**Responses to Reviewer 4**

Below the complete reviewer comments are shown in black font along with detailed responses to each comment in blue font.

Review:

The study is investigated the ability of available satellite information on ocean surface physical and biological properties to constrain and improve simulated subsurface biogeochemistry in the Gulf of Mexico. The study also shows an example of using complementary Argo data. In this respect the paper nicely suits the frame of the journal. Generally, the manuscript is well structured and clearly written. The figures are of a good quality. I have got just minor comments (please see below) the authors might still want to consider before publishing the manuscript.

Response: We thank the reviewer for the constructive comments and suggestions which will be very helpful as we revise the manuscript.

Title: an edit is required "…assimilation of satellite physical and biological observations …". It is a bit confusing: the title emphasizing the use/role of satellite information in improving subsurface biogeochemistry, however Argo data are also assimilated. Could the title reflect the use of Argo data?

Response: We feel the title is appropriate as is and would like to keep it. Although we also assimilate Argo profiles in DAargo run, this is only core-Argo data (i.e. temperature and salinity) and not BGC Argo data. The key point of this study is to evaluate the impact of assimilating satellite observations on the subsurface distributions, especially of biological properties. Moreover, the additional improvements obtained by assimilating Argo profiles are limited.

Specific comments:

Line 15:16. reads as BGC-Argo data are also assimilated complementary to the satellite data. Reads a bit contradictory to the title or vice versa the title reads contradictory to the statement.

Response: The satellite data are assimilated into the coupled model and the BGC-Argo data are used to calibrate the biological model. We will revise this sentence into:

*'This study investigates to what degree the assimilation of satellite observations in combination with a prior model calibration by sparse BGC Argo profiles can improve subsurface biogeochemical properties'*

P1, L18: "… into a three-dimensional biogeochemical model …"

Response: We will revise it as suggested. As also suggested by the reviewer 2, we will rephrase the lines 16-21 in our original manuscript into:

*"…… The multivariate Deterministic Ensemble Kalman Filter (DEnKF) has been implemented to assimilate physical and biological observations into a three-dimensional coupled physical-biogeochemical model, of which the biogeochemical component has been calibrated by the BGC-Argo floats data for the Gulf of Mexico. Specifically, observations of sea surface height, sea surface temperature, and surface chlorophyll were assimilated, and profiles of both physical and biological variables were updated based on the surface information……"*

P2, L30-31: How was the tuning done? To a certain extent it is still a kind of assimilation of the information.

Response: We agree. From a broad perspective, any practice to constrain a model by observations can be referred to as data assimilation. In this study, when we say data assimilation, we specifically refer to state estimation. We will state this clearly in the revied manuscript. Available BGC Argo float data are still insufficient for three-dimensional state estimation but can be used for *a prior* tuning of model. As a result, the well-tuned model can reproduce the key processes, e.g. the feedbacks between chlorophyll and light intensity in our study, and maintain the improvement obtained in update steps. The alternative light parameterization was tuned subjectively by BGC-Argo floats. However, the *a priori* tuning can be also done in other ways, such as parameter optimization.

P2, L41: “discretization and numerical schemes” instead of “discretion”
Response: We will revise it as suggested.

P2, L45: “.” is missing in the end of the sentence.
Response: We will add this as suggested.

P2, L48: suggest to add “(e.g. Chla)” after “satellite data of ocean colour have been the major source of observations”
Response: We will revise it as suggested.

P2, L51: correct reference is Pradhan et al., 2020
Response: Thank you for pointing out this. We will correct it as suggested.

P4, L85-86: are the mentioned five BGC-Argo floats really independent if used for the model calibration (model optimisation even though by "trails-and-error")?
Response: We understand the reviewer’s concern that the BGC-Argo float data are not fully independent because they have been used to optimize the biological parameters in Wang et al. (2020). Here the ‘independent observations’ refers to the unassimilated observations. We believe this is consistent with the common understanding in the data assimilation literature and therefore we will keep it as it is but we will point out clearly that the data has been used in prior tuning in the revised manuscript.

P5, L31: “observational operator” instead of “measurement operator”
Response: We will revise it as suggested.

P6, L53-54: The specified (assumed) observational errors for SSH and SST are quite small, which could lead to overfit to the data with possible deterioration of the state for other model variables.
Response: The observational errors of SSH and SST are based on previous studies (Song et al., 2016; Yu et al., 2018, 2019) and have been applied successfully in Yu et al. (2019). With respect to the overfitting issue, we compared model results below the surface with unassimilated Argo and BGC-Argo profiles. The subsurface temperature, salinity, and NO3 were largely improved. Other biological variables were also improved although the improvements were limited. These results give us some confidence that the observational errors are appropriate.

