, data accuracies of the input variables (P, T, S, O₂) are considered in the final error estimation ($\pm 7.5 \mu\text{mol kg}^{-1}$ for C_{ant} concentration) using a Monte-Carlo method. This is detailed in Appendix B. As this was not clear enough in the text, we clarified it as follow (in bold the text added): "To quantify the uncertainty associated with our C_{ant} concentrations, we randomly generated 100 C_{ant} fields by perturbing P, T, S and O₂ fields, **based on their accuracies (Section 2.1)**, using a Monte Carlo method (Metropolis & Ulam, 1949). This means that for a single C_{ant} profile of a specific float, we generated an ensemble of 100 Monte Carlo realizations. This procedure, **detailed in Appendix B**, was applied to all profiles considered in this study. The mean standard deviation of the C_{ant} realizations, taken here as the overall uncertainty, was $\pm 7.5 \mu\text{mol kg}^{-1}$."

For section 2.2, I would rather that you shorten the neural network explanation and instead give more context on your workflow, i.e. the method of Asselot et al. 2024. Even if you reuse existing neural networks, the details about the neural network training are insufficient. It would be great to include some plot showing neural network performance – especially for the Irminger Sea. I believe it's beneficial to be open about where the neural network is struggling and where we have the most confidence.

We have added a description of the workflow following Asselot et al. (2024), clarifying how the Argo T, S, and O₂ observations are used sequentially to estimate the biogeochemical and carbonate variables required for the C_{ant} calculation (Section 2). We have also added information on the performance of the neural networks and their associated uncertainties, with particular attention to the North Atlantic/Irminger Sea region (Section 2.4).

Minor comments:

L20: Are the amount of Deep Argo floats, compared to normal Argo floats, sufficient to simply expand from 2000 to full depth? Have you noticed any transition (of data or uncertainties) at 2000 m due to the changing data density?

In total we have 29 Argo-O₂ floats in our database, comprising 5 Deep-Argo-O₂ floats. There is no visible transition at 2000 dbar when considering vertical C_{ant} concentration (Figure 3a). In section 2.1 we added the sentence:

“The absence of any vertical discontinuity in the basin-scale average C_{ant} concentration calculated from those data (see Section 3.1) suggests that the difference in sampling density does not impact the vertical C_{ant} distribution.”

L32: Perhaps you should make clear in the abstract that you’re using existing neural networks – some readers will only read the abstract first and may conclude that you have developed new ones. See also L155

We added the word “existing” to make it clear.

L51-53: Despite this being a plain language summary, it would be more accurate to say that C_{ant} can’t be measured directly and needs to be estimated from different methods. I know you use “studied” to create some ambiguity, but it still sounds like research ships go out and measure C_{ant} directly. Just a slight rephrasing is enough

We agree with your suggestion. However, since Biogeosciences doesn’t require a “plain language summary”, this paragraph has been removed in the revised manuscript.

L63-65: This first sentence sounds a lot like Asselot et al. 2024

We rephrased this sentence.

L67: You write about further consequences of increased C_{ant} uptake further down (from L78), perhaps that’s where the sentence about the effective radiative forcing would fit better

On L67 we talk about the consequences of C_{ant} on the atmosphere (radiative forcing) while on L78 we talk about the consequences on the ocean. We keep the two sentences separated to avoid confusion.

L69: in the most recent version this estimate rises to 29% for the Global Carbon Budget 2025 - but you should still throw in an “approximately”

We updated this number.

L76: I would specifically add “observation-based” to those products, to distinguish them from the models. Also, you can cite Friedlingstein here again as this is the main review (or “current state”) of models vs. observations and the difference between them (figure 10 in Global Carbon Budget 2025)

Changed.

L87: I would say “estimate”, not compute. That is also what was done in Asselot et al. 2024 – at the same time, you again risk sounding a lot like Asselot et al. 2024

Changed.

L89: You start with the abbreviation without introducing it first. It should be “Dissolved inorganic carbon (DIC)”. Also, perhaps it would be useful to include a bit more detail on the types of carbon in the ocean. First, only C_{ant} is talked about, later, DIC is mentioned without explanation. If you use C_{ant} , you should also make clear what C_{nat} would be and that DIC holds both. This can be done in a very short way, but would provide the reader with more context.

We added “...measured dissolved inorganic carbon (DIC, which represents the total sum of natural (C_{nat}) plus anthropogenic (C_{ant}) carbon).”

L89-91: Include a reference – especially because this wording could also be applied to the C^* method (Gruber et al. 1996)

References added.

L99: there are more, e.g. the eMLR(C^*) method of Clement and Gruber 2018

The reference is added.

L110-113: C_{ant} is not measured by ships directly, it's been estimated using data from these time series

We replaced “documented using” by “indirectly estimated from”.

