

## RC2: 'Comment on bg-2023-48', Anonymous Referee #2, 05 Apr 2023

Citation: <https://doi.org/10.5194/bg-2023-48-RC2>

The authors couple a BGC-optical model with a 1-dimensional physical ocean model for a site in the NW Mediterranean Sea, with the goal of better resolving dynamics of CDOM. It is nice to see an emphasis on modeling CDOM in a coupled physical-BGC model. However, some work is required to improve the manuscript and make it more accessible to readers.

We appreciate the constructive comments and suggestions from the Reviewer. We present our point-by-point responses to the Reviewer's comments below. The Reviewer's comments are in black, our responses follow each comment in green. In each response, we detail the changes we propose to make to the manuscript in order to address these comments, and include the proposed modified text and/or figure.

### general comments

I have a couple of methodological questions concerning the optimization procedure. Firstly, the authors declare the estimated 25 parameters "optimal" without too much evidence to back it up. Interdependence between parameters (which the authors mention in the manuscript) and other factors can lead to complex cost functions with local minima. Based on the manuscript and appendix A, it appears like the one large optimization experiment was conducted, using a large number of iterations. I would encourage performing a few more experiments starting from different initial parameter values, in order to assess the new parameter estimates are similar to the initial ones. Here it could be enough to perform fewer iterations and examine the convergence. If this type of experiment was already performed, this information should be included in the manuscript so that readers (like me) are not left wondering if the reported results are based on parameter values from a local minimum.

Differential Evolution is a global optimization method that searches for the absolute extremum of the function over its entire definition domain. The risk of becoming trapped in a local minimum is comparatively lower compared to direct search approaches. As a parallel search technique, it has some built-in safeguards to forestall misconvergence. For example, by running several vectors simultaneously, superior parameter configurations can help other vectors escape local minima. In fact, DE shows consistent convergence to the global minimum in consecutive independent trials. A complete evaluation about the resilience of the DE algorithm to fail into local minima is given in Storm and Price 1997. We are also following the recommendations of this publication for the values of the metaparameters NP, F and CR.

We will add the following statement in Section 2.4.2 'Parameter optimization method':

'Genetic algorithms are particularly robust in overcoming local minima (Storn and Price, 1997). As a parallel search technique that runs several vectors simultaneously, DE sampling method can help vectors escape local minima and therefore avoid misconvergence.'

Nevertheless, we do agree with the reviewer that it would be interesting to see different trials of this optimization experiment. We disagree in the fact that only the iteration over a few generations will be necessary. From our point of view, this analysis would require to repeat the whole optimization experiment completely several times. Otherwise, we do not think the results will be meaningful. These new set of experiments would be time consuming and computationally expensive. We will consider doing these additional experiments provided the reviewer still considers them essential, and we are given the necessary time.

Secondly, the authors perform two experiments (named "Nutrients" and "Constant") in which certain parameters are held constant at some average values. Fig 12 shows some of the results, indicating that the constant values have a worse fit to data. However, one could argue that this is not an entirely fair comparison, as the constant values were not allowed to be optimized. That is, in how far would results and their interpretation change if the values of the constant parameters were adjusted, perhaps alongside all other parameters. I understand that the experiments were mainly conducted to examine the effect of underlying mechanisms that make these parameters non-constant, and I am not sure if this requires new experiments, but it would be nice to see at least some comment or discussion about the role of the constant value that was selected.

We do agree that Experiment 1 does not compare three possible optimized configurations, and as such, the two non-optimized ("Nutrients" and "Constant", names will change in the revised version) are in disadvantage with respect to *Optimized*. But the objective of the experiment is to highlight the vertical and temporal differences in CDOM production among formulations. Therefore, we will clarify this objective including the following statement in Section 3.3.1 In situ production of CDOM and synchronization with phytoplankton Chl-a: 'The constant % values in *Nutrients* and *Constant* were chosen to homogenise the values of  $f_{R2}^{X2}$  at the beginning of the year between the formulations and to highlight the vertical and temporal differences in  $f_{R2}^{X2}$  between the three formulations.'

Parts of the manuscript are well written and easy to follow, but in some parts I had to read sentences many times over to try to understand what is meant. Let me start with the notation: In the differential equations, many terms are abbreviated using a vertical line with a dx/dt part on the left and a 3-letter description and a variable on the right. This notation is never explained and, at first, I was under the impression the vertical line had a mathematical meaning (e.g., "evaluated at"). After going over equation 1 and accompanying text again, I noticed that this assumption was probably wrong. Instead, the variable on the bottom right of the vertical line may be the other variable this part of the equation is dependent on. I still don't exactly know, because this notation is never described in the manuscript. A couple of sentences would really help to make this notation accessible.

We will clarify notation by including new text in the first paragraph in Section 2.1.2 'The BGC model BFM and the bio-optical component': 'Following Vichi et al. (2007), the dynamical equations of this section are written in 'rate of change' form, where the right hand side contains the terms representing significant processes for each living or non-living component. For each process, the state variable subject to change is indicated before the vertical bar and the superscript and subscript after the bar are the 3-letter

acronym of the process represented and the state variable involved in the process, respectively.'

Similarly, the variable notation is not explained properly, and the reader is left guessing or is required to find more information in other parts of the manuscript. A small example: There are 4 zooplankton, which are (oddly) denoted  $Z^{\{3\}}$  to  $Z^{\{6\}}$ . Why is there no  $Z^{\{1\}}$ ? What are the kind of zooplankton is represented by each variable? Why do only two of them appear in Eq 8 and 9? Some information can be obtained from Fig 1, but this information should be much more front and center, so that these questions do not even arise. Again, a few extra sentences would make the manuscript a lot more accessible. I mention a few more instances when I encountered text without enough explanation in my specific comments below.

We will modify the notation for the four zooplankton groups (line 168) from numbers 6, 5, 4 and 3 to acronyms: HETN (heterotrophic nanoflagellates), MICRO (microzooplankton), OMNI (omnivorous mesozooplankton) and CARNI (carnivorous mesozooplankton), respectively. This hopefully will avoid the question about the lack of 1 and 2, and will clarify the kind of zooplankton each variable represents.

