Review of the BG Discussion paper (**bg-2020-137**) "Assessing the value of BGC Argo profiles versus ocean colour observations for biogeochemical model optimization in the Gulf of Mexico" by Bin Wang et al.

Dear Editor, I read with great interest this manuscript related to the application of an optimization procedure combining satellite and BGC-Argo floats measurements with a biogeochemical model of the Gulf of Mexico. The manuscript is well organized and provides important insight to improve model parameterizations combining state-of the-art sensors data-streams. I found particularly important the results showing the importance of combining chlorophyll and bbp measured by BGC-Argo to optimize the model in terms of POC and export. In my opinion it is also important the results that, to a large extent, the optimization procedure carried out on the 1D model can be translated to the 3D model implementation.
Therefore, I suggest the publication of this manuscript after minor revisions reported below. My comments are indented in blue.

Minor revisions

1. Introduction

Pg2 line 45. because the number of parameters increases **exponentially** with the number of state variables (Denman, 2003).

I would expect a polynomial [quadratic] not an exponential increase of the parameter number vs state variable number. In fact, Denman (2003) provides a polynomial formula for the increase of fluxes vs state variables and elaborates the estimate of increase of parameters basing on this formula.

Pg3 lines 68-70. This is especially problematic in oligotrophic regions where the maximum chlorophyll concentration (referred as the deep chlorophyll maximum, DCM) is pronounced near the base of the euphotic zone because of photo-acclimation (Cullen, 2015; Fennel and Boss, 2003).

Could Authors explain better what they mean? In general DCM can appear in mathematical terms even without photo-acclimation (e.g. Varela et al. 1994, Ryabov and Blasius 2008). I suggest Authors to be more specific on this point. Nonetheless, the effect of photo-acclimation mentioned here could produce an enhancement of the DCM feature [already explainable by mechanisms other than photo-acclimation] and could be particularly important for the specific area of the GoM.

References
Varela, R. A., A. Cruzado, and Joaquín Tintoré. "A simulation analysis of various biological and physical factors influencing the deep-chlorophyll maximum structure in oligotrophic areas." *Journal of Marine Systems* 5.2 (1994): 143-157.

Ryabov, Alexei B., and Bernd Blasius. "Population growth and persistence in a heterogeneous environment: the role of diffusion and advection." *Mathematical Modelling of Natural Phenomena*3.3 (2008): 42-86.

2. Study Region

I would suggest Authors to add information about the recirculation times of the GoM waters with

respect to the Atlantic Ocean boundary conditions (BC), this would give an idea on the relevance of the BC in the experiments.
Moreover it would be interesting to know what are the observed dominant plankton species, or plankton functional types, in the open-ocean part of GoM [pico-phytoplankton and flagellates?]. This could be useful to figure out the implications of the choice of considering one phytoplankton in the model when compared to data.

3. Methods
3.1. Biological observations

Pg 5 lines21-22 SeaWiFS (Sea-viewing Wide Field-of-view Sensor), MODIS (Moderate-resolution Imaging Spectroradiometer), and MERIS (medium-spectral resolution imaging spectrometer) products.

What about VIIRS? Is it included in the database?, please check.

Pg 5 line 26 and particulate backscattering

Pg 5 lines 31-32 Satellite estimates were therefore corrected following the regression equation shown in Figure S2a (Figure S1c).

Pg 6 lines 45-48 The bbp700 from the floats is weakly correlated with the satellite estimates (R_2=0.11) and generally lower by a factor of ~0.45 than the satellite estimates (Figure S2b). The bbp700 profiles were therefore multiplied by 2.2 before being converted to bbp470 following the equ. 1.

In one case Authors consider the BGC-Argo the ground truth (Pg 5 lines 31-32) in the other Authors correct BGC-Argo with respect to satellite data (Pg 6 lines 45-48). The procedure seems a bit circular, could Authors explain better this part?

Pg 6 line 44 extent

3.2. 3D Model description

Pg 7 lines 74-75 Medium-Range Weather Forecast ERA-Interim product with a horizontal resolution of 0.125^{o} (ECMWF reanalysis, https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era-interim).

The resolution reported in the link above for ERA-Interim is 80 km this does not match with the 0.125^{o} (~12 km) reported in the manuscript, could Authors double check?

3.3. 1D Model description

Pg8 line 10-12. The 1D model, which is similar to that used by Lagman et al. (2014) and Kuhn et al. (2015), covers the upper 200 m of the ocean with a vertical resolution of 5 m and is configured at one location in the central Gulf (see Figure 1).

Authors should add some comment about the choice of 5 m vertical resolution. What is the 3D model vertical resolution in the same region of the GoM? Why not taking 1m resolution that probably is the vertical resolution of data acquired by BGC-Argo floats?

Pg9 line 13-15. A higher diffusion 213 coefficient (K_{Z1} =max(H^2_{MLD} /400,10))is applied in the turbulent surface layer and a lower diffusion coefficient (K_{Z2} =K_{Z1}/2) is assigned to the quiescent bottom layer.

What is the unit of measure of $K_{1,2}$, please add this information in the text. If it is m2/s it seems very high, because even with H_{MLD} = 0 → K_{Z1} =10 → K_{Z2} =5. Please, add also unit of measure of H_{MLD}, meters?

3.4. Parameter optimization method

Pg 11 line 63. and n is the number of base-test pairs including

I suggest to use consisting rather than including that in my opinion is confusing. In fact *n* accounts exactly to what specified (i.e. base –test pairs etc etc)_and not more.

