Reviewer 1

We kindly thank the reviewer for his valuable comments and suggestions. The accepted changes from the attached pdf file are added in the updated version of the manuscript (highlighted in **fuchsia**). Here we are replying to the major comments and questions from the pdf file.

Major Remarks

The paper needs some more editing to be easily readable. Its organization could be improved a lot using tables describing the different configuration used (e.g. Which specific IOP choice for each, which property profile is used for each) and to display the results.

Thank you, that was very helpful. By having added tables to the updated version of the manuscript we hope to have made results much clearer and the whole paper more readable.

The neglect of Raman needs significant justification, in particular at 490nm where it could significantly affect measured K_d and R_{rs}.

Thank you for having pointed this out, as well as for having provided us with the adequate reference. We accounted for Raman scattering by correcting R_{rs} values according to Lee et al. [2013]. The inclusion of Raman scattering results in up to 3.5, 4 and 8.5% difference in terms of R_{rs} at 412, 443 and 490 respectively, as seen from the Raman factor values in Fig.1. This has resulted in a minor change which has not impacted the choice of the final modelling configuration in terms of scattering. The three-stream radiative transfer modelling configuration does not include Raman in the computation of in-water irradiances, thus the error still exists in the irradiance and K_d values. We are not familiar with any study that would similarly correct K_d values as it was done for R_{rs}. In the future we might be able tackle this issue at more depth, possibly by comparing the three-stream model with more refined configurations such as Hydrolight.

The neglect of a_ϕ at 380nm deserves more justification. Works of Bricaud, among others, suggest significant phytoplankton absorption there due to MAAs, particularly near the surface.

We acknowledge our text might have been a bit confusing and will thus rewrite it. a_ϕ was indeed considered at all wavelengths whenever a_ϕ models were included in the simulations, therefore also at 380 nm. It had a substantial contribution at all 3 wavelengths and hence also at 380 nm (the inclusion of a_ϕ results in a 32% decrease in RMSE compared to the pure water configuration, whereas bias decreases by 37.5% and the correlation increases by 66%), which is in line with your remark. We will underline this more clearly in the updated manuscript version.

The effect of S and T on water IOPs is very different in the bands investigated. S primarily contribute to backscattering but if a mean salinity is used, the change may be rather small. T affects primarily absorption in the NIR. Again, using an average T may be more than sufficient anywhere else.

We agree that average values of T and S would have been sufficient. As reported below, by using mean values for each profile the performance stays the same, exactly as you have pointed out. The aim of the inclusion of precise in-situ values was however more to quantify the difference in the model performance when considering T-S correction or else, without taking the intermediate step of using only mean values. It was not costly in terms of computational time, and we have the depth profiles of T and S, thus we opted for the latter choice. The skill improvement when in-