Only two of them appear in Eq 8 and 9 because mesozooplankton ( $Z_3=CARNI$  and  $Z_4=OMNI$ ) do not excrete dissolved carbon, only particulate. That's because the excretion of  $Z_3$  and  $Z_4$  is supposed to represent the production of fecal pellets and egestion (sloppy feeding). We will clarify as follows:

'Labile DOC ( $R_c^{(1)}$ ) is produced by the excretion of microzooplankton ( $Z^{(MICRO)}$ ) and nanoheterotrophs ( $Z^{(HETEN)}$ ) (*exc*), and therefore represents the sources of DOC associated to the zooplankton-mediated mortality of phytoplankton and bacteria, and it is consumed quickly by bacteria. The released fraction of carbon by mesozooplankton excretion (both  $Z^{(OMNI)}$  and  $Z^{(CARNI)}$ ), on the other hand, is assumed to have no dissolved products and therefore directed to the state variable  $R_c^{(DET)}$ .'

Additionally, and in response to some of the specific comments, we will modify the superscripts of other state variables, as follows:

$P^{(1)}$  (diatoms) =>  $P^{(DIATOM)}$

$P^{(2)}$  (nano-flagellates) =>  $P^{(NANO)}$

$P^{(3)}$  (picophytoplankton) =>  $P^{(PICO)}$

$P^{(4)}$  (dinoflagellates) =>  $P^{(DINO)}$

$R^{(6)}$  (detritus) =>  $R^{(DET)}$

$O^{(3)}$  (dissolved inorganic carbon) =>  $O^{(DIC)}$

$N^{(1)}$  (phosphate) =>  $N^{(PO_4)}$

$N^{(3)}$  (nitrate) =>  $N^{(NO_3)}$

$N^{(4)}$  (ammonium) =>  $N^{(NH_4)}$

The names of the simulations and experiments could be improved: The simulation names "Nutrients", "Constant" do not help the reader understand what is being modified or held constant. In comparison, The "Optimized" experiment has a better name. Then the

experiment names EXP-X<sup>{2}</sup> and EXP-X<sup>{3}</sup> are maybe less helpful than the simulation names: What is the significance of "X" in the name, and why start counting at 2? Just using "Experiment 1" and "Experiment 2" would raise fewer questions, but a short descriptive name would be even better.

Thanks for the suggestion, we will change the name of the simulations to give them short descriptive names. *Nutrients* and *Constant* will be changed to indicate what is being held constant.

Nutrients => Constant leakage.

Constant => Constant *dpp*.

And for consistency *Bleaching* will be also modified to indicate in which sense photobleaching rate is being modified:

Bleaching => Intense bleaching.

Regarding experiments EXP-X(2) and EXP-X(3), we will use short descriptive names and avoid numbering them:

EXP-X(2) (experiment 1) => EXP-Biology

EXP-X(3) (experiment 2) => EXP-Physics

There were a two instances showing a clear break in the train of thought, and I would suggest reordering the text: Firstly, Section 2.4.1 "Parameter optimization method" briefly introduces the parameter optimization method and then goes into detail (often difficult to follow, see specific comments) about which parameters were chosen to be optimized, before returning to the optimization method. Here, I would suggest moving the part about the parameters into its own subsection.

This Section 2.4.1 will be split into two: 2.4.1 'Parameters optimized' that describes which parameters have been chosen to be optimized, and 2.4.2 'Parameter optimization method' that describes the DE algorithm.

Secondly, Section 2.4.2 briefly introduces the hypothesis testing experiments without motivating these experiments well. Pages later, in Section 3.3.1, additional context is provided that is very useful to better understand the setup of the experiments. I suggest merging information from the two texts, so that the parameter changes performed are easier to understand and well motivated from the start.

Section 2.4.2 is now Section 2.4.3 'Simulations and hypothesis testing experiments'. We will reintroduce the hypothesis testing experiments:

'The model configuration described in Section 2.3 and the optimized parameter values obtained as described in Section 2.4.2 represents the *Optimized* model configuration. We used a series of hypothesis-testing experiments to investigate the mechanisms that determine the vertical and temporal patterns of CDOM distribution at the BOUSSOLE site. We conducted three more simulations in which we changed some of the assumptions of the *Optimized* configuration. All simulations are summarized in **Table 2**, all were run for 6-years with two years of spin-up and analyzed the results of the last four years (January 2011 to December 2014).'

Also, in response to the specific comment I 479 we will reformulate the description of each experiment.

I really like the figures, which look good and are informative. Here, I just have one general comment, I think it would be useful to consistently add description and units to their respective axis or the color bar. For example, in Fig 12, move the "a<sub>CDOM</sub>(450)" and units in (d) and (e) to the y-axis, just like in panels (a-c). In Fig. 3 and similar, move property name and units to the color bar. This step would reduce text in the panels and avoid repeating units.

Thanks for the suggestions, we will make the following changes to labels and axis:

Figure 12: panels d and e, move label to y axis; panels b and c, add 'surface' and 'DCM'.

Figures 3, 4, 5 and 6: move variable name to color bar.

Figure 7: move variable name to y axis.

Figure 13: move variable name to y axis, keep depth in the panel.

### specific comments

I 18: "The absorption coefficient of CDOM, [...], is measurable in situ and remotely": Can it indeed be measured reliably remotely and distinguished from other (living) light absorbing constituents?

No, remotely CDOM cannot be distinguished from detritus, not at least in current products. We will rewrite this part of the abstract: 'The absorption coefficient of CDOM,  $a_{\text{CDOM}}(\lambda)$ , is measurable in situ and can be retrieved remotely, although it is not distinguishable from the absorption of non-algal particles. These observations can be used as indicators for the concentration of other relevant biogeochemical variables in the ocean, e.g. dissolved organic carbon.'

Similarly, we will also rephrase the second paragraph of the introduction: ' $a_{\text{CDOM}}(\lambda)$  can be measured in situ at selected locations and be retrieved on a global scale from remote-sensing platforms, although remotely it is not distinguishable from the absorption of non-algal particles.'

I 40: "the spatio-temporal dynamics [...] reflects the functioning of the carbon cycle": What exactly is meant here, I would suggest rephrasing.

We will rewrite this sentence to: 'Furthermore, as the largest pool of reduced carbon in the oceans, it plays an important role in the global carbon cycle (Legendre et al., 2015).'

I 41: "non-absorbing": I would suggest changing this to "not light absorbing".