L122: You should make clear that this is what you're attempting here – this sounds like getting C_{ant} from neural networks is the default method

We rephrased by “Argo-based C_{ant} estimates rely on existing neural networks to reconstruct the biogeochemical variables required for the C_{ant} calculations from water mass characteristics...”.

L127: Sounds a lot like “finest spatio-temporal resolution to date” in Asselot et al. 2024

We rephrased by “...Here, we extend this approach to all available Argo-O2 profiles in the Irminger Sea from 2011 to 2021. This expanded dataset provides the basis for reconstructing the first monthly time series of C_{ant} in the basin, allowing us to estimate monthly C_{ant} inventories and investigate potential changes in C_{ant} storage rates over time.”

L133: Why was the last access in 2021? Wouldn't it be helpful to include more data in the training?

This was a typo, data was last accessed and downloaded in December 2023.

If the reviewer is referring to the neural network training, we would like to clarify that the training phase was not performed as part of this study. The training phase is performed during the development of the neural networks. We used previously developed, trained, and validated neural networks to estimate the required biogeochemical variables. This is now clarified in Section 2.2.

L133-134: Why is it relevant to say that real-time profiles have been replaced by delayed mode? This sounds like you were using real-time profiles at some point. Why? Wouldn't it be better to have more QC prior to neural network estimation?

We agree that this sentence was confusing. We replaced it by “We only used data adjusted in delayed mode. We retained data identified as “good” in the Argo data stream (quality control (QC) flag =1 or 2) and excluded data that had not undergone quality control (QCflag =0), data flagged as “bad” (QC flag=3 or 4), or those that were missing, changed or interpolated (QC flag=5, 8 and 9) (Wong et al., 2025).

L136: Why do you include “probably good” QC flags?

See answer above and to the previous related comment.

L141: Not necessary to repeat the choice of years again

Removed.

L155: It would be good to highlight in a sentence explaining that you’re using “existing neural networks (i.e., ESPER_NN, CANYON-B and CONTENT)” - like in Asselot et al. 2024, where the language around using existing algorithms is much more clear. If a reader hasn’t read Asselot et al. 2024 first, it initially seems like you are applying new neural networks.

We replaced by “existing neural networks”.

L165: Why was GLODAPv2.2020 chosen? There are 3 more recent versions than this available. Also, was the complete, global dataset used, or just a subset?

We apologize for the misunderstanding and realize that the wording may have led to confusion. We did not perform the training phase for either of the neural networks. Instead, we used the existing neural networks, which had already been trained and validated by the teams that developed them. For ESPER_NN in particular, the authors conducted the training with GLODAPv2.2020 dataset and they used the global dataset. To avoid any ambiguity, we have revised the corresponding sentences in Section 2.2.

L167-169: It is not immediately clear where those uncertainties apply to, how they were calculated and what they mean for your own estimates. Refer to the supplementary, if that’s where they come from.

We agree and thank the reviewer for this suggestion. We added a reference to the paper describing ESPER_NN and reporting these uncertainties (Carter et al., 2021). These uncertainties, which represent the error relative to actual measurements, are not directly incorporated into our uncertainty estimate. Instead, we performed a Monte Carlo analysis to quantify the overall uncertainty associated with our method. However, new section 2.4 *Evaluation of neural network predictions against observations* provides the necessary context for understanding what these uncertainty values represent and account for.

Carter, B. R., Bittig, H. C., Fassbender, A. J., Sharp, J. D., Takeshita, Y., Xu, Y., Álvarez, M., Wanninkhof, R., Feely, R. A., & Barbero, L. (2021). New and updated global empirical seawater property estimation routines. *Limnology and Oceanography: Methods*, 19(12), 785–809. <https://doi.org/10.1002/lom3.10461>

L168: A bit more context on the workflow would be good. Reading Asselot et al. 2024 first lets you follow everything much better, but if a reader sees only this manuscript it is confusing why DIC and alkalinity is thrown in here

We rephrased by “carbonate variables needed to compute C_{ant} (i.e. total alkalinity (AT) and DIC)”.

L193: Specify where in supporting information

We specified Appendix B.

L240-241: It would be helpful for the reader to explain a bit what this does, instead of just citing the associated manuscript

The section on the scaling method and the “atmospheric perturbation” concept (now Section 2.7) has been revised and clarified.

L307-308: Could this be an artifact from your method?

We modified the sentence as follows: “Over this 4.5-year period, the C_{ant} storage rate exhibits a sign reversal around mid- to late 2019 (Figure 4c) suggesting a transition from increasing accumulation rate prior to this point to decreasing accumulation rate thereafter. This change should be considered with caution considering the uncertainty on C_{ant} inventory and deserves further investigations.”

L325: Caption refers to Fröb et al. 2018, but figure refers to Fröb et al. 2016

We changed the figure legend to Fröb et al. (2018).