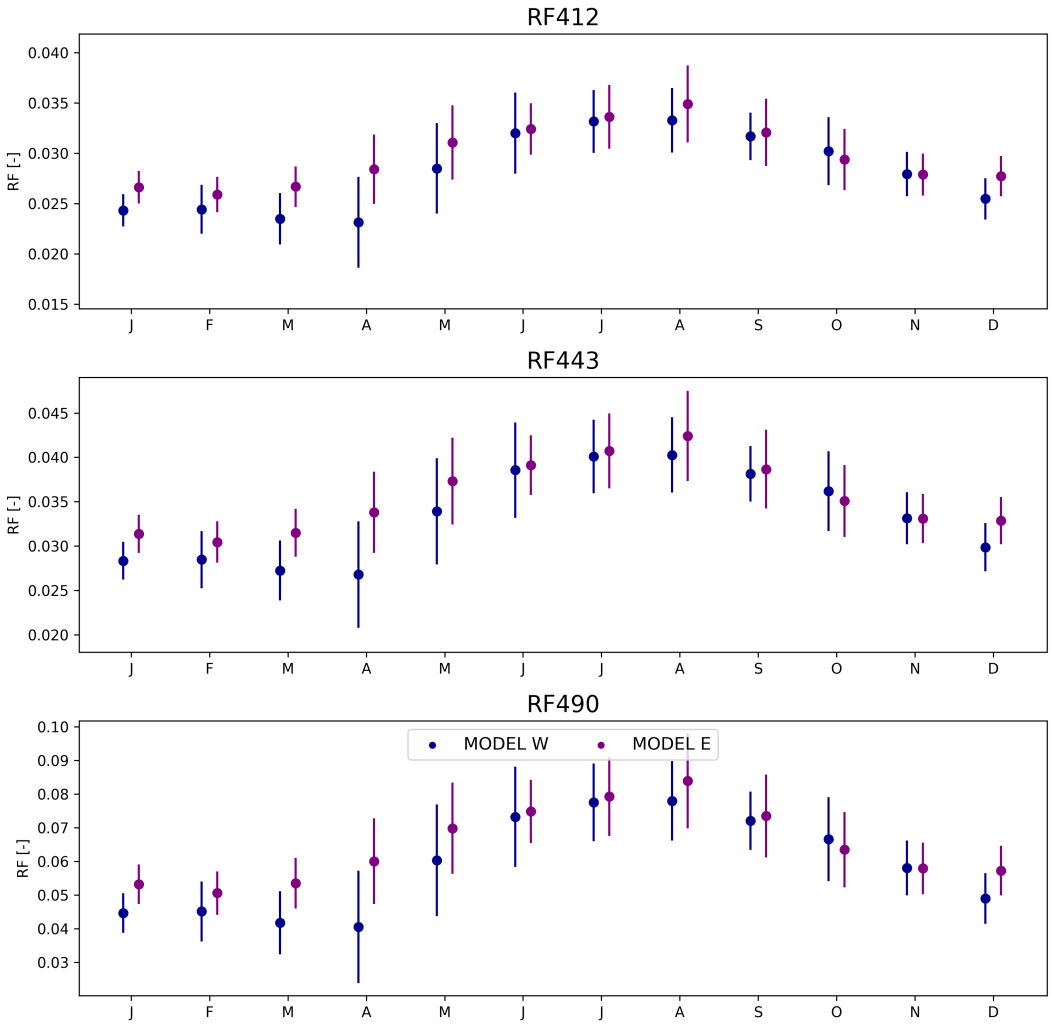

Figure 1: Monthly climatological values of the Raman factor according to Lee et al. [2013] for western and eastern Mediterranean

cluding T and S profiles from floats was small, however present, compared with the configuration with excludes the use of T and S data altogether.

The final result that the best choice is $\widetilde{b_{bp}}(\lambda)$ =0.002 and η=0 deserves attention. It does not seem consistent with expectation from other studies.

We noticed that the reviewer didn't receive our last updated version, which corrected the R_{rs} calculation compared to the previous one (i.e. transmission from in- to above-water). This resulted in more sensible values also in terms of the best scattering configurations. The chosen configuration with a slope of 3 is consistent with the range of values from the reported literature, reaching best agreement for the Eastern Mediterranean. Exceptions are seen during summer months, when the most adequate slope amounts to 4, and for the west, where modelled and observed values align better with a slope of 2. This result is encouraging as it suggests that there are different scattering regimes at play in the two basins, most likely stemming from a different particle size distribution. This will be further highlighted in the updated text. Still, the authors are aware that the selection of the proper scattering model proves to be the weakest point of the paper. As already pointed out in the text, with the lack of in-situ L_u measurements we are certainly not in the position to quantify the skill of different scattering models as thoroughly as we could assess the skill of absorption models that were evaluated both in terms of in-situ E_d and in-situ and remote sensing K_d values. Changing b_{bp} models didn't result in major changes in K_d values, so R_{rs} was the only AOP available for the model skill validation. This could be properly tackled by including instruments with radiometric upgrades, such as ProVal, and also by upgrading the three-stream radiative transfer model to a full one as Hydrolight.

It is not clear why there seem to be no configuration with Fchl providing the vertical profile for a_ϕ , Fdom for a_{CDOM} and b_{bp} for b_{bp} and NAP. If this one does not work well, please explain why you think it does not? What are the implications?

The final modelling configuration is precisely the one that you just described. For a_ϕ we consider Chl-shape, for a_{CDOM} the fDOM profile, and for a_{NAP} and b_{bp} the $b_{bp}(700)$ one. This configuration is also the one giving a much higher skill compared to Case 1 descriptions, following Chl as a proxy. We will underline this in the updated text, as well as add tables that should help clarify this point and remove any doubts.

Comments from the attachment

L20: Isn't this well known?

It is, however PAR models and Case 1 parameterizations are still largely at play, notwithstanding technical improvements of both in-situ and remote sensing technologies. Thus we felt it was important to underline the necessity to include multi-spectral methods also in numerical modelling.

L75: Not clear why? So what if they are shallower than the DCM?

It's not a problem but we kept it for consistency. The aim of these simulations was to capture all the features along the water column. As we are describing them with Chl profiles, especially in Case 1 configurations, we wanted to consistently follow the maximum range of DCM values in the Mediterranean. We could have kept them, but chose to perform this first test with a consistent depth range among the analysed profiles. We are not saying the inclusion of shallower profiles would have been wrong, but it's just an internal consensus for the scope of this study.

L77: Why?