Thanks for the suggestion, we will rewrite this sentence to: 'Although most of the dissolved exudates that form the DOM are non light absorbing (Mühlenbruch et al., 2018), a fraction of DOM absorbs light mainly in the ultraviolet (UV) and blue spectral range of the electromagnetic radiation.'

I 50: "There are several possible factors determining...": "Besides the presence of CDOM, there are other possible factors influencing..."

Thanks for the suggestion, we will rewrite this sentence as: 'Besides the presence of CDOM, there are other possible factors influencing this optical behavior, including: the particular pigment ratios in the Mediterranean phytoplankton community (Organelli et al., 2011), the abundance of small coccolithophores (Gitelson et al., 1996), and the influence of Saharan dust (Claustre et al., 2002).'

I 63: Since vertical transport is mentioned here, why not horizontal as well? Maybe change "terrestrial and atmospheric inputs" to something like "terrestrial and atmospheric CDOM sources and physical transport"?

Thanks for the suggestion, we will rewrite this sentence as: 'CDOM cycling is essentially controlled by in situ biological production (Romera-Castillo et al., 2010), terrestrial and atmospheric sources and physical transport, microbial consumption (Nelson and Gauglitz, 2016; Legendre et al., 2015; Stedmon and Markager, 2005), as well as deep ocean circulation and/or vertical mixing (Coble, 2007), and it is photoreactive and efficiently destroyed in the upper layers of the water column by solar radiation (Mopper and Kieber, 2000).'

I 73: Is the NW Mediterranean Sea the study region?

No, in this paragraph we were referring to the higher-than-average DOM deposition fluxes reported for the whole Mediterranean Sea, thanks for pointing out. We will rewrite this sentence as: 'In the basin, the higher-than-average CDOM concentrations seem to be sustained by allochthonous sources, ...'

I 75: "depositions from the atmosphere are 2–5 times larger in the Mediterranean Sea than in the oceans": higher than "in the oceans" is not very specific: higher than anywhere in the ocean, higher than some average?

We do agree, we will rewrite this sentence as: '... as fluxes of DOM depositions from the atmosphere are 2–5 times larger in the Mediterranean Sea than those estimated for the global ocean, which explains the abundance of humic-like substances (Santinelli, 2015; Galletti et al., 2019).'

I 85: What does "probably controlled" mean? Perhaps change to "mostly controlled" or "its dynamics are dominated by"?

and

I 86: One can maybe make the counterpoint that photoacclimation and nutrient limitation are also often controlled by vertical transport. Maybe distinguish between being controlled by and directly caused by, as the preceding sentence suggests?

We will rewrite this sentence as: 'While  $a_{CDOM}(\lambda)$  dynamics at the surface is mainly explained by vertical transport of CDOM from depth and by photochemical degradation,

Chl-a concentration at the surface is mainly driven by photoacclimation of phytoplankton and nutrient limitation.'

I 92: "observations in the NW Mediterranean Sea suggest direct CDOM production by phytoplankton": Is this then equivalent to the way it is handled in other models, i.e. the fixed fraction of the dpp of phytoplankton?

This is equivalent to a fraction of the dpp of phytoplankton, not necessarily a fixed fraction.

I 169: Explain the symbols and superscripts used here (" $O^{(3)}$ " etc.).

We will change the numbers on the superscripts for more self-explanatory names, as follows:

'The implementation of the BFM in FABM comprises 54 state variables. These include representations of dissolved inorganic carbon ( $O^{(DIC)}$ ), inorganic forms of nitrogen ( $N^{(NO3)}$ ), ammonium ( $N^{(NH4)}$ ) and phosphorus ( $N^{(PO4)}$ ), four phytoplankton types ( $P^{(DIATOM)}$ ,  $P^{(NANO)}$ ,  $P^{(PICO)}$  and  $P^{(DINO)}$ ), heterotrophic bacteria (B) and four grazers ( $Z^{(HETN)}$ ,  $Z^{(MICRO)}$ ,  $Z^{(OMNI)}$  and  $Z^{(CARNI)}$ ), three pools of dissolved organic matter ( $R^{(1)}$  to  $R^{(3)}$ ) and CDOM ( $X^{(1)}$  to  $X^{(3)}$ ) differentiated by reactivity, and particulate organic matter ( $R^{(DET)}$ ).'

I 171: "A subscript appended to each module ...": This sentence requires more explanation. What are the modules? Is Chl-a considered an elemental constituent? I think I understand the rest of this paragraph, but this sentence is difficult to follow and seemingly unrelated to the preceding and following sentences.

We will clarify as: 'The subscript appended to each living and non-living component indicates the elemental constituents among carbon (C), nitrogen (N), phosphorus (P) and silica (S, only in  $P^{(DIATOM)}$ ), and the content in Chl-a only in PFTs.'

I 184: What is "OAC"?

Generally, OAC accounts for 'optically active constituents'. In this work, we refer to those as optical constituents and therefore the use of OAC is a typo. We will rewrite this sentence as: '... the optical treatment of constituents (PFTs, CDOM and detritus) ...'

Eq 1: "; i = 1 to 4" -> " for i = 1,2,3,4"

i = 1,2,3,4 will be included in Eq. 1 (phytoplankton) and i = 1,2,3 in Eq. 13 (cdom).

Eq 1: Why does zooplankton indexing start at 3?

We will modify the notation for the four zooplankton groups (line 168) from numbers 6, 5, 4 and 3 to acronyms: HETN (heterotrophic nanoflagellates), MICRO (microzooplankton), OMNI (omnivorous mesozooplankton) and CARNI (carnivorous mesozooplankton), respectively. The counter j will take values from 1 to 4, it will be modified in Equations 1, 8 and 9.

I 191: As an aside, I find the "prd" abbreviation of "predation" slightly confusing, as I always think of "production" first. I would suggest using "grz" for "grazing" instead. 'prd' is the 3-letter acronym that BFM uses conventionally to refer to predation. We will substitute 'prd' by 'grz' for clarity.

Eq 2: Underneath, it reads like both temperature and nutrients are dimensionless, when really these temperature- and nutrient-dependent factors are dimensionless. I would suggest changing "ft\_P" to "ft\_P(T)" where T is model temperature. The same applies to fn\_P, which has the benefit of telling the reader which nutrients are entering this equation and control GPP. Currently, it might appear to a reader as if GPP is independent of other model variables.

