Reviewer 2

We thank the reviewer for having highlighted some important key points which we found very helpful in improving the quality of the text. We corrected the typos and kept the Major Points and Specific Comments sections below, to which we reply point-by-point. In the updated version of the manuscript, the changes based on the reviewer's suggestions can be tracked in **blue/violet**.

Major Points

Some of the method details and corresponding results are a bit hard to follow. Sometimes it's unclear what exact IOP parameterization is used where, and what parameterization the different results in the text are referring to. I think this could be simplified with a table or two summarizing these details.

We agree, and it has been suggested also by another reviewer. We included tables and rewrote the text, hoping to make it clearer and more easily readable.

Maybe I'm missing something in your methodology completely, but it is unclear to me why you test the individual OSM (optically significant material) constituent IOP models separately. You know that for your float data, you're never in pure water or pure water + 1 OSM. You're almost always going to be in your 6th IOP model group (i.e. all constituents). So you need to test your different individual OSM-IOP model parameterizations within this 6th IOP model group because it is the total IOPs that are important. For example, when testing your a_{NAP} model parameterizations (your 2nd IOP model group), you may get a better result (i.e. an irradiance that is closer to the measured value) for one of the parameterizations that is actually less correct. This is because that parameterization could be less correct in a way that is more correct in terms of the actual total IOP e.g. in this NAP case, perhaps your NAP model is incorrect, but it accurately captures CDOM, and CDOM is the dominant OSM in that sample, so overall, the total absorption is more correct.

It's a good point to address, and it has been actually one of the major debates among co-authors as we weren't sure whether to demonstrate the skill by "deactivating"one IOP at the time from the final configuration or vice-versa, like it was decided upon at the end. We agree that it's never the case that we have just one IOP present, but the aim of this analysis was to approach a more "realistic" situation only after a proper IOP model for each OSM, where the choice of the latter was conditioned with the highest skill in terms of a point-by-point match-up. The scope of the adopted approach, i.e. activating one IOP at the time, was twofold: first, as explained above, to choose a model which within that same OSM gives the best performance, and second, to quantify the impact each of the model range has at each of the tested wavelengths. It was therefore not aimed at saying that this is close to a realistic in-water propagation, but rather try to see how much does each of the OSM contribute to the relative improvement compared to the base. For example: how much only the inclusion of CDOM at specific wavelengths contributed to the light absorption versus NAP? What's the impact of different vertical shapes, e..g. Chl versus $b_{bp(700)}$ in case of NAP simulations? It enabled us to have a clearer picture on how much each of the OSMs absorbs/scatters light at a specific wavelength and their importance in terms of the relative contribution, also by assessing the impact of different depth shapes.

Specific Comments

L7: Lefering et al (2020) used model, satellite and glider data for the type of three-platform comparison done here. [Development of bio-optical model for the Barents Sea to quantitatively link glider and satellite observations, Philosophical Transactions A: Mathematical, Physical and Engineering Sciences https://doi.org/10.1098/rsta.2019.0367]

Thank you, none of us was aware of the paper. We changed the sentence accordingly.

Fig 1 and 2: units missing on the colour bar – number of floats? Number of days with profiles present over the 5 year period?

Indeed, another reviewer suggested to merge the two figures, thus we followed all pointers, which will be included in the next version of the manuscript.

L75: you haven't defined K_d yet

We modified the sentence accordingly, thank you!

L75: The details on the K_d calculation are not clear. It seems like you are doing a depth averaged K_d rather than taking the derivative of the Ed profile (to get a profile of K_d). If so, what depth are you averaging over? Why do you need to have the first Ed measurement shallower than 1m?