BGC-Argo radiometric measurements have a 20 cm frequency in the first meter. If we saw that

the first 5 points of the profile went missing, we got suspicious of the overall quality of the profile, even though it passed as "good" in terms of the standard QC. When calculating K_d values, the absence of a sufficient number of points close to surface depths resulted in a lower performance of the curve fit. We decided to work with less profiles for which we saw robust fits and dismiss the ones where there were fewer values at the surface. We had to apply additional criteria *a-posteriori* as the standard QC of radiometry still retained profiles with noisy behaviour and dubious depth shapes, which made it difficult to obtain sensible K_d for the analysis. We attach an example of dubious profiles which passed the official radiometric QC procedure, Fig.2 and Fig.3. In the first figure, the shape of E_d at 412 nm caused problems in calculating K_d due to constant values from the surface to almost 10 m depth. The second plot has two different issues: a spike at 412 nm with increasing values, which was somewhat removed with an additional QC (running mean and median filters), whereas at 490 nm there seem to be constant values at the first 10 meters which like in the first example caused issues in computing the depth derivative.

L79: Aren't you forcing a 'good' agreement by deleting profiles where there isn't one?

As already explained in the previous point, this criterion is explained by the noisy or oddly-shaped E_d profiles which generated dubious values of K_d. When calculating K_d values, the absence of a sufficient number of points close to surface depths resulted in a lower performance of the curve fit. It might be argued that we are forcing a "good" agreement, but the truth is that for such a massive data set (we are working with more than 1000 profiles and thus more than 100000 points for the first 150 m) it is better to add some stricter criteria and work with less profiles than vice-versa. This is, after all, a modelling experiment.