We will clarify the formulation of the temperature- and nutrient-dependent factors as follows:

'Gross primary production of each phytoplankton species (mg C m<sup>-3</sup> d<sup>-1</sup>) is computed as follows:

$$\frac{dP_c^{(i)}}{dt} \Big|_{O(3)} = r_{0p}^{(i)} \times ft_p^{(i)}(T) \times fn_p^{(i)}(\text{nutrients}) \times \left( 1 - \exp \left( \frac{-\phi^{(i)} \times \theta^{(i)} \times \int_{387.5}^{800} a_{PS}^{*(i)}(\lambda) \times E_0(\lambda) d\lambda}{r_{0p}^{(i)}} \right) \right) \times P_c^{(i)}$$

where  $r_{0p}$  is the maximum productivity rate (d<sup>-1</sup>), regulated by dimensionless temperature- ( $ft_p(T)$ ) and nutrient-dependent ( $fn_p(\text{nutrients})$ ), nutrients are N<sup>(PO<sub>4</sub>)</sup>, N<sup>(NO<sub>3</sub>)</sup> and N<sup>(Si)</sup> for P<sup>(DIATOM)</sup>, and N<sup>(PO<sub>4</sub>)</sup> and N<sup>(NO<sub>3</sub>)</sup> for the rest of the phytoplankton types) factors.'

I 195, 197: Sometimes "mg Chl", sometimes "mg Chl-a" is used. To stay consistent with previous notation, I would suggest using "mg Chl-a" and "(mg Chl-a)<sup>-1</sup>" when in the denominator.

Thanks for pointing this out. We will correct Chl to Chl-a throughout the text, including when it is mentioned in the Chl-a:C ratio.

Eq 4: I presume, P\_P is the phosphorous quota (subscript P) of phytoplankton (regular P).

Is the superscript denoting the phytoplankton type missing, or am I misunderstanding?

Similarly, is the "P\_N<sup>{(k)}</sup>" meant to be "P\_{N<sup>{(k)}</sup>"? That is, the superscript "(k)"

appears to be attached to the P, but should it be attached to the subscript N instead?

Thanks for spotting this. The superscript (i) denoting the phytoplankton type was missing in the P quota and inserted incorrectly as (k) in the N quota. This Equation will be corrected as:

' $G_p^{balance}$  is calculated as:

$$G_p^{balance} = \min \left( G_p, \frac{1}{p_p^{min}} \times \frac{dP_p^{(i)}}{dt} \Big|_{N^{(PO4)}}, \frac{1}{n_p^{min}} \times \sum_{k=3,4} \frac{dP_N^{(i)}}{dt} \Big|_{N^{(k)}} \right)$$



where  $p_p^{min}$  and  $n_p^{min}$  are the minimum phosphorous and nitrogen quota, respectively, and the two nitrogen sources are  $N^{(NO3)}$  and  $N^{(NH4)}$ .

Eq 5: What is i set to here, is there perhaps a sum missing?

Thanks for spotting this, (i) is included because Eq. 5 indicates how  $f_{R2}^{X2}$  is computed for each phytoplankton type, and not the coloured fraction of the total DOC of phytoplankton origin as the text says. We will keep referring to the coloured fraction of the total DOC of phytoplankton origin as  $f_{R2}^{X2}$  (that is shown in Figure 1 and Figure 12), and will use  $f_{P(i)}^{X2}$  to refer to the coloured fraction of the DOC produced by phytoplankton type i. Eq. 5 will therefore be rewritten to:

$$f_{P(i)}^{X2} = f_{P(i)}^{maxX2} \times \frac{\theta^{(i)}}{\theta_{chl}^{0(i)}} \times \frac{\beta_P^{(i)} \times \frac{dP_c^{(i)}}{dt} \Big|_{\frac{gpp}{O^{(3)}}}}{\frac{dP_c^{(i)}}{dt} \Big|_{\frac{dpp}{R_c^{(2)}}}}$$

and we will change equations 6 and 7 accordingly as:

$$\frac{dR_c^{(2)}}{dt} = \sum_{i=1,2,3,4} \left( \frac{dP_c^{(i)}}{dt} \Big|_{\frac{dpp}{R_c^{(2)}}} \times (1 - f_{P(i)}^{X2}) \right) - \frac{dB_c}{dt} \Big|_{\frac{upt}{R_c^{(2)}}}$$

$$dX_c^{(2)}/dt = \sum_{i=1,2,3,4} \left( \frac{dP_c^{(i)}}{dt} \Big|_{\frac{dpp}{R_c^{(2)}}} \times f_{P(i)}^{X2} \right) - \frac{dB_c}{dt} \Big|_{\frac{upt}{R_c^{(2)}}} - \frac{dX_c^{(2)}}{dt} \Big|_{\frac{deg}{R_c^{(3)}}}$$

The maximum fraction of coloured DOC in each phytoplankton type that is prescribed as a fixed parameter will be renamed as  $f_{P(i)}^{maxX2}$ , and Table 2, Table A1 and Figure A2 will be updated with the new name of the parameter.

I 231: Why do the other zooplankton not excrete labile DOC? Are they too small/large?

This information is missing from the manuscript.

Line 232 states that 'Large zooplankton excretion ... is composed by particulate only and directed to the state variable  $R(6)$ '. That's because the excretion of Z3 and Z4 is supposed to represent the production of fecal pellets and egestion (sloppy feeding).

We will clarify as follows: 'The released fraction of carbon by mesozooplankton is assumed to have no dissolved products and therefore directed to the state variable  $R(6)$ '.

Eq 14-16: Out of curiosity, why are different reference wavelength used in the 3 equations (450 nm, 440 nm and 550 nm)?