K_d was calculated with a non-linear least squares fit procedure. Values were obtained by fitting an exponential function to the E_d profile (either modeled or measured by floats). The depth range considered for obtaining K_d is the first optical depth, i.e. the e-folding depth of E_d at a specific wavelength. The existent quality-control procedure for radiometric quantities still retains a noisy behaviour, and it resulted in profiles for which it was difficult to obtain sensible values of K_d. For this reason we needed to discard them *a-posteriori* by adding a few criteria. One condition demanded values present also near the surface, and the other limited the maximum divergence of K_d-calculated and measured E_d values. The former was chosen for the main reason that we are mostly interested in profiles which have as many near-surface measurements, as we're later linking them with remote sensing quantities. If we saw that the profile had removed all 5 values in the first 1 m (Argo has a 0.20m frequency for the first 1 m in terms of radiometry), we got suspicious of the overall quality of the profile, even though it passed as "good" in terms of the standard QC. The latter criterion is on the other hand explained by the noisy 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. For better clarity, we attach two examples of dubious profiles which passed the official radiometric QC procedure, Fig.1 and Fig.2. 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.

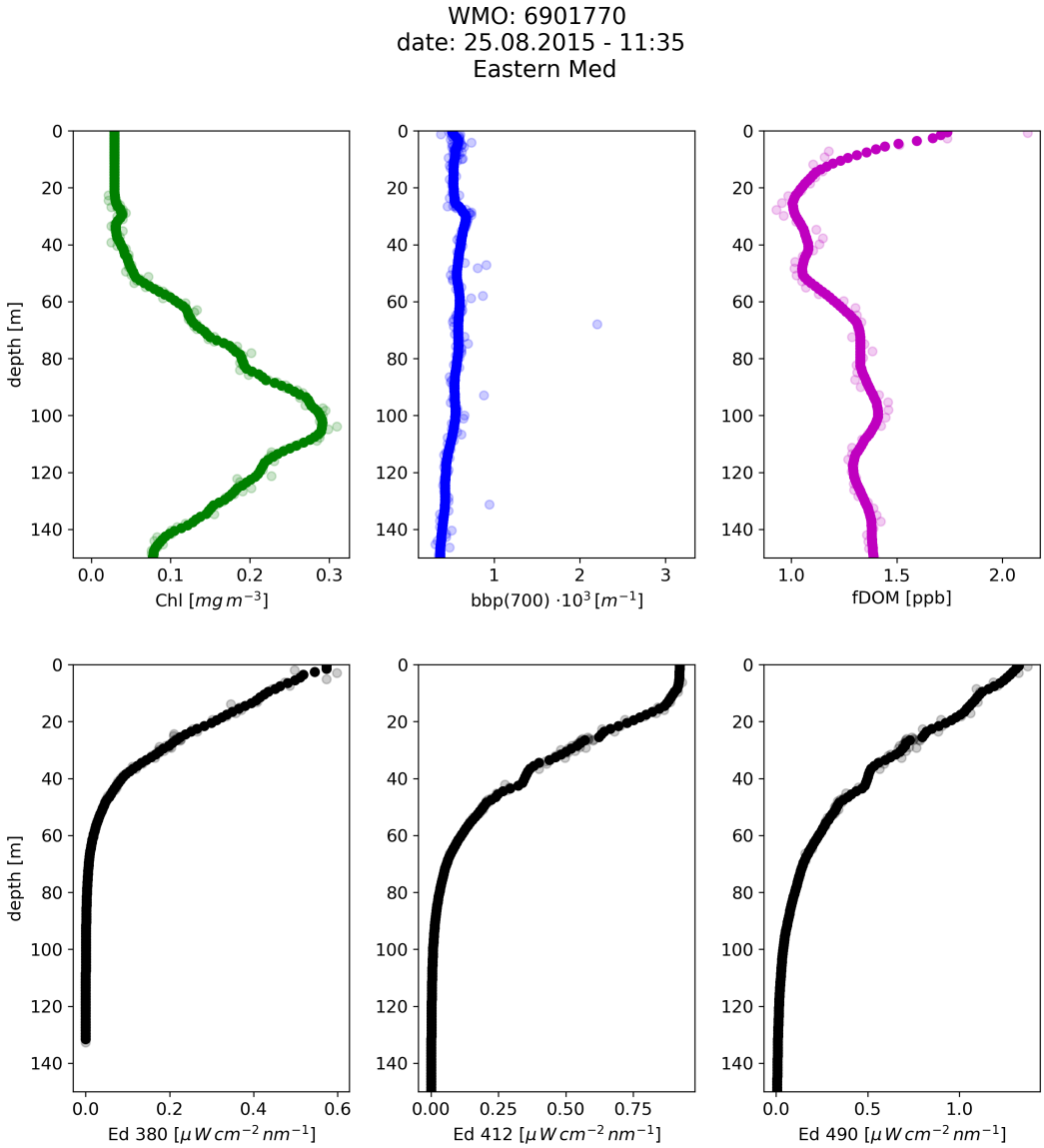

Figure 1: Example of a BGC-Argo profile in the Eastern Mediterranean with radiometric values that caused issues in obtaining K_d and was therefore discarded. Pale dots are values prior the QC applied to this study (running mean and median filter).