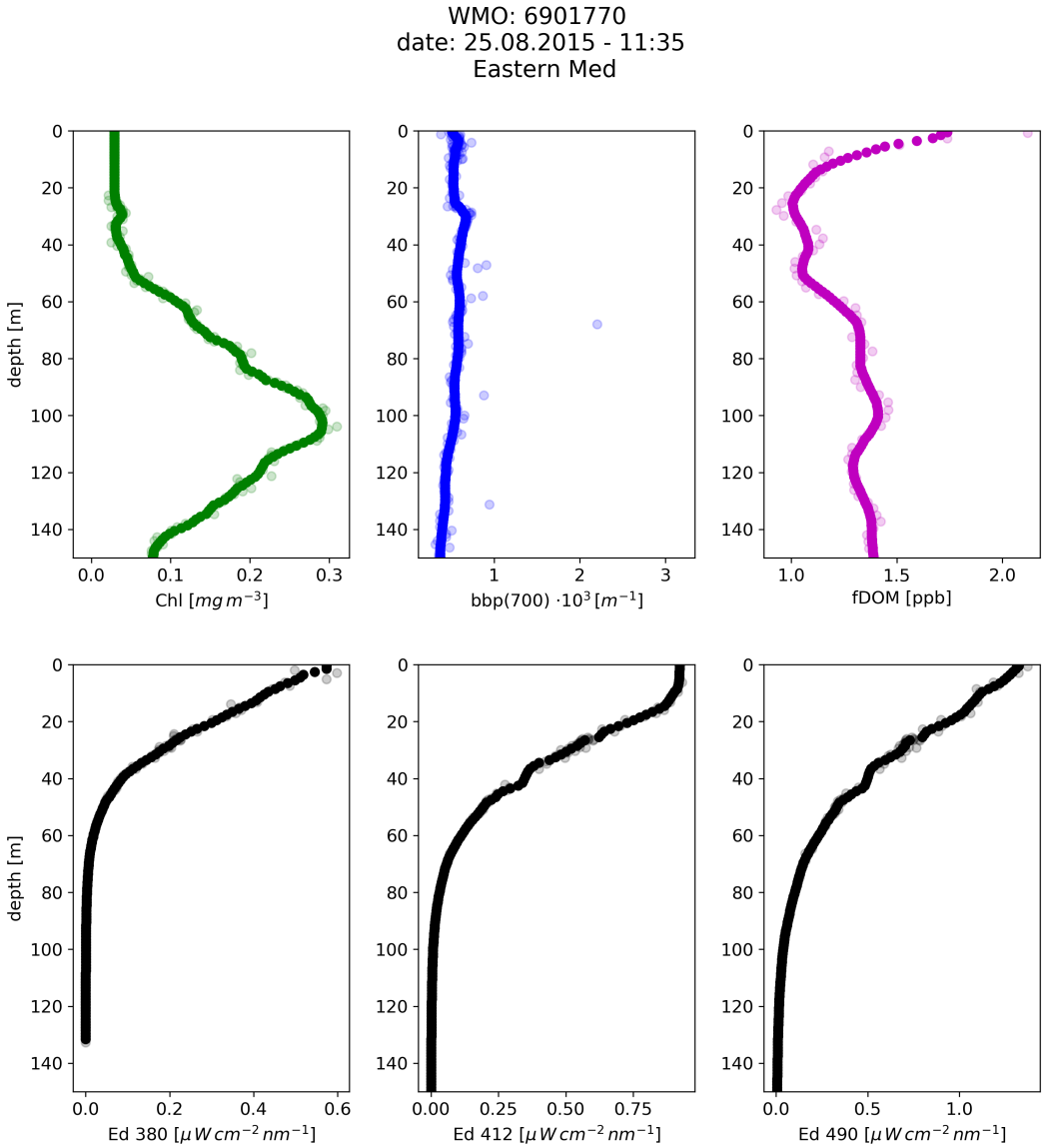

Figure 2: Example of a BGC-Argo profile in the Eastern Mediterranean with radiometric values that caused issues in obtaining K_d and was therefore discarded. Pale dots are values prior the QC applied to this study (running mean and median filter).

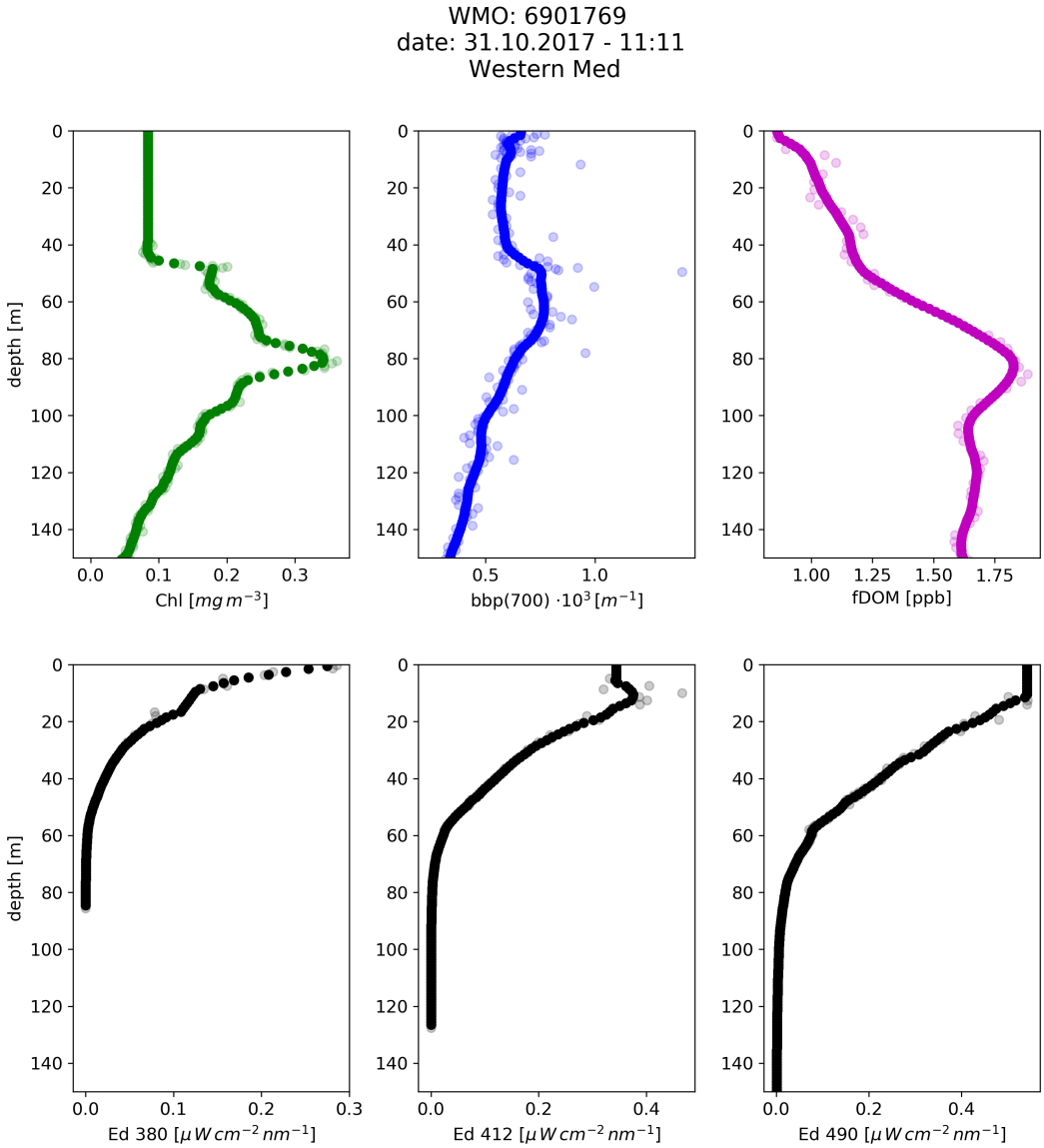

Figure 3: Example of a BGC-Argo profile in the Western Mediterranean with radiometric values that caused issues in obtaining K_d and was therefore discarded. Pale dots are values prior the QC applied to this study (running mean and median filter).