Because the reference wavelength is among the parameters obtained by fitting non-linear models to observed  $a_{CDOM}(\lambda)$ ,  $a_{NAP}(\lambda)$  and  $b_{NAP}(\lambda)$  spectra. In the basic formula for the specific absorption coefficient:

$$a=c2 \times \exp(-S(\lambda-\lambda_{ref}))$$

$c2$ ,  $S$  and  $\lambda_{ref}$  are all coefficients obtained when fitting the non-linear model to observed spectra (of both  $a_{CDOM}(\lambda)$  and  $a_{NAP}(\lambda)$ ). For the absorption of CDOM,  $c2$  scales the

absorption of CDOM at a reference wavelength ( $\lambda_{\text{ref}}$ ) of 450 nm (taken from Dutkiewicz et al. 2015, originally from the parameters fitted by Kitidis et al. 2006). For the absorption of NAP,  $c_2$  scales the absorption of NAP at a  $\lambda_{\text{ref}}$  of 440 nm (taken from Gallegos et al. 2011, originally from the parameters fitted by Bowers and Binding 2006).

The same happens in the basic formula for the scattering coefficient:

$$b = c_2 \times (\lambda_{\text{ref}}/\lambda)^e$$

where  $c_2$ ,  $e$  and  $\lambda_{\text{ref}}$  are all found when fitting the non-linear model to observed spectra.

In this case,  $c_2$  is the mass-specific scattering coefficient at a  $\lambda_{\text{ref}}$  of 555 nm (originally from parameters fitted by Babin et al., 2003 to phytoplankton scattering spectra, Gallegos et al. 2011 used the same  $\lambda_{\text{ref}}$  and adapted  $c_2$  to represent the scattering of several types of non-algal particles).

I 356: "()" are missing for  $a_{\text{PH}}$  and  $a_{\text{DG}}$ .

We will rewrite this part as: IOPs ( $a_{\text{PH}}(443)$  and  $a_{\text{DG}}(443)$ ) are derived ...'

Fig. 2: x- and y-axis labels are missing. Consider coloring land gray in the top left map.

Thanks for the suggestion. We will add '°E' and '°N' to the x- and y-axis respectively. We will color land grey in the small map of Figure 2.

I 378: "assumed equal": Why "assumed"?

'assumed' will be removed and the sentence left as follows: 'The parameter values related to the partition between dissolved and particulate excretion in  $Z^{(\text{MICRO})}$  and  $Z^{(\text{HETEN})}$ , and the mortality of  $Z^{(\text{CARNI})}$  and  $Z^{(\text{OMNI})}$ , are equal to the up-to-date values derived in Álvarez et al. (2022), together with all the optical parameters.'

I 389: "and restored at monthly frequency": Does this imply nudging/relaxation?

Yes, it does. We will rewrite this sentence as follows to clarify: 'Observed vertical profiles of T and S ... were ... used as initial conditions and constrained by a monthly nudging, in order to reproduce the intensity and timing of the mixing as closely as possible to observations.'

I 419: I would suggest removing the "while" here.

This 'while' will be removed as a result of the reorganization/reformulation of Section 2.4.1 'Parameters optimized' and Section 2.4.2 'Parameter optimization method' suggested in a previous comment.

I 425: What does "assumed true" mean, assumed that the true value is known and not included in the optimization? I would prefer "assumed known" or "assumed constant". Please make it explicit that this implies that the parameter in question is not included in the optimization. This needs to be made clear initially, the phrase is used a lot.

These parameters have not been included in the optimization. We will rephrase to clarify at the beginning of Section 2.4.1 'Parameters optimized' that all parameters not included in the optimization are assumed constant. We will also reformulate the whole section slightly in order not to repeat as much 'assumed constant'.

I 426: Subscript "(5 to 6)" and "(1 to 4)" are unconventional and could be confusing some readers. I suggest writing them out fully, i.e. (I am not including superscripts here):  $f_Z(5)$ ,  $f_Z(6)$ ,  $f_P(1)$ , ...,  $f_P(4)$ , and  $f_B$ .

This sentence will be changed to: 'We assumed constant the absorption coefficients of the three reactivities at 450 nm ( $a_{X(1)}^{450}$ ,  $a_{X(2)}^{450}$  and  $a_{X(3)}^{450}$ ) and optimized parameters related to the production of CDOM ( $f_Z^{X1}$ ,  $f_Z^{X1}$ ,  $f_P^{X2}$ ,  $f_P^{X2}$ ,  $f_P^{X2}$ ,  $f_P^{X2}$ ,  $f_P^{X2}$ ,  $f_P^{X2}$ ) and those related to the photobleaching of  $X_C^{(3)}$  ( $b_{X(3)}$  and  $I_{X(3)}$ ).'

I 432: "All optimized parameters appear in some process independently:": As a reader, I was now expecting a brief review of where the parameters mentioned in the previous sentences appear in equations. Instead, new parameters are mentioned and added to the list, of which two are not actually appearing in some process independently. Does the "all parameters" include the previously named ones? It would be helpful to break up this long sentence into two, don't use "all parameters" (one could use "the following" or similar).

Thanks for pointing out, we agree that the parameter appearing in an independent process is a condition that all parameters included in the optimization must meet, and in fact is met in our study. We will move this consideration to the beginning of the section, as it applies to all parameters. For phytoplankton, we will reformulate to avoid the use of "all parameters" and to clarify that we included in the optimization the two parameters that appeared in two different processes (the absorption cross section and the maximum Chl-a:C quatum), and chose one of the two that appear together in photosynthesis (the absorption cross-sections of photosynthetic pigments and the photochemical efficiency).

I 435: "We assumed true  $a_{PS}$  and optimized  $\phi_C$ , because these parameters are not well documented in literature being more difficult to measure.": Does this mean  $\phi_C$  was included in the optimization and  $a_{PS}$  was not, because they both appear in the same equation, and they may be dependent on each other?

Yes, the absorption cross-sections of photosynthetic pigments ( $a_{PS}$ ) and the photochemical efficiency ( $\phi_C$ ) only appear multiplied in the photosynthesis equation. Their product is the equivalent to the so-called initial slope of the production-irradiance (PE) curve ( $\alpha$ ), used in most of the non-spectrally resolved formulations of the PE relationship.

I 441: "A total of 25 optical and BGC parameters were optimized": I think it would be helpful to mention this at the beginning of the paragraph.

We will move this sentence to the beginning of Section 2.4.1 'Parameters optimized'.