P7, L79-80: Inflation is normally introduced to account for uncertainties in approximation of model error (due discrepancies in the forcing or internal model parameter/parameterisations), which consequently alters the ensemble spread.
Response: We agree. We will revise it into:
*‘In addition, ensemble anomalies are inflated by 1.05 at each update step to account for the unrepresented model uncertainties (Anderson and Anderson, 1999).’*

P7, L89-91: It would be nice to provide a reference to a study on model sensitivity to these particular parameters? A motivation and a reference to a procedure of parameter perturbation would support. Please also make it clearer whether the parameters are perturbed just to introduce more stochasticity to the system (e.g. Pradhan et a. 2019, 2020) or the data assimilation experiment considers also parameter estimation (Doron et al., 2011, Simon et al. 2015).
Response:

The sensitive biological parameters were selected by sensitivity tests. Specifically, we ran a 1D version of this model multiple times by incrementally perturbing one parameter at a time and setting other parameters unchanged (hereafter as **Test case**). Each parameter ($p$) was perturbed by multiplying factors ranging from 0.25 to 1.75 with an increment of 0.25. Then parameters were sorted based on their sensitivity

which was quantified by the normalized absolute differences from the unperturbed run (hereafter as **Base case**):

$$Q(y,p) = \frac{1}{m}\sum_{i=1}^{m}\frac{1}{n}\sum_{j=1}^{n}\frac{|y_{Base} - y_{Test}|}{y_{Base}}$$

where *m* is the number of parameter increments (here 7) and *n* is the number of base-test pairs including all 1D model grid cells throughout the whole simulation period. Based on the sensitivity tests, the most sensitive parameters were selected and perturbed in our data assimilation experiments. The details about the 1D model can be found in the section 3.3 of Wang et al. (2020). In our revised manuscript, we will add explanation about the choice of sensitive biological parameters.

In this study, the sensitive biological parameters were perturbed to introduce model uncertainties but were not updated. We will make it clearer in the revised manuscript.

P8, L1-3: Please consider rephrasing this sentence. The length of the state vector should not affect crucially the computational cast. Normally the time required for the analysis (independent on the length of the state vector) takes much less than the computational expenses required for running the ensemble itself. It is worth providing another argument for justification of the choice of model variables included to the state vector.
Response: As suggested, we will rephrase this sentence into:
*'Although the DEnKF can update all variables based on their cross-covariance, we limit updates to two physical variables (temperature and salinity) and four biological variables (nitrate, chlorophyll, phytoplankton, and zooplankton) which are key to the coupled physical-biogeochemical system. Other variables will be adjusted by model dynamics.'*

P8, L4-12: It is worth showing both criteria RMSD and unbiased RMSD (+ additional? bias). In this case it would be clearer for the reader for which model variables the solution deviate systematically or randomly from the observations, whether the data assimilation allows to reduce bias (if any) or random part of the obtained differences between model and observations.
Response: Thank you for this suggestion. We will report the bias and correlation coefficient in our revised manuscript as suggested.

P9, L24: "This figure shows" or "This comparison shows"
Response: We will revise as suggested.

Figure 2 could be slightly increased.
Response: We will revise as suggested.

Supplement, Figure S3: a legend or more detailed figure caption is required.
Response: We will revise as suggested.

**Reference**

Song, H., Edwards, C. A., Moore, A. M. and Fiechter, J.: Data assimilation in a coupled physical-biogeochemical model of the California current system using an incremental lognormal 4-dimensional variational approach: Part 3—Assimilation in a realistic context using satellite and in situ observations, Ocean Modelling, 106, 159–172, doi:https://doi.org/10.1016/j.ocemod.2016.06.005, 2016.

Wang, B., Fennel, K., Yu, L. and Gordon, C.: Assessing the value of biogeochemical Argo profiles versus ocean color observations for biogeochemical model optimization in the Gulf of Mexico, Biogeosciences, 17(15), 4059–4074, doi:10.5194/bg-17-4059-2020, 2020.

Yu, L., Fennel, K., Bertino, L., Gharamti, M. El and Thompson, K. R.: Insights on multivariate updates of physical and biogeochemical ocean variables using an Ensemble Kalman Filter and an idealized model of upwelling, Ocean Modelling, 126, 13–28, doi:https://doi.org/10.1016/j.ocemod.2018.04.005, 2018.

Yu, L., Fennel, K., Wang, B., Laurent, A., Thompson, K. R. and Shay, L. K.: Evaluation of nonidentical versus identical twin approaches for observation impact assessments: an ensemble-Kalman-filter-based ocean assimilation application for the Gulf of Mexico, Ocean Science, 15(6), 1801–1814, doi:10.5194/os-15-1801-2019, 2019.