L92: You are assuming that the diffuse cannot contribute to the direct.

This is indeed how the three-stream works and at present there isn't a way to circumvent this problem, unless another radiative transfer model is used.

L93: This model ignores Raman scattering, a major process in oligotrophic waters. You should at least mention it and suggest by how much it may be biasing your results.

Indeed, as discussed above, we included the correction of R_{rs} from Lee et al. [2013].

L137: Not clear. I can see that you set a scaling factor, but this means IOP(labda) is the value at the maxima of a property. How is this value determined?

We acknowledge it's not clear enough and we shall rewrite this part and thus explain it better. The equation shows the assigned depth variability of any IOP considered. We obtain $IOP(\lambda)$ from the assigned equation, any equation solving the IOP spectra - a_ϕ , a_{CDOM}, a_{NAP}, b_p - as described in each following subsection. X(z) denotes the value of any measured biogeochemical or optical quantity (in our case Chl, fDOM and $b_{bp}(700)$) at a specific depth z, which is then divided by the maximum value along the profile. This enables to follow exactly the shape of Chl in case of Case 1 models or a_ϕ, fDOM shape for a_{CDOM} and $b_p(700)$) shape for a_{NAP} or b_{bp}.

L155: Some (Twardowski and others) think of Mason's values to have problems. The latest 'consensus' values for water absorption are in the IOCCG report on absorption.

Indeed, the authors are aware of the report, and studied it carefully beforehand. The Addendum on page 3 [1] on top says that "A recent study by Mason et al. [2016] includes measurements of absorption in this spectral range". We assumed, being a novel study, as well with some of the authors being the ones from older reference publications, that there was no dispute regarding its accuracy. Here we simply quantified the difference between one and the other model, without making conclusions about which one works better. We chose this one as a reference in subsequent simulations due to the fact that is more recent, assuming that the latest technology development enabled more accurate spectral measurements.

L182: In essence you are using the old water absorption and putting the difference into CDOM because the results are not consistent with your expectation. You have to be careful that you are not forcing agreement.

Perhaps it might look as an agreement forcing, however we aimed solely at a procedure consistency. a_{CDOM} values in the original publication from Morel and Gentili [2009] used pure water values of Pope and Fry [1997] for their Case 1 model to eliminate the water contribution. This is why we applied the correction with values from Mason et al. [2016], to account for the difference in the absorption spectra used as an exercise to quantify the difference when performing this test. It was an experiment, we are not trying to say that this is closer to "reality".

L196 : Are you assuming phytoplankton and NAP do not absorb at 380nm. If you are, say so. If not, you need to explain how you get from $K_{bio}(380)$ to $a_{CDOM}(380)$.

This is indeed a very good point, which raised some doubts also among the authors. In the first version of the manuscript we assumed that the contributions of NAP and phytoplankton are much smaller. As discussed in Organelli and Claustre [2019], there are some previous studies in the clearest oligotrophic world oceans that have shown that CDOM dominates the light absorption budget at 380 nm (pg. 6 of the paper). In the absence of coincident light absorption data to prove this statement, other possible sources that affect light attenuation in the UVs, such as light absorption by mycosporine-like amino acids and NAP, can be excluded or considered negligible. NAP light absorption at 380 nm contributes less than 20% to total non-water absorption in clear oligotrophic waters [Bricaud et al., 2010]. As we are aware that this is not at all correct, it's the best shot in the absence of a_{CDOM} from fDOM data which do not exist for the current sensor configuration as far as we know.

However, now that you sensibly raised this point, we decided to run an additional set of simulations by changing the relative contribution of $K_{bio}(380)$, i.e. by assigning a factor ranging between 0.5 and 1 and thus assessing the relative contribution of such a model, thus leaving some uncertainty in the method to use $K_{bio}(380)$ as a proxy for a_{CDOM} only. Results show that in terms of E_d, the impact is negligible (the performance increases a bit at 412 nm at the expense of 380 nm), however by looking at the monthly climatological scatter plots in terms of K_d, modelled values approach much more closely the measured ones. Of course we are aware that we do not

[1] https://ioccg.org/wp-content/uploads/2020/09/absorption_protocol_final-incl-cover_rev.pdf