After the modifications stated in comments from I425 to I441, the rewritten Section 2.4.1 will read as follows:

'Observations used for optimization were all collected at the BOUSSOLE site at monthly temporal resolution and roughly at the same number of discrete depths and included pico-, nano- and micro-phytoplankton Chl-a (pico-Chl-a, nano-Chl-a and micro-Chl-a, respectively),  $a_{PH}(450)$ ,  $a_{NAP}(450)$ ,  $a_{CDOM}(450)$  and  $S_{CDOM}^{350-500}$ . The potential number of independent parameters included in the optimization problem is limited by the observations available and by the fact that optimized parameters must appear in some process independently. A total of 25 optical and BGC parameters were optimized (Table A1 in Appendix A), all of them were given a positive range of values for initialization and evolution, without conditions or trade-off for the consistency between pairs of parameters. The rest of the parameters not included in the optimization kept constant values (Table 1). The only observations related to CDOM were  $a_{CDOM}(\lambda)$  that depend both on the mass-concentration and on the mass-specific absorption coefficients across the electromagnetic spectrum. We assumed constant the absorption coefficients of the three reactivities at 450 nm ( $a_{X(1)}^{450}$ ,  $a_{X(2)}^{450}$  and  $a_{X(3)}^{450}$ ) and optimized parameters related to the production of CDOM ( $f_Z^{X1(MICRO)}$ ,  $f_Z^{X1(HETEN)}$ ,  $f_P^{X2(DIATOM)}$ ,  $f_P^{X2(NANO)}$ ,  $f_P^{X2(PICO)}$ ,  $f_P^{X2(DINO)}$  and  $f_B^{X3}$ ) and those related to the photobleaching of  $X_C^{(3)}$  ( $b_{X(3)}$  and  $I_{X(3)}$ ). The spectral slope of the CDOM pool has been associated with aromaticity and average molecular weight of the CDOM compounds (Blough and Green, 1995) but appears to be less linked to CDOM concentration. Therefore, the spectral slopes between 350 and 500 nm for the three reactivities ( $S_{X(1)}^{350-500}$ ,  $S_{X(2)}^{350-500}$  and  $S_{X(3)}^{350-500}$ ) were optimized to ensure the proper simulation of  $a_{CDOM}(\lambda)$  in wavebands other than 450 nm. For phytoplankton, independent observations regarding mass-concentrations (Chl-a) and optical properties ( $a_{PH}(\lambda)$ ) were available, therefore both types of parameters were optimized. We optimized the absorption cross-sections ( $\underline{a}_{PH}^*$ ) and the maximum Chl-a:C quota ( $\theta_{chl}^0$ ) as they appear independently in light attenuation (Eq. (17)) and in photoacclimation (Eq. (19) in Álvarez et al., 2022) and in CDOM production (Eq. (5)), respectively. The absorption cross-sections of photosynthetic pigments ( $\underline{a}_{PS}^*$ ) and the photochemical efficiencies ( $\phi_c^0$ ) of the PFTs appear only in photosynthesis (Eq. (2)). We kept constant  $\underline{a}_{PS}^*$  and optimized  $\phi_c^0$ , because these parameters are not well documented in literature being more difficult to measure. For detritus, we optimized the reference absorption coefficients at 440 nm ( $a_R^{440}$ ) and kept constant all BGC parameters that alter the mass-concentrations. Given that  $a_{NAP}(\lambda)$  observations were available, optical parameters related to the spectral shape of absorption by detritus ( $S_R^{350-500}$ ) could have been included in the optimization, but we decided to maintain this parameter constant given the small contribution of  $a_{NAP}(\lambda)$  to the total non-water absorption as compared to CDOM and phytoplankton. All parameters related to scattering and backscattering both of phytoplankton ( $\underline{b}_{PH}^*$  and  $bbr_{PH}$ ) and detritus ( $b_R^{550}$ ,  $e_R$  and  $bbr_R$ ) were not included in the optimization.'

I 449: "O\_1, ..., O\_n, O\_N": This notation is confusing, why include n and N here, is n=N-1? The same applies to P.

We will remove this notation. The sentence will read: 'For any variable observed  $m$  from a total of  $M$ , there will be  $N$  number of pairs consisting of the simulation  $P_n$  and the corresponding observation  $O_n$ .'

Eq 27: Similar confusion, I assume  $n$  iterates from 1 to  $N$  here. So it should be " $n = 1, \dots, N$ " or (more conventional) " $n=1$ " and the " $N$ " is placed on top of the sum. The same applies to equation 29.

We will correct the  $\Sigma$  on both equation 27 and Eq. 29 to place " $n=1$ " and " $m=1$ ", respectively, on the bottom of the sum and " $N$ " and " $M$ ", respectively, on top of the sum.

Eq 29: So  $N_m$  is the number of observations for variable  $m$ ? This needs to be made explicit, and I suggest introducing this notation right away in Eq 27, so that it is clear what  $SSQ_m$  is. I probably do not understand the terms here and what the subscript  $m$  indicates, it appears like  $SSQ_m/(2 \times \sigma_m) = SSQ_m/(2 \times SSQ_m/(N_m-1)) = (N_m-1)/2$ ? This is very confusing.

We will clarify that  $m$  refers to each variable observed of a total of  $M$ , and  $n$  refers to each pair observation-prediction from a total of  $N$ . In the corrected notation, the number of observations of a given variable  $m$  is denoted as  $N_m$  and the sum of the squares as  $SSQ_m$ . Regarding the formulation of the  $\ln$ Likelihood, in the current version of the manuscript we gave the general formulation for the  $\ln$ Likelihood as it is computed by ParSAC. This formulation can use a standard deviation for the observations that is manually provided by the user, or, if the  $sd$  is not specified (this work), ParSAC estimates it from the model-observation differences (Eq. 28). In the first case, the  $SSQ/2sd^2$  terms ( $sd$ -dependent) in the  $\ln$ Likelihood (Eq. 29) are not constant, and can therefore not be dropped. We agree with the reviewer that in our particular application, where we do not provide a known  $sd$  for each variable and compute them instead from the  $SSQ$  (second case), the last term in the  $\ln$ Likelihood can be dropped. Therefore, for the sake of clarity, we will simplify Eq. 27, 28 and 29, to just two equations as follows:

'ParSAC formulates such fitness (i.e. probability that the candidate parameter values are the true parameter set representing reality) as a multi-objective function calculated as the log-transformed likelihood between the outcome of the model and the observations provided, assuming the observed values are log-normally distributed, with the median equal to the model prediction. For any variable observed  $m$  from a total of  $M$ , there will be  $N$  number of pairs consisting of the simulation  $P_n$  and the corresponding observation  $O_n$ . The sum of the squares of the residuals for each variable observed  $m$  ( $SSQ_m$ ) is computed as:

$$SSQ_m = \sum_{n=1}^{N_m} [P_n - O_n]^2$$

For  $M$  variables observed, all differences between model and observations are combined as:

$$\ln Likelihood = \sum_{m=1}^M -N_m \times \ln(SSQ_m)$$

I 460: The DE algorithm is not explained very well. For example, it is not clear from the text which of the symbols denotes the target vector and which is the mutant vector. What does "crossing" with the mutant vector mean? In "All target vectors in the first generation are crossed with the mutant vector", it sounds like the same mutant vector is applied to all target vectors. If I did not know how DE worked, I would not have a better idea after reading this section. This description needs to be rewritten.