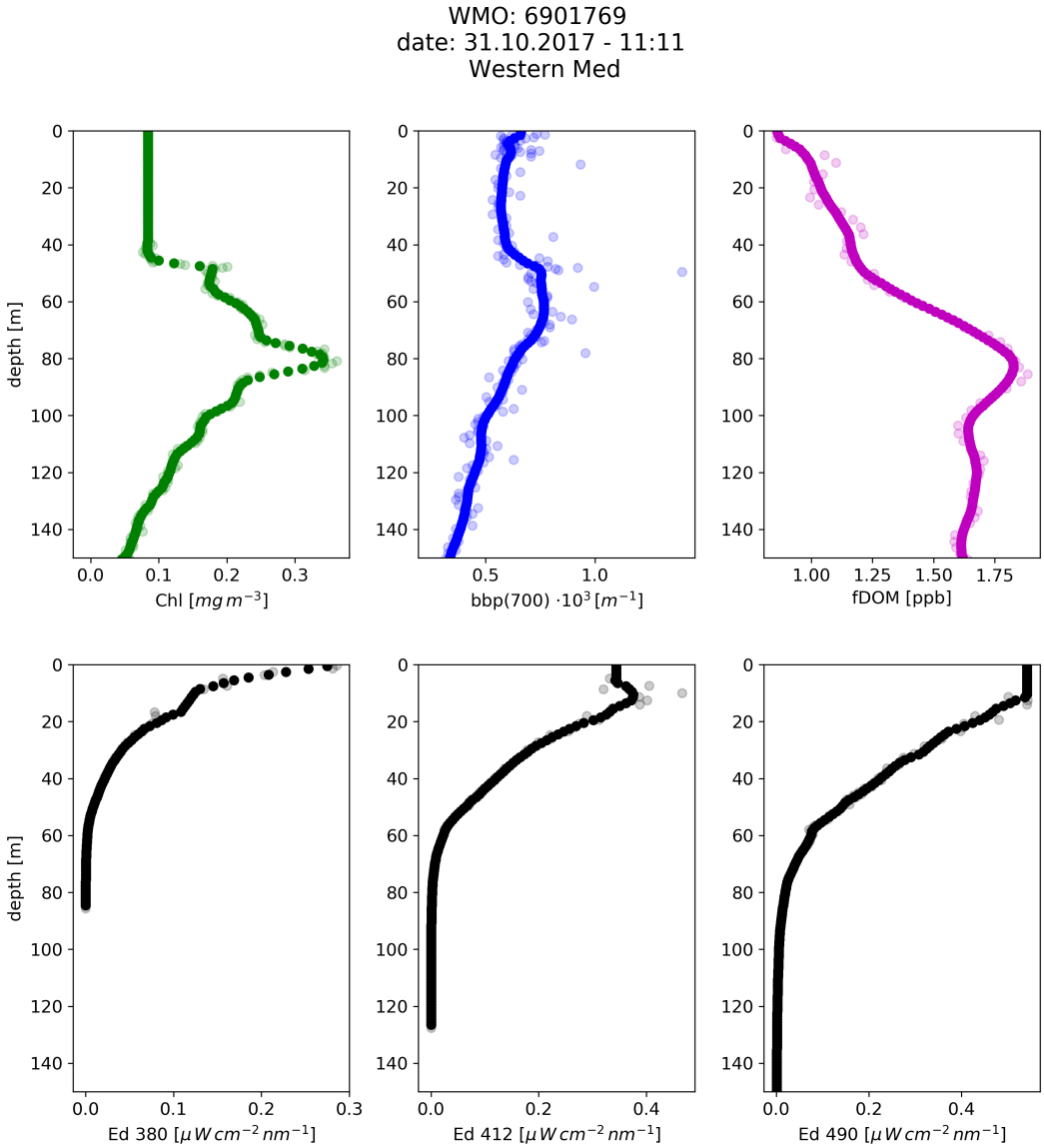

Figure 2: Example of a BGC-Argo profile in the Western Mediterranean with radiometric values that caused issues in obtaining K_d and was therefore discarded. Pale dots are values prior the QC applied to this study (running mean and median filter).