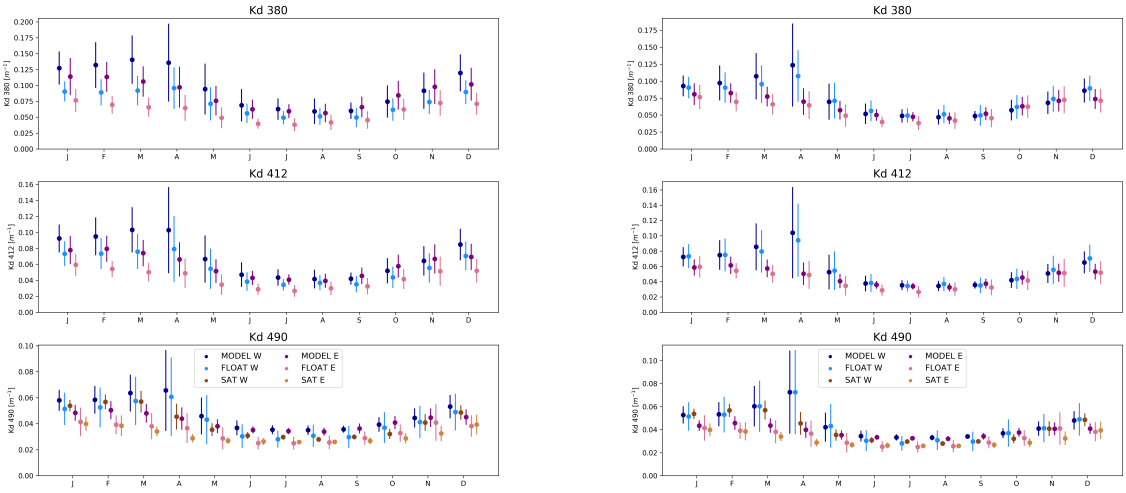

Figure 4: Monthly climatology of K_d values with $a_{CDOM}(380) = K_{bio}(380)$ * f, where f=1.0 (left) and f=0.5 (right).

include Raman which has a certain impact, however we are comparing also with remote sensing products and achieve the greatest 3-platform consistency when halving $K_{bio}(380)$, i.e. using the factor 0.5 instead of 1.0, Fig.4 - right. We will add this part in the updated text, hoping to make it clearer to the readers as well. Thank you, it was a very useful observation, which we think will increase the study's quality, as well as enable to achieve some sort of closure!

L202: what does this light level represent? Surface, ML, DCM? Be clear.

The experiments were carried out with strains of organisms adequate for surface applications, with several of them suitable for both surface and ML. For more details you can see Organelli et al. [2017]. We will include this information in the text accordingly.

L223: This is only true if Dmin=0 and Dmax=infinity, See Boss et al., 2001. Applied Optics. Original derivation is by Volz, 195x cited by van de Hulst.

OK, we will include this explanation in the text, thank you!

L230: These high values imply presence of very small particles that Organelli et al. has shown are an artifact of assuming Mie.

We will include this explanation as well, thank you!

L233: Note that Antoine et al. [2011], claims this ratio has a spectral change similar to λ^{-1}. Did you implement that?

We do not use the spectral dependency of the backscattering ratio. It is a constant value. As for the remark relative to the paper Antoine et al. [2011], pg. 5 it says: "The particle backscattering coefficient at 555 nm is extrapolated from the measurements at 589 nm **assuming** a λ^{-1} spectral dependency. For both data sets, the spectral dependency of $b_{bp}(\lambda)$ is expressed through the slope c, ..." which we then estimated between 0 and 4. It seems to the authors of this paper that the λ^{-1} assumption was used only to deduce $b_{bp}(\lambda)$ values at another wavelength.

L256: All these results can be condensed in a table, making it clearer and easier to compare.

Thank you, we condensed them below each of the barplots, hoping to make the manuscript more readable.

L267: Why do you say so? Just because something is published does not say it is so. Only after other replicate these measurements will we have confidence in them. Also, these measurements do not account for any absorption by ions be it from oxygen to nitrate etc'.

Please see reply above for comment on L155. Perhaps the text suggested that the authors claim this reference to be more accurate, we will modify the explanation accordingly.

L271: How can that be? Salinity and temperature corrections are mostly impacting in the NIR?

They do have a slight impact also in the other parts of the VIS. It's not large in absolute terms, but relatively speaking yes. Be aware that we had so many points of validation (between 100.000 and 170.000) that resulted in a small, but significant improvement. We are modelling more than 1000 profiles at the same time with 150 points each.

L278: This suggest that you cannot really separate NAP and CDOM and that you may want to lump them to a single term. Using 0.0087 suggest you can really tell the different of ± 0.0001 m^{-1} absorption. I don't think you can and suggest you use 0.009 or even 0.01 to reflect the accuracy of the measurements and their variability.