The range of variability of Wphy spans many orders of magnitude and it is the most sensitive parameter, could Author comment on that?

3.5. Parameter optimization experiments

Pg 11 line 69-71. For the parameter optimization of the 1D model, satellite chlorophyll within a 3´3 pixel (12 km´12 km area) around the 1D station and climatological monthly averages of the profiles from the bio-optical floats were used.

Did you considered all the BGC-Argo data available or only the ones near the virtual mooring of your 1D experiment? How you decided the BGC-Argo float to include in the optimization procedure?

Pg 11 line 81-84. Prior tests have shown that the available observations cannot simultaneously constrain the sinking rates of small and large detritus (w_{SDet} and w_{LDet})Therefore, a constant ratio of 0.1 between these two parameters (w_{SDet} =0.1 w_{LDet}) was imposed and only one of the two was optimized.

I suggest Authors to be clearer: I cannot get why the two parameters cannot be constrained. In general, after you completed your optimization procedure you could perturb the parameter *0.1* (ratio between WsDet and WlDet) to see if it corresponds to a minimum for your metrics F(p) or if there are better values other than 0.1. Or the system is unstable if you don't take *0.1*? Please explain.

Pg 12 line 92 please remove *and* from the equation.

4. Optimization of 1D models

General question: how you define the DCM depth in the presented analyses?

4.1. Observations and base case

Pg 12. Lines 107-109. Unlike the surface chlorophyll, the vertically integrated chlorophyll as well as the phytoplankton and POC over the upper 200 m tend to be more constant with much less seasonality (Figure 3b-d).

This statement refers to observation? Please specify. Is it possible to add the error bar to the dots of Figure 3 as in the case of Figure 4?

Pg 12 line 14 in June and gradual shoaling after July (Figure 3e), reflecting the seasonality of the solar radiation.

With the term *reflecting* Authors mean that there is direct causality or correlation?

Pg 13 line 18 However, it fails to reproduce the deepening of the DCM in June

This deepening is related to a physical process (change in some environmental regulating factor) or to a biogeochemical process?

4.2. Results of the optimizations

4.2.1 Model-data misfits

4.2.2 Experiment A

Pg 13 line 41-42 The optimal parameter sets (A4, B2, and C4), which are selected based on case-specific misfit from these three groups, will be used in subsequent analyses …

Why Authors decide to use B2 rather than B4 that show a smaller total misfit? Is it better to take a realization with better case-specific misfit or better total misfit?

4.2.3 Experiment B

Pg 14 line 47-49 However, the vertical structure of chlorophyll deteriorates relative to the base case (Figure 4a) because of inappropriate estimates of the initial slope (α=0.0101; see table 2) and the maximum ratio of chlorophyll to carbon (θ_{max}=0.0191; see table 2).

If Authors can judge a-priori that the values for α and θ_{max} resulting from the optimization are not appropriate, why they didn't consider different parameter ranges in the optimization procedure from the beginning, excluding bad values?

Pg 14 lines 57-59. A straightforward interpretation is that the addition of subsurface observations reduces the model's degrees of freedom to fit one single observation type (surface chlorophyll).

Does this imply that a model with more parameters (e.g. more phytoplankton species) would fit better?

Pg15 line 70 In contrast to the observations where detritus dominates POC …

What observation? Please add reference.

4.2.4 Experiment C

Pg 15 lines 76-78 As shown in Figure 4a, the annually averaged depth of DCM of 80 m coincides with the observed DCM, primarily because experiment reproduces the deepening of the DCM in summer.

Interesting. Can Authors explain if there is a specific parameter/mechanism that controls this dynamical deepening of the DCM? Or it is a complex combination of parameters values generating this emergent property?

5. 3D biogeochemical model

Could Author explain better the manual correction described at pg 17 lines 24-25? If they consider the corrected values more realistic why they didn't narrowed the parameter variability range in the optimization experiment? Or the 1D vs 3D implementations do not allow this?
For example, is there a simple explanation for the need to set K_{NH4} to 0.01 in the 3D experiments? It would be useful for readers interested in applying this methodology in other areas.

It would be useful to know how the PP from the model is computed: integrating down till the bottom, considering thee MLD?

6. Discussion
6.1 Trade-offs between different observations for parameter optimization

In this section Authors use a number of times the following terms *poorly constrained*, *weakly constrained* and *unconstrained, un-optimized fields, not optimized but well defined*. Some definitions can be grasped from section 3.5. In my opinion it would make things more simple to have the formal definition of these terms and to know if, in same cases, they are equivalent/synonym.

Pg 21 lines 36-42 Although this cross-validation at different times and locations may give some indication of overfitting, it cannot determine whether the model reproduces observation through wrong mechanisms because a small misfit of cross-validation can be caused by missing validations of key variables or fluxes, e.g. ignorance of phytoplankton and PP in the experiment B, while a large misfit can be a result of the intrinsic heterogeneity of biological parameters in different times (Mattern et al., 2012) and locations (Kidston et al., 2011), e.g. underestimation of coastal surface chlorophyll in the experiment C.

In my opinion the sentence above is not very easy to follow, could Authors simplify?

Pg21 lines 54-55 On the other hand, counter examples exist where the 3D simulations outperform the 1D model (Hoshiba et al., 2018).

Could Authors explain better this sentence? Outperform with respect to what aspect?