L78: Are you saying that you tested all your different model parameterizations but then just removed the results that had an Ed difference of greater than 30%? What's the justification for this?

With this condition we tested the irradiance profiles from in-situ measurements, meaning that first we discarded the profiles, and only then continued with the in-water simulations. As mentioned in the reply above, we had to further limit the number of chosen profiles as the current quality-control procedure still retained profiles with noisy behaviour or questionable shapes which prevented us from obtaining K_d values from the non-linear least squares fit procedure.

L95, 99: missing backscattering coefficients

Thank you, we modified the lines accordingly.

L142: What is the "this" at the start of the sentence "This is however present…" Can you please reword to make it clearer?

Thank you, we rewrote the sentence, hoping that it's clearer now.

Eq 16 / L177: I don't understand this. Why are you estimating a_{CDOM} using Pope and Fry [1997] when you are using Mason et al. [2016] elsewhere?

The Morel and Gentili [2009] model back then 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 spectrum used. It was just an additional test to check the difference in the performance when applying this additional correction due to the modified pure water absorption values.

L178-179: in the text you have a_{ORIG}, but in Eq 16 you have a_{wORIG} – can you please make it consistent?

We corrected that as well, thank you.

L195: No other OSMs affecting K_{bio}? Can you expand on this in the text please.

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. We are more confident in this result also due to the fact that we are comparing our results with remote sensing products as well, and achieve the greatest 3-platform consistency when halving $K_{bio}(380)$, i.e. using the factor 0.5 instead of one, Fig.3 - 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!

Section 2.3.4: How do the min and max Chl values affect the results? You didn't see any features/performance issues at the low and high ends of your data?

We checked the range of Chl values in our data set. Only 5 profiles out of 1126 had values with maximum Chl values above the range (i.e. 5.71, 5.77, 5.82, 5.96, 5.53 $mg\,m^{-3}$ in the North-Western Mediterranean during spring blooms), thus we simply followed the limits of the algorithm by correcting minimum and maximum values in such way that were in the range of acceptable values. We plotted outputs and haven't observed any features. We thought of underlining this range of values to name one of the inherent limitations of this algorithm, although it seems to correspond rather well to the range observed from BGC-Argo measurements. Perhaps we should mention that