We might do so, but we are using different vertical shapes. We believe that it's not to our advantage to lump them together because we assume that NAP has the shape of $b_{bp}(700)$ and CDOM that of fDOM when not using Case 1 models, knowing that detritus and CDOM do not have the same vertical variability nor dynamics. We also prove that the biggest contributor to the model performance improvement is the shape, not the range of values chosen. This is why we are insisting on keeping them separate because it's the strongest point of this study. We picked the minimum value (following the exact value from the paper for consistency) because the measurements in Babin et al. [2003] were done at the coastal areas. Floats are in the open sea. We could have invented the relative proportions of each of the IOP's contributions and ran for all possible combinations, but this is not the aim of the paper, as there are inherent limitations to the three-stream model itself. We are assessing the state-of-the-art IOP regional (when available) models and plugging them in the RTM, then estimating the relative contribution by each. It's an experiment based on the knowledge that has been available so far.

L281: The measurements were done at 700nm and the skill is in the UV? How come? Isn't that very sensitive to your choice of eta? And since eta can vary from 0 to 4 that could be all over the map...

Perhaps we weren't clear enough and will try to rewrite this part. The shape of the model follows the vertical profile of $b_{bp}(700)$. This has nothing to do with the spectral response. We are using $a_{NAP}(\lambda)$ spectra, $b_{bp}(700)$ is just a metric for its depth variability. The skill in the UV is in terms of $a_{NAP}(\lambda)$ IOP models, not $b_{bp}(\lambda)$.

L289: Does this mean you are using the vertical profile of $b_{bp}(700)$ for all the IOPs? Is this sensible? This is very confusing and you need to describe the different cases much more clearly, best if using graphical means to convey what you did.

No, we use the shape of $b_{bp}(700)$ here only for the NAP model. The only other example is when we test particle scattering model with either $b_{bp}(700)$ or Chl shape. We will change the sentence to remove any doubts, as well as add the tables for clarity.

This likely means that the shape of Chlorophyll from Fchl + correction is problematic. This is a major result.

It means that for CDOM models fDOM works better than Chl. The model behaved better when using fDOM shape instead of Chl for the absorption of CDOM, after the shape has been smoothed to remove spikes.

For sure there are uncertainties in the fluorescence-to-Chl conversion, but this is impossible

to quantify without HPLC data that would be collected synoptically for at least some profiles.

The fluorescence-to-Chl conversion factor follows Roesler et al. [2017] (i.e. with an average value of 2). The values of the conversion factors for the Mediterranean Sea are however between 1.6±0.3 and 1.7±0.2, thus we checked also the model sensitivity by modifying the obtained Chl accordingly. We achieved this by multiplying the profile by xCHL = $\frac{MeanF}{MedF}$, where $MeanF = 2$ and $MedF$ is between 1.3 and 1.9. This would result in xCHL range between 1.05 and 1.45. However, it didn't result in a major change in the simulation outputs and was hence not shown in the manuscript, especially when we use Chl only for the a_ϕ model. Apart from that, it's not possible to account for photophysiology. Not without a proper biogeochemical model, but even then we would need to have some additional in-situ (quasi-)synoptic measurements whenever available to validate this with.

Not clear how MLD relates to the optical parameters described here.

In Section 2.3.3. at line 190, following Organelli and Claustre [2019], we used both the MLD and the first optical depth to calculate $K_d(380)$. Here we simply say that we show the skill when using only the first optical depth because this range proved to work better than from the MLD calculation.

L302: In the spectral range where you work S matters only for b_{bp} (and do not vary much if you fixed it at a mean salinity) and T effect are not significant. Am I wrong?

Indeed, you're right in assuming that we could have chosen to work with mean values of T and, we checked the performance and it doesn't affect it. However, initially we didn't fix it at a mean salinity. We use the point-by-point values of both, but we're not saying it's mandatory.

L319: Is this spectrally constant? Didn't Antoine suggest λ^{-1} (many others do think it is better approximated by a constant) but this needs to be clarified.

As discussed above, we corrected this in the last version of the manuscript, which resulted in the following : the most adequate scattering model was a non-Case 1 type, i.e.derived from $b_{bp}(700)$ measurements following equations 21 and 22, with a maximum backscattering ratio $\widetilde{b_{bp}}$ of 0.015 and a spectral slope η of 3. Yes, the backscattering ratio is spectrally constant. We removed the (λ) in eq. 22 - left side, which might have confused the readers.

L326: Iterate what model output you use to compare with R_{rs} and K_d

Thank you, we will add this in the updated version as well.

L327: 490 nm is likely the most affected by Raman

Yes, indeed, please see the clarification above, with Fig.1.

L345: Didn't you make it such?

True, but as mentioned above, this is the only option available for modelling a_{CDOM} when lacking proper measurements.

L350: You forced a_ϕ=0 at 380, which is not consistent in observation of surface waters of region with strong radiation where MAAs are important.

No, we didn't, as already explained at length in the major remarks' response. Whenever a_ϕ was accounted for, it was included at all wavelengths, as for all the other IOPs: whenever any IOP model was tested, it was done so for the whole spectrum considered.

