Responses to Reviewer #2

"A synthetic optical database generated by radiative transfer simulations in support of studies in ocean optics and optical remote sensing of the global ocean"

Hubert Loisel, Daniel Schaffer Ferreira Jorge, Rick A. Reynolds, and Dariusz Stramski

We appreciate the constructive comments by the Reviewer. Here we provide our detailed point-by-point responses and a description of any actions taken in regard to these comments. The Reviewer's comments are shown in italicized font; our responses follow each comment in normal font. Line numbers and figures indicated in our responses refer to the revised manuscript unless otherwise noted.

The manuscript by Loisel and co-workers presents a synthetic data set of so-called apparent optical properties (AOPs) generated through simulation of the radiative transfer (RT) in oceanic waters. The goal is to help the development and test of inversion algorithms that use these AOPs as input (generally the remote-sensing reflectance, Rrs) and attempt to derive quantities such as the inherent optical properties (IOPs, here the absorption and backscattering coefficients) or biogeochemical / ecological properties such as the phytoplankton chlorophyll concentration (Chl). The Rrs are generally derived from satellite ocean colour observations, although the inversion can also be applied to Rrs derived from in-situ radiometry measurements.

Numerous such data sets exist and the justification for proposing this new one is that the IOPs used as inputs to the RT simulations are made more representative of the real world by using probability distribution functions that are consistent with what global in-situ or satellite data sets reveal.

I guess this is an important feature if the data set is then used in its entirety to, e.g., train ML-based algorithms such as neural networks. In such a case, it is indeed critical that the training data set is as realistic as possible. If the data set is used by expert users who can appropriately select what they need in this simulated data (e.g., they can pick up whatever range of the various input parameters that they think is relevant to their development), then the justification for this new data set is a bit weaker.

In any case, I think it is a very useful data set. I appreciate that the authors have included not only the Rrs spectra but also most apparent optical properties as well as the full radiance distributions. This will undoubtedly make this dataset of a broader interest.

The effort to make the synthetic data set better representative of the real global ocean, which is largely made of clear and moderately clear waters, is also a very good one. As clearly shown by Fig. 6, previous synthetic data sets did not match at all what the global ocean is. They were largely skewed towards high values of most IOPs, which is rather typical of coastal waters. Still, they were often used to develop and test algorithms that are then applied globally. To my opinion, this situation has led to frequent overstatement about the performance of inversion algorithms applied to satellite ocean colour.

Response: We thank the Reviewer for positive comments on our manuscript.

The use of the Hydrolight RT code is a relevant choice, in particular because it allows Raman scattering and Chl fluorescence to be simulated. I would however recommend that the authors make clearer in the paper that modelling Chl fluorescence includes several assumptions so that any algorithm development using the part of the spectrum affected by Chl fluorescence will be largely depending on those assumptions. They allude to this on lines 577-578 (by mentioning that the quantum efficiency of Chl fluorescence was set to a default value) but a better description of this part and of the associated limitations should be included.

Response: We agree and added two references showing that this parameter is variable (from about 0.01 to 0.05). The following sentence has been added in the revised manuscript (lines 568-570): "The quantum efficiency of chlorophyll-a fluorescence, which may exhibit significant variability (nearly 5-fold between about 0.01 and 0.05) in ocean waters (Maritorena et al., 2000; Morrison et al., 2003), was also set to its default value of 0.02 in the HydroLight code."

The selection of possibly realistic combinations of absorption by phytoplankton, coloured dissolved organic matter (CDOM) and non-algal particles is important if users of the data set aim at retrieving these quantities. However, the authors know that RT in the ocean, for given boundary conditions, is entirely driven by only two inherent optical properties: the total absorption coefficient and the total volume scattering function (VSF). It does not matter for RT whether a given total absorption at a given wavelength is generated by various proportions of phytoplankton or water or anything else. The important parameters are therefore the ratio of particulate to molecular scattering (defining then the total VSF) and the ratio of total absorption to total scattering or total attenuation. Therefore, although I recognise that it is useful to see IOP distributions for components like phytoplankton and CDOM (Fig. 2), it would be equally useful to see the distributions of the parameters mentioned above. With what is presented, we cannot be sure that they are well covered.

Response: We agree and following Reviewer's comment we have added in the revised manuscript the information about these ratios (lines 620-627), and a new figure (Figure 8 in the revised manuscript). This addition reads as follows "The radiative transfer is driven mainly by two ratios of IOPs which are the scattering to absorption ratio, $b(\lambda)/a(\lambda)$, which controls the number of scattering events (Morel and Gentili, 1991), and the molecular to total scattering ratio, $b_w(\lambda)/b(\lambda)$, which is the parameter controlling the weighted sum of the particle scattering and molecular scattering phase functions (Morel and Loisel, 1998; Loisel and Stramski, 2000). Figure 8 shows the distribution of these two ratios at 440 nm for the synthetic dataset. The $b_w(440)/b(440)$ and $b(440)/a(440)$ ratios range between about 0 and 0.2 and 0.5 and 10, respectively, which is consistent with previous models developed for Case-1 waters (Figs. 2 and 3 in Morel and Gentili, 1991; Fig. 2 in Morel and Loisel, 1998)."