We will rewrite this section as follows:

'Each individual in a given population of size  $S=288$  is a 25-dimensional target vector that represents a candidate solution to the problem. The parameter values in the target vectors in the first generation are initialized randomly across the parameter space. The parameter values were **sampled uniformly between minimum and maximum values**, which are listed in **Table A1** (Appendix A). The DE algorithm creates new generations of individuals by applying cycles of mutation, crossover and selection operating on the target vectors. Mutation: once all simulations in a generation are completed, the DE algorithm creates a new set of  $S$  mutant vectors. Each mutant vector is created selecting randomly three targets ( $x_{r1}$ ,  $x_{r2}$  and  $x_{r3}$ ) and applying a mutation operator that consisted on  $x_{r1} + 0.5 \cdot (x_{r2} - x_{r3})$ . Crossover: to increase the diversity of the parameter vectors, crossover is done to generate the final trial set of individuals. Crossover is performed between each target vector and its corresponding mutant vector, retaining the target vector in the population with a probability of 0.9, otherwise introducing the mutation only with a 0.1 probability. Selection: the final phase is selection where the  $\ln$ Likelihood of the trial vector is compared to the  $\ln$ Likelihood of the corresponding initial target. The trial vector replaces the target if the  $\ln$ Likelihood value obtained from the trial is higher or equal than the  $\ln$ Likelihood obtained from the target vector. After the selection procedure, the size of the population is again  $S$ . The cycle of mutation, crossover and selection is carried out iteratively and eventually, the DE algorithm provides an estimate of the optimal parameter values that minimizes the misfit between the model output and the observations (i.e. maximizes the sum of  $\ln$ Likelihoods in the population).'

I 466: "optimal set of 25-parameter values": The DE algorithm aims to find optimal parameter values, but that does not mean that the resulting parameters are indeed optimal in that they maximize the likelihood. Hopefully, they are close to the global optimum. In other words, I would suggest using "estimate of the optimal parameter values".

We will take the reviewer suggestion and the sentence will read: 'After the simulation of a new generation, the cycle of mutation, crossover and selection is carried out iteratively and eventually, the DE algorithm provides an **estimate of the optimal parameter values** that minimizes the misfit between the model output and the observations (i.e. maximizes the sum of  $\ln$ Likelihoods in the population).'

I 476: "reversed some of the assumptions of the optimized configuration": What does "reversed" mean here? I think "changed" might be more suitable.

We will use 'changed' instead of 'reversed'. Additionally, this first part of the Section 2.4.3 'Simulations and hypothesis testing experiments' will be rewritten in response to one of the general comments.

I 479: "The comparison of the results of Optimized with Nutrients and Constant constituted and experiment..." This sentence is difficult to follow, I suggest rephrasing. I had to read the previous sentences several times as well: one uses "removed the dependence >>of<< f on  $\theta/\theta_{chl}$ ", the other "removed both dependencies >>from<< f, on  $\theta/\theta_{chl}$ ", making it very difficult to understand what is dependent on what, and what is being removed. Here, I strongly suggest stating the aim of each experiment first and then go into details of the implementation.

Section 2.4.3 'Simulations and hypothesis testing experiments' will be rewritten in response to one of the general comments. The two experiments will be described separately as follows:

'In **EXP-Biology** we investigated the role of nutrient and light limitations on controlling the biogenic in situ production of biodegradable CDOM and therefore on generating the observed  $a_{CDOM}(\lambda)$  values. This experiment consisted of comparing the results of *Optimized* with two additional simulations. In **Constant leakage**, we considered the proportion  $X_C^{(2)}:R_C^{(2)}$  in *dpp* as a constant fraction of leakage and therefore removed the dependence of  $f_{R2}^{X2}$  on  $\theta/\theta_{chl}^0$ . In **Constant dpp**, we considered the fraction  $X_C^{(2)}:R_C^{(2)}$  in *dpp* constant and therefore removed both dependencies from  $f_{R2}^{X2}$ , on  $\theta/\theta_{chl}^0$  and on leakage/*dpp*.

In **EXP-Physics** we investigated the role of vertically-transported CDOM in maintaining CDOM concentrations in the surface ocean. This experiment consisted of comparing the results of *Optimized* with an additional simulation, **Intense bleaching**, where we prescribed the same bleaching parameters for  $X_C^{(3)}$  ( $b_{X(3)}$  and  $l_{X(3)}$ ) as for the other two CDOM pools and quadrupled the CDOM:DOC fractions in all biotic sources ( $f_{Z(MICRO)}^{X1}$ ,  $f_{Z(HETEN)}^{X1}$ ,  $f_{P(DIATOM)}^{X2}$ ,  $f_{P(NANO)}^{X2}$ ,  $f_{P(PICO)}^{X2}$ ,  $f_{P(DINO)}^{X2}$  and  $f_B^{X3}$ ). By comparing *Optimized* with **Intense bleaching** we evaluated the extent to which the *Optimized* simulation was able to represent the role of  $X_C^{(3)}$ , and thus of any source of allochthonous CDOM, in generating the observed  $a_{CDOM}(\lambda)$  values.'

Fig 3: Mention that these are climatological values in the caption.

The caption will be rewritten to:

'Figure 3. **Annually-averaged** vertical profiles as a function of time of (a-b) temperature, (c-d) nitrate and (e-f) phosphate, obtained from observations at the DYFAMED site (left panels) and from the *Optimized* simulation (right panels). The white line in the temperature panels represents the mixed layer depth (mld) obtained in (a) with a threshold of 0.2 °C on temperature (de Boyer Montégut et al., 2004) and in (b) with a threshold of  $1 \cdot 10^{-5} \text{ m}^2 \text{ s}^{-2}$  on turbulent kinetic energy. The line on (c) and (d) represents the nitracline (nitrate concentration equals 2 µM).'