L385: The need to add S for bb is well known from the work of Morel. The need to add T is new and I am not sure justified. Can you provide data using a mean temperature vs. the measured temperature to convince me that it is needed for UV and blue wavelengths?

It's the same skill if you take the mean value along the profile or else the exact values of T and S. The improvement is referring to the increase in skill when actually including T and S data or not, and not the shape itself.

L389: This means that NPQ correction or otherwise relation off fluorescence to a_ϕ maybe more complex.

For sure, the photophysiological response cannot be captured, and we cannot estimate the uncertainty relative to NPQ. We used a range of difference fluo-to-Chl conversion factors, which resulted in a negligible impact on the model skill and was thus not shown.

L397: The current measurements do not allow for separate functional group. Here you parametrized them as function of chlorophyll based on another study but could no evaluate it.

Yes, correct, this sentence was referring to future directions.

Figure 9: How many (N=?) used for each?

We will add the following summarizing table:

month	West	East
1	58	42
2	49	48
3	86	69
4	58	70
5	75	110
6	43	48
7	32	38
8	19	26
9	22	16
10	17	22
11	31	25
12	70	42

References

Zhongping Lee, Chuanmin Hu, Shaoling Shang, Keping Du, Marlon Lewis, Robert Arnone, and Robert Brewin. Penetration of UV-visible solar radiation in the global oceans: Insights from ocean color remote sensing: PENETRATION OF UV-VISIBLE SOLAR LIGHT. *Journal of Geophysical Research: Oceans*, 118(9):4241–4255, September 2013. ISSN 21699275. doi: 10.1002/jgrc.20308. URL http://doi.wiley.com/10.1002/jgrc.20308.

John D. Mason, Michael T. Cone, and Edward S. Fry. Ultraviolet (250–550 nm) absorption spectrum of pure water. *Applied Optics*, 55(25):7163, September 2016. ISSN 0003-6935, 1539-4522. doi: 10.1364/AO.55.007163. URL https://www.osapublishing.org/abstract.cfm?URI=ao-55-25-7163.

A. Morel and B. Gentili. The dissolved yellow substance and the shades of blue in the Mediterranean Sea. *Biogeosciences*, 6(11):2625–2636, November 2009. ISSN 1726-4189. doi: 10.5194/bg-6-2625-2009. URL https://www.biogeosciences.net/6/2625/2009/.

Robin M. Pope and Edward S. Fry. Absorption spectrum (380–700 nm) of pure water II Integrating cavity measurements. *Applied Optics*, 36(33):8710, November 1997. ISSN 0003-6935, 1539-4522. doi: 10.1364/AO.36.008710. URL https://www.osapublishing.org/abstract.cfm?URI=ao-36-33-8710.

Emanuele Organelli and Hervé Claustre. Small phytoplankton shapes colored dissolved organic matter dynamics in the north atlantic subtropical gyre. *Geophysical Research Letters*, 46(21): 12183–12191, 2019.

Annick Bricaud, Marcel Babin, Hervé Claustre, Joséphine Ras, and Fanny Tièche. Light absorption properties and absorption budget of southeast pacific waters. *Journal of Geophysical Research: Oceans*, 115(C8), 2010.

Emanuele Organelli, Caterina Nuccio, Luigi Lazzara, Julia Uitz, Annick Bricaud, and Luca Massi. On the discrimination of multiple phytoplankton groups from light absorption spectra of assemblages with mixed taxonomic composition and variable light conditions. *Applied Optics*, 56(14):3952, May 2017. ISSN 0003-6935, 1539-4522. doi: 10.1364/AO.56.003952. URL https://www.osapublishing.org/abstract.cfm?URI=ao-56-14-3952.

David Antoine, David A Siegel, Tihomir Kostadinov, Stéphane Maritorena, Norm B Nelson, Bernard Gentili, Vincenzo Vellucci, and Nathalie Guillocheau. Variability in optical particle backscattering in contrasting bio-optical oceanic regimes. *Limnology and Oceanography*, 56(3): 955–973, 2011.

Marcel Babin, Dariusz Stramski, Giovanni M Ferrari, Herve Claustre, Annick Bricaud, Grigor Obolensky, and Nicolas Hoepffner. Variations in the light absorption coefficients of phytoplankton, nonalgal particles, and dissolved organic matter in coastal waters around europe. *Journal of Geophysical Research: Oceans*, 108(C7), 2003.

Collin Roesler, Julia Uitz, Hervé Claustre, Emmanuel Boss, Xiaogang Xing, Emanuele Organelli, Nathan Briggs, Annick Bricaud, Catherine Schmechtig, Antoine Poteau, et al. Recommendations for obtaining unbiased chlorophyll estimates from in situ chlorophyll fluorometers: A global analysis of wet labs eco sensors. *Limnology and Oceanography: Methods*, 15(6):572–585, 2017.