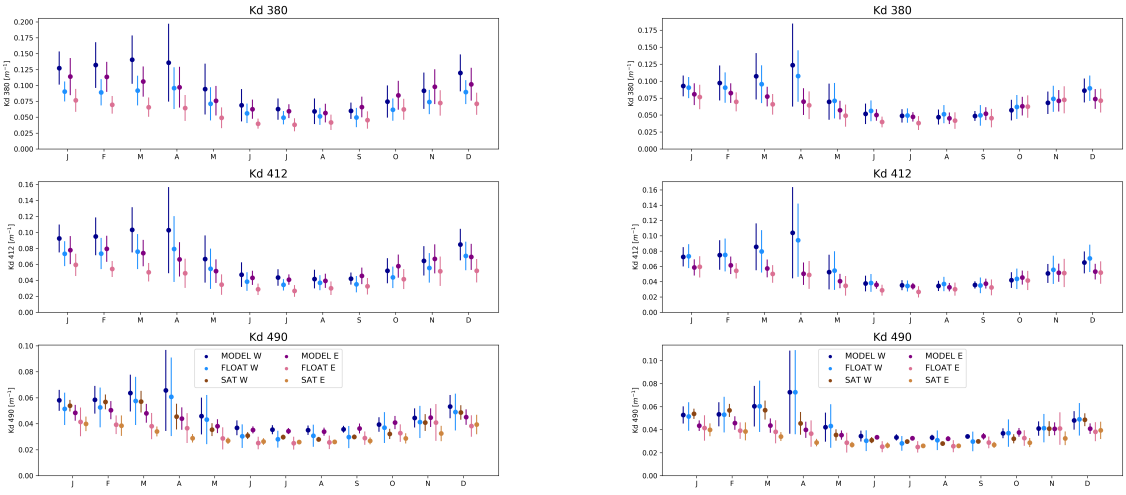

Figure 3: Monthly climatology of K_d values with $a_{CDOM}(380) = K_{bio}(380)$ * f, where f=1.0 (left) and f=0.5 (right).

we also carried out some sensitivity tests in terms of the fluorescence-to-Chl conversion factor, which follows Roesler et al. [2017] (i.e. with an average value of 2). The values of the conversion factors for the Mediterranean Sea range 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.

Eq 20: using Chl as a b_p model – what about NAP?

Indeed, we tested both in terms of shape, but not as a proxy for the b_p estimation in the equation. See line 227: "both Chl and $b_{bp}(700)$ shapes were taken into consideration to account for the depth variability".

Are all IOP models derived for the "surface" and then extrapolated? What's the "surface"?

Surface values are the ones taken at the shallowest depth from float measurements. We've written this sentence under eq.24, but will be specified more clearly in the updated version.

Fig 6 and L310: what configurations are used for each of the OSMs here? The best a_{NAP} and a_{CDOM} results from Figs 4 and 5? You described a range of different scattering models – which one is shown in this figure?

The configurations used are the models that gave us the highest skill in each of the OSM IOP simulations. For NAP and CDOM those are indeed the best models in Figs. 4 and 5 of the paper. For b_p we plot the model which also gave the highest skill, although this could not be quantified in terms of E_d or K_d, but rather with R_{rs}, as shown in the next subsection. As explained also to another reviewer, we are least confident in the estimation of scattering models. 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 (and of course even less so on E_d), 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.

L311-312: Suggest including that these decreases in RMSE and bias are in reference to the pure water simulation.

Thank you for having pointed this out, we've added the clarification at the end of the sentence.

L323-326: what run is this paragraph referring to? I'm assuming the run with water + all OSMs. If so, what configurations are used for each of the OSMs? Can you please include those details?

We've rewritten this part, hoping to make it clearer now. We've also included tables to make the simulations descriptions, as well as results, more legible.

Fig 7: The y-axis label isn't defined anywhere

Thank you, we changed the plot and wrote "model" instead.

Fig 8 – 10: what is the depth bin or range for the model and float data shown in these figures?

We calculated the K_d values for the 1st optical depth, i.e. the e-folding depth at each wavelength, both for model and in-situ values.

I'm not sure I'm following the a_{NAP} argument presented in section 3.2. You state the K_d comparisons improve when you remove the a_{NAP} component – this makes sense because of your assumption that K_{bio} was only CDOM driven. Then you say you need the a_{NAP} component to get the best R_{rs} retrievals, but offer no explanation. What's going on here? It seems your R_{rs} model is inconsistent with your K_d model?

This is indeed the least quantitative, and most descriptive part of the paper. As explained above, in the absence of upwelling radiometric measurements, the quantification of scattering models was possible only in terms of R_{rs}. We wanted to show all possible AOPs available to validate our model choice, and indeed in terms of K_d, as you rightly wrote, we did overestimate a_{NAP} through K_{bio}, which was now modified. But since NAP have, unlike CDOM, a significant role in terms of scattering, we showed that what might have been the best metric in terms of K_d, which is predominantly proportional to absorption, might not result in the same skill when considering a different metric, such as R_{rs}, which is impacted much more also in terms of backscattering b_b. The fact that R_{rs} is proportional to b_b and inversely proportional to the sum of a and b_b makes the choice of the combination with most adequate IOP models slightly more complex compared to absorption models in the previous section. We will rewrite this paragraph to make the discussion clearer. It should be noted that there are also some inherent limitations to the fact that we are not using a full radiative transfer model like Hydrolight, and instead opt for a three-stream configuration.

References

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.

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/.

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.

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.