Fig 3: Units in the color bar and in the panel are not identical (they are equivalent, but why create confusion).

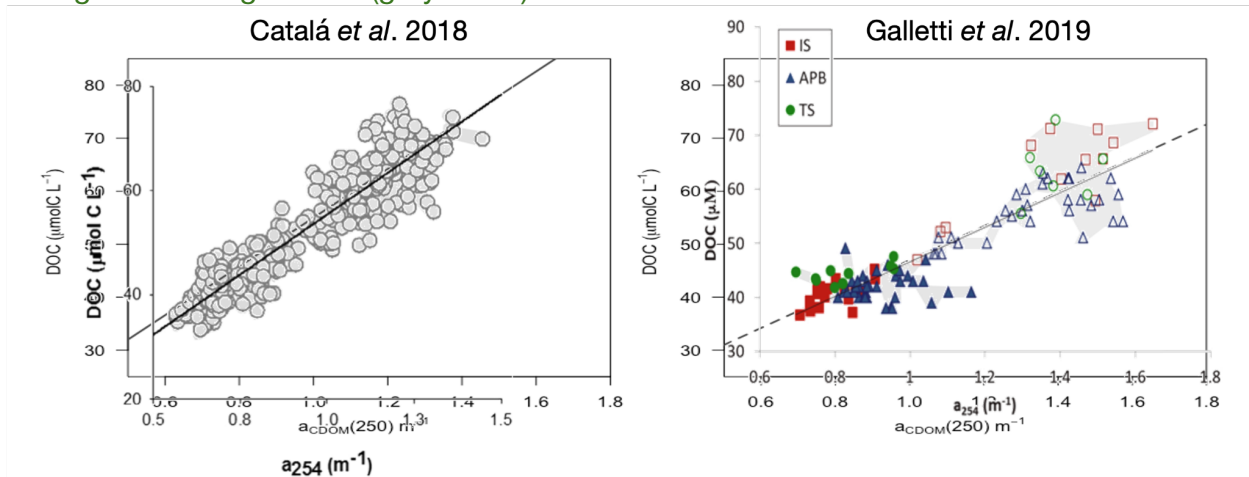
We will remove the units from the panels and leave the units in the color bar, both in the nutrient panels of Figure 3 and the DOC panel of Figure 5.

I 568: Out of curiosity, what are the correlations between in situ and satellite observations? Thanks for the question, we will evaluate the task of intercalibration exercise between observing systems in future works. For a detailed comparison between remotely and in-situ collected observations at the Boussole site, we suggest figures 40 (Chl-a, page 52) and 52 (CDOM, page 65) of the following report: 'Climatological characterization of Ocean Sites for OC-SVC-BOUSSOLE-D4' <https://www.eumetsat.int/media/49886>

Fig 7: What happened in the model in early 2013 when there appears to be a drop of model chlorophyll to nearly zero? It is the effect of the intensive winter mixing that homogenizes concentrations quite deep in the water column and takes chlorophyll almost to zero. It also happens in early 2012, although not so intensively. We have stated in the text that these periods have not been included in the metrics computation.

Fig 8 and 9: it might be helpful to change the order of legend entries to match the extra axes. We will reorder the legend elements in order (top to bottom): Bc, aCDOM, TChl-a, DOC.

Fig 11b: Why are the ranges of reported values so patchy? Because these are areas outlining clouds of points. We show below the original figures from Catalá et al. (2018) and Galletti et al (2019) with the ranges that we included in the background of Figure 11b (grey areas).



I 683: "1) depends": Missing "it".

We will correct this sentence as follows: 'Therefore, the fraction  $X_c^{(2)}$ :  $R_c^{(2)}$  in the *Optimized* simulation was variable in time and depth for two reasons: 1) **it** depends on the relative



proportions of DOC from leakage or exudation and 2) it depends on the  $\theta$  to  $\theta_{chl}^0$  ratio of phytoplankton.'

I 684: "Two additional simulations considered an average value of 6.2% of CDOM in the leakage flux": writing "considered an average value" is not very precise, the reader needs to know exactly what is happening, for example: "Two additional simulations use a constant fraction of CDOM in the leakage flux, setting the CDOM fraction to 6.2%, which is the average value obtained from the Optimized simulation."

Thanks for the suggestion, this sentence will read: 'Two additional simulations use a constant fraction of CDOM in the leakage flux (*Constant leakage*, **Table 2**) and a constant fraction of CDOM in the total dpp flux (*Constant dpp*, **Table 2**), setting these CDOM fractions to 6.2% and 1.8%, respectively, which is the average values obtained from the *Optimized* simulation.'

Fig. 12: Panel (d) says "0-9m" but (e) is missing units: "40-60". It would also be useful to have this information in panels (b) and (c).

In Figure 12, we will add 0-9m to panel (b), DCM to panel (c) and 40-60m to panel (e).

I 799: "This is in better agreement with observations when the proportion of coloured material in the dpp depends on nutrient and light constraints": Add a reference to the model here, and change the "when" to "if", otherwise readers might think the "when" means "at times when" which would completely change the meaning of the sentence. Proposed new version: "This is in better agreement with observations if the proportion of coloured material in the dpp is modelled to depend on nutrient and light constraints"

Thanks for the suggestion, we will take it and use the proposed version.

I 809: Light limitation of what?

We will modify this sentence to: 'Light limitation to phytoplankton CDOM production acts through the elemental ratio of Chl:C relative to the maximum value...' For consistency we will also update the corresponding sentence for nutrient limitation in L801 to: 'Nutrient limitation to phytoplankton CDOM production acts through the relative proportion between leakage and *dpp*...'

I 836: "Adding optical components to BGC models greatly improves the comparability of simulated and observed types of variables": What exactly is meant here, it becomes easier to compare model and observations when the model has optical components?

No, we meant it increases the points of convergence 'simulated variable-observed variable'. We will rephrase as: 'Adding optical components to BGC models greatly increases the amount of data that can be used for model validation and calibration.'