We thank the reviewer for their thoughtful comments. Here we offer detailed responses to all questions. Reviewer's comments are in black, our replies are in blue.

Responses to Reviewer #1 comments:

General comments:

Despite the increasing significance of BGC models, the model validation is limited to the comparison with satellite estimates of surface properties, the climatological data, and/or sparse in-situ observations. In recent years, the fast growing BGC-Argo network provides opportunities to evaluate BGC models in an unprecedented spatial and temporal resolutions. Since there is a large number of floats at the global scale, it becomes difficult to evaluate the global model through the point-to-point comparison which has been used in the regional model. This study suggests some BGC-Argo-based metrics to evaluate a global model and provides some diagnostic plots to display these metrics. This manuscript is well structured and easy to follow. I would suggest to publish after minor revision.

REPLY: Thanks for the positive assessment of our work.

Specific comments:

P5 Line 18-23: The BGC-Argo-based POC concentrations were obtained from the filtered bbp signals and therefore should be the small, slow-sinking POC (i.e., 0.2-20μm) (Dall'Olmo and Mork, 2014; Lacour et al., 2019). In the following model description (P6 Line 15-16), the authors mentioned that their POC model had two size classes. Which modelled POC class was compared with the BGC-Argo based one?

REPLY: In the PISCES model, POC corresponds to the sum of the two size classes of detritus, phytoplankton and zooplankton (Levy et al. 2013). Based on the reviewer's suggestion, we will compare the smallest sizes of simulated detritus and phytoplankton to match the small and slow-sinking POC observed by the BGC-Argo floats.

Based on Roesler et al. (2017), the BGC-Argo based chlorophyll were suggested to be divided by a factor of 2 due to the systematic error in fluorometers. It seems that the authors did not apply this correction to the chlorophyll. If not, this can partially explain the model underestimation of surface chlorophyll in the high-chlorophyll regions (please see the Figure 4). The authors should include some descriptions on how they process the BGC-Argo based chlorophyll.

REPLY: The gain adjustment of 0.5 is already implemented in the “adjusted“ chlorophyll data (Bittig et al., 2019). We have not applied any processing to the BGC-Argo data apart from those already applied at the Data Assembly Center levels as described in the given references. We will clarify this point in the revised version of the manuscript.

P5 Line 25-30: I am concerned with the comparisons between the global model and the estimates from CANYON-B neural network which is also a model. Although it has been validated with some independent observations (e.g. the GO-SHIP cruise data and BGC-Argo floats), differences between the global model and the CANYON-B neural network may come from the CANYON-B neural network’s deviation from the observations.

REPLY: We agree with the reviewer that we do not provide justification for mixing together BGC-Argo data with CANYON-B estimates. We propose to add the following paragraph in the Data section to justify our choice.

“ *Finally, we complemented the existing BGC-Argo dataset with pseudo-observations of* NO_3*,* PO_4 *, Si, and DIC concentrations as well pH and* pCO_2 *using the CANYON-B neural network* (Bittig et al., 2018)*. CANYON-B estimates vertical profiles of nutrients as well as the carbonate system variables from concomitant measurements of floats pressure, temperature, salinity and* O_2 *qualified in “Delayed “mode together with the associated geolocalization and date of sampling. The CANYON-B estimates of* NO_3 *and pH were merged with measured values on the rationale that CANYON-B estimates have RMS errors (NO3 = 0.7 μmol/kg , pH= 0.013) (Bittig et al., 2018) which are of the same order of magnitude than the BGC-Argo observations RMS errors (NO3 = 0.5 μmol/kg , pH= 0.07) (Mignot et al., 2019; Johnson et al., 2017). We also verified that RMS errors of CANYON-B estimates are at least 4 times lower than the RMS difference between the model and CANYON-B estimates, so that the comparison of simulated properties with the neural network estimates leads to an evaluation of the model performance. We believe it is reasonable to draw conclusions on the model uncertainty from CANYON-B estimates as long as the pseudo-observations errors are much lower than the model-pseudo observations RMS difference. However, caution should be considered when errors are comparable.*”

P8 Line 1-2: Since the POC concentrations vary a lot (~ 2 orders of magnitude) within the mesopelagic zone, the averaged POC concentrations will be skewed to the upper layers right below the mixed layer. In addition, the reference is not appropriate here since the upper bound of mesopelagic zone was defined as the base of productive layer (the maximum of mixed layer and the euphotic zone) in Dall'Olmo and Mork (2014).

REPLY: We agree with the reviewer. We used a log10-transformation to represent the data to account for the skewness in this layer. We will also add in the revised version that our definition of mesopelagic POC differs from the Dall'Olmo and Mork (2014) reference.

P8 Line 13: What is the definition of H? I guess it is the mixed layer depth (MLD).

REPLY: It is an omission on our part. H is the mixed layer depth. We will replace H by MLD.

P13 Line 26-29: I don't agree with this sentence "However, this seems to have a limited effect on the export of POC …". First, the conclusion here is anti-intuitive because the authors mixed up the POC concentration with the POC export flux. In the north Atlantic, the POC concentration is largely determined by the small, slow-sinking POC. Although the relative contributions of small, slow-sinking POC has been recently addressed, the POC export flux is dominated by the gravitational sinking flux of large, fast-sinking POC, which is estimated by multiplying the concentration and sinking velocity. Therefore, the large differences in the POC export flux can be hide by the similar POC concentrations. Second, the lower sPOC but the similar levels of POCmeso (Figure 6) can be a result of the suboptimal parameter values, e.g. the underestimated remineralization rate. However, this cannot deny the sensitivity of POCmeso to the primary productivity. This is very likely that the POCmeso will vary a lot if the authors change the modeled primary productivity.

REPLY: We agree with the reviewer that this paragraph bears lots of assumptions. We will revise this paragraph and we will remove the conclusions about the POC export flux. As pointed out by the reviewer, we don't have sufficient information to assess the skill of the model in simulating the export of POC from sPOC and POCmeso.

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 CO2 variables and nutrient concentrations from T, S, and O2 data using Bayesian neural networks, Front. Mar. Sci., 5, 328, 2018.
Bittig, H. C., Maurer, T. L., Plant, J. N., Wong, A. P., Schmechtig, C., Claustre, H., Trull, T. W., Udaya Bhaskar, T. V. S., Boss, E., and Dall'Olmo, G.: A BGC-Argo guide: Planning, deployment, data handling and usage, Front. Mar. Sci., 6, 502, 2019.
Dall'Olmo, G. and Mork, K. A.: Carbon export by small particles in the Norwegian Sea, Geophys. Res. Lett., 41, 2921–2927, https://doi.org/10.1002/2014GL059244, 2014.
Johnson, Plant, J. N., Coletti, L. J., Jannasch, H. W., Sakamoto, C. M., Riser, S. C., Swift, D. D.,

Williams, N. L., Boss, E., Haëntjens, N., Talley, L. D., and Sarmiento, J. L.: Biogeochemical sensor performance in the SOCCOM profiling float array: SOCCOM BIOGEOCHEMICAL SENSOR PERFORMANCE, J. Geophys. Res. Oceans, 122, 6416–6436, https://doi.org/10.1002/2017JC012838, 2017.

Mignot, A., D’Ortenzio, F., Taillandier, V., Cossarini, G., and Salon, S.: Quantifying Observational Errors in Biogeochemical‐Argo Oxygen, Nitrate, and Chlorophyll *a* Concentrations, Geophys. Res. Lett., 46, 4330–4337, https://doi.org/10.1029/2018GL080541, 2019.