A few more detailed comments

Title: a data base of what? Maybe that could be said (Although the title is already quite long)

Response: The title has been modified. We added "optical" in front of "database" which we think is adequate, especially that the database consists of numerous optical quantities including various IOPs, AOPs, and radiometric quantities, so it is impractical to attempt to create a title that would list all different categories of optical quantities included in the database.

The entire section 2 is quite dense and not that easy to read. Some sub-headings might help better figure out the various steps followed in generating the IOP set. I guess many details could also be moved into a supplemental section, at least if the Journal allows it.

Response: We agree and following Reviewer's comment have added four sub-headings (four subsections) in section 2 to facilitate reading. However, we believe it is important to have sufficiently detailed description of IOP dataset and how it was created because the database obtained through RT simulations depends critically on IOPs. Readers who may be less interested in the IOP-related details can skip this section with only minor implications to reading or understanding of RT-related section.

Line 115: the use of Zhang et al to calculate the seawater scattering coefficient implies that a temperature and a salinity have been chosen. This should be indicated, and it should be made clear that this data set *cannot be used for freshwater, for instance. The manuscript title says "global ocean" but we know that not all users might be that careful.*

Response: We believe that this point is already clear in the manuscript but for further clarification we added the following sentence in section 5 (Summary): "Because of the use of absorption and scattering properties of pure seawater (assuming the salinity of 35‰) in the simulations, the present database cannot be used for applications to freshwater environments and also special caution should also be exercised for applications when water salinity is significantly less than 35‰ because of decrease in pure seawater scattering".

Paragraph 199-214: the in-situ data bases that have been used should be listed and acknowledged. I know it can be a bit cumbersome because there are many, but a Table would make this efficiently. Citing papers that have previously used these data bases is not the correct way of doing it.

Response: We agree that citing papers that have previously used these databases is not the best approach. We therefore removed such papers from the citation list but added several new citations to papers which describe the original collection of in situ data. Papers that describe the compilation of specific datasets and a list of general database sources used in our study are also provided. The revised description now reads (lines 214-220): "Many in situ data of IOP coefficients used in the present study were collected in previous studies (e.g., Reynolds et al., 2001; Babin et al., 2003; Loisel et al., 2007 ; Claustre et al., 2008; Huot et al., 2008; Stramski et al., 2008; Lubac et al., 2008 ; Loisel et al., 2009; Bricaud et al., 2010; Loisel et al., 2011; Antoine et al., 2011; Neukermans et al., 2012; Uitz et al., 2015; Neukermans et al., 2016; Reynolds et al., 2016; Aurin et al., 2018; Reynolds and Stramski, 2019; Stramski et al., 2019). Some data are described in publications devoted to compilation of various datasets (Valente et al., 2019; Casey et al., 2020) and are included in several databases (e.g., SeaBASS, CoastlOOC, BOUSSOLE, and GOCAD)."

As the Reviewer notes, because of the large number of individual datasets it would be quite cumbersome to provide a complete list, even in the form of a dedicated Table (several pages). As the majority of this assembled IOP dataset (e.g., coefficients for $a_g(\lambda)$, $a_{dg}(\lambda)$, $b_{bp}(\lambda)$) is not used in any quantitative way to develop the synthetic IOP dataset, but instead primarily serves as a means for comparison with the derived synthetic IOPs, we believe a detailed description of all data sources is unnecessary. With regards to measurements of $a_{ph}(\lambda)$, although these data are used to provide realistic spectral shapes of the phytoplankton absorption coefficient, as described in section 2.3 of the revised manuscript the vast majority of assembled data required additional modifications and refinements (e.g., spectral interpolation, extrapolation to the UV) to suit the purposes of our study. In addition, a large number of the initial assembled dataset of $a_{ph}(\lambda)$ (~50%) failed to meet certain criteria required for our use and were subsequently rejected and not included in the study. Finally, the distribution of spectral shapes obtained from these final accepted spectra was further modified to ensure a distribution representative of the global ocean. For these numerous reasons, we feel that a detailed description of all original sources of the in situ IOP dataset would, in addition to being impractical, have little to no value to the reader.

The spectral values of the particulate backscattering coefficient resulting from summing up the phytoplankton and non-algal particle backscattering coefficients derived through Eqs. 8 and 11 should be displayed for some Chl values (or the spectral slope as a function of Chl could be displayed). That is an important parameter and it is unclear how it looks like.

Response: Figure R2-1 below gives some insight into this question by showing a_{ph} vs. the spectral slope of b_{bp}. We have not included this figure in the manuscript but added some relevant text (lines 489-492): "The spectral slope of $b_{bp}(\lambda)$, γ, where $b_{bp}(\lambda)$ is obtained as the sum of $b_{b\text{-}ph}(\lambda)$ and $b_{b\text{-}d}(\lambda)$, has a mean and standard deviation of 1.