L339: While the higher spatio-temporal resolution is an advantage, I would not necessarily say that this is more robust – after all, you are working more with neural network-generated estimates compared to ship-based data (moreover: estimates based partly on estimates)

We agree that the higher spatio-temporal resolution does not necessarily imply greater robustness of the underlying C_{ant} inventory estimates themselves, especially considering the use of neural network-generated estimates for some of the biogeochemical variables. By “robust,” we intended to refer specifically to the statistical robustness of the estimated trend. To clarify this distinction, we have rephrased the sentence in the manuscript as follows : “..., the high spatio-temporal coverage of Argo- O_2 data allows for a more statistically constrained estimate of the C_{ant} storage rate”.

L371-373: Yes, they align closely, but can you still say something about where they don't match? What may create differences? Is there anything your method has been lacking? I'm not doubting these curves, I just wish that there'd be more of a discussion regarding neural network performance and limitations

As mentioned here above, we have included in the revised manuscript a dedicated section 2.4 *Evaluation of neural network predictions against observations* where we assess in more detail neural network performance and expand the discussion of the methodological limitations and uncertainties associated with neural networks. We have also added a paragraph in the discussion section on the limitations of the neural networks.

We have also modified the description of the results in Section 3 to highlight any discrepancies between Argo-based and OVIDE-based C_{ant} estimates, and to relate them, where relevant, to the performance of the neural network or the sampling strategy.

L384: How do the values compare if you do the same?

If we remove the first 100 m to compare our OVIDE-based estimates with the ones of Fröb et al. (2018), we obtain a difference of 14.0 mol/m^2 . We added a sentence in the revised manuscript.

L423-424: These aims would've been good to state at the beginning

Following both this minor comment and the major one, the last paragraph of the introduction (old lines 124-127, now 87-96) was modified to better account for how the present analysis relates to, and

extends upon, the results of Asselot et al. (2024): “Taking advantage of the full spatio-temporal coverage provided by the Argo-O₂ array and the sustained efforts since the early 2010s to maintain the network in the subpolar North Atlantic gyre (Figure 1), this study investigates the evolution of C_{ant} in the Irminger Sea over the 2011–2021 period. Asselot et al. (2024) demonstrated the feasibility of estimating C_{ant} from Argo-O₂ observations using a neural network-based back-calculation method. Their study used three floats drifting in the North Atlantic subpolar gyre. Here, we extend this approach to all available Argo-O₂ profiles in the Irminger Sea from 2011 to 2021. This expanded dataset provides the basis for reconstructing the first monthly time series of C_{ant} in the basin, allowing us to estimate monthly C_{ant} inventories and investigate potential changes in C_{ant} storage rates over time.”

L450-451: How do you explain it? Again, could this be an artifact from your method? Was there anything suspicious in the individual components (e.g. the estimates from ESPER-NN and CONTENT?)

We are now more cautious in the text about the reversal in C_{ant} storage rate observed in 2019 in the deep layers of the Irminger Sea (see above). We modified the discussion accordingly.

L465-475: This paragraph makes sense, and I can see the advantages – but it sounds like this belongs more in the original Asselot et al. 2024 paper. You are not creating a new method here

We agree that we are not creating a new method here. This paragraph is just to show the advantages of using Argo floats (and neural networks, which have their limitations) compared to ship-based measurements. This paragraph indicates that Argo floats and ship-based measurements are complementary. It can be seen as a reminder for readers that haven't read Asselot et al. (2024).

Supporting information

L34 This is partly repeated from the main text. More importantly, the main text states “accuracy of Argo data is 0.02°C for temperature” but in the supplementary it's 0.002°C. Which is true?

The value 0.002°C is the right one. The main text has been changed.

L70-71 “The standard deviation fluctuates between $\pm 2.7 \mu\text{mol kg}^{-1}$ and $\pm 16.8 \mu\text{mol kg}^{-1}$ ” it would be useful to know where the values have been higher/lower

The value of $16.8 \mu\text{mol kg}^{-1}$ represents a local peak that doesn't bring much information. Consequently we revised our sentences to give more meaningful values: “The overall mean standard deviation is $7.5 \mu\text{mol kg}^{-1}$. The uncertainty is relatively homogeneous with depth, with mean standard deviations of $7.1 \mu\text{mol kg}^{-1}$, $7.6 \mu\text{mol kg}^{-1}$ and $7.8 \mu\text{mol kg}^{-1}$ between 0 and 500 dbar, 500 and 2000 dbar and below 2000 dbar, respectively. The maximum standard deviation of $8.8 \mu\text{mol kg}^{-1}$ occurs at approximately 2750 dbar. These results indicate that the propagated uncertainty associated with the Argo-O₂ input variables remains relatively consistent throughout the water column, with no pronounced increase toward the surface or at intermediate depths.”