We thank the reviewer for their constructive feedback. We report here our responses to the specific points.

The text in *blue italics* are the comments from the referee. The text in "black regular" is our response to the comment, and the text in *red italics* is our text that appears in the revised manuscript.

Comments from the referee:

*Many of my comments from an earlier revision are resolved, and the only outstanding issue in my mind is the quantification and description of how parameter uncertainty is incorporated into the model, either now or in the future. I saw in your response that you 'are not pursuing a precise 1-to-1 match between the model and BGC-Argo, but exploring notable differences that should be improved regardless of the concentration differences'*

*When looking at the manuscript and performance metrics, it appears as if there is no uncertainty or weighting scheme involved in the calculations (equations 1-4). Incorporation of uncertainty and other parameters is a goal of this framework eventually (although not in this manuscript), I recommend the authors write more specifically and more quantitatively about how uncertainty (specific magnitudes for different variates) can be incorporated into the framework. For example Argo chl, oxygen, bbp have vastly different relative uncertainties ranging from 15% to over 200% in some cases. The tuning meteorological climatologies used also have different uncertainties. A plan for how to tune model parameters would be a nice edit in the new paragraph added (through the last revision) about uncertainty.*

We agree on the necessity of including parameter uncertainty in a study of this kind. The reviewer, however, agrees that it is beyond the context of this study to conduct further experiments on uncertainty. We see this paper as the presentation of the framework, how it is set up, and ways it can be used as a tool for model improvements. We believe we have established this aim here. That being said, while this study does not go into details about parameter fine tuning, that would indeed be the natural application of this framework where uncertainties would play a major role. For this reason, we have expanded the text on uncertainty following the reviewer's recommendations, starting with examples of BGC-Argo uncertainties of various sensors, how we approached this issue in this study (employing ADJUSTED BGC-Argo variables), and how we will approach it in the follow-up study (i.e. increase BGC-Argo tracks, employ multiple BGC-Argo variables for the tuned model parameters to better represent the environmental constraints.) We envision an ensemble suite of model runs, and together with multiple BGC-Argo tracks, we will have a very large dataset to apply the statistical analyses. And as a final step, we will refine the search for optimal parameter sets by narrowing it down for the variables with lower uncertainty and expanding it for those with higher uncertainty. And as a further follow-up study, a data-assimilation scheme can be included. All this information is added to the Concluding remarks section, and the following is the added/modified text:

Lines 514 – 552 On the topic of tuning, a planned follow-up to this study is the application of more systematic parameter tuning approaches 515 (e.g., Gharamti et al., 2017) compared to the relatively simple exercise presented here. In the future, the number of buoys that are used in the experiments should be increased, preferably establishing a region-wide sample set, and a detailed sensitivity analysis should be made for a wider parameter set. For instance, we note that the model chlorophyll *a* is on the higher end with the parameter values chosen in this study (see Section 3.3.3). Dedicated parameter tuning approaches should consider uncertainties in the BGC-Argo data as well as the uncertainty range of the tuned parameters. Because this study focus on improving the model formulation (i.e. increasing growth rates under low-light conditions), rather than model parameter fine tuning, a dedicated assessment of BGC-Argo data errors was not included. To limit the effects of observation uncertainties, we only included (see Sec. 2.1.1) the "ADJUSTED" BGC-Argo variables (i.e. temperature, salinity, pressure and chlorophyll *a*) which provide either a "real-time-adjusted" or a "delayed-mode" data control and correction (Bittig et al., 2019). Despite applying a level of correction, there are still observational errors present. For example, the measured fluorescence chlorophyll *a* to chlorophyll *a* ratio can vary due to various factors and can lead to uncertainty as high as ±300% (Roesler et al., 2017). However, some of these errors can be reduced to a maximum of 0.12 mg Chl $m^{-3}$, with an average reduction of ±40% (Johnson et al., 2017; Bittig et al., 2019). On the other hand, the BGC-Argo estimated POC uncertainty is lower and but can be as high as 40 mg C $m^{-3}$, about 50%. In the case of oxygen, the sensors show a strong drift (order –5% $year^{-1}$ between calibration and deployment), this can be corrected (to approx. 1.0 – 1.5 µmol $kg^{-1}$) with surface measurements adjustments along-the-track (Bittig et al., 2018). Similar uncertainties that exist for all other BGC-Argo variables should also be accounted for in model validation studies.

We envision an ensemble simulation approach for model biogeochemical parameter tuning as a follow-up study where we construct a suite of ensemble experiments with systematic perturbations of selected model parameters within a ± uncertainty range from the respective reference parameter value. However, depending on the number of modified parameters and BGC-Argos, the number of experiment can be in the range of thousands which raises the question of how to select the parameter set(s) that yield the best results objectively. The statistical analyses that have been performed in this study is done on a limited number of BGC-Argos and a single biogeochemical variable (i.e. chlorophyll *a*) and may turn out inconclusive for a fine-tuning parameter study, given the BGC-Argo uncertainty. Newer BGC-Argos are equipped with multiple biogeochemical sensors, making the statistical analysis of a parameter fine-tuning experiment more robust as the number of experiments increases while accounting for multiple BGC-Argo variables including their associated uncertainty ranges. The inclusion of multiple BGC-Argo variables statistics would enhance the ecosystem representation of the parameters, and multiple variables would provide more constraints toward realistic representations. At that stage of the analyses, the uncertainty range of the observed variables can be included, and the search for the better performing experiments could be narrowed down for the less uncertain variables (e.g. POC, oxygen) and widened for the more uncertain ones (e.g. chlorophyll *a*). In addition, instead of directly incorporating the concentrations of the full experiment in the statistical analyses, it may provide valuable insights to separately assess the timing of seasonal events driven by the mixed layer dynamics. Alternatively, comparing correlations between the model and the depth location of key features in the BGC-Argo profiles, such as the nutricline, would give an insight to the mixing and production dynamics (e.g., Salon et al., 2019). These approaches would reduce the influence of observation errors, but would rely on the consistency of the sensor along-the-track. Finally, a more elaborate data assimilation scheme that take into account model variable and parameter uncertainties, such as that based on Ensemble Kalman Filter could be considered to this framework in an idealized setting to investigate whether or not the current model parameterization is suitable to represent the observed real world process (e.g., Singh et al., 2022).

Lines 556 – 558 – this sentence was modified Our study demonstrate the possibility to the design and apply such approaches through considering different (1) regional coverage and (2) ever-growing time-extent of BGC-Argo

data, allowing us to investigate the model discrepancy on a large-scale but also on local scale when considering (3) high-resolution depth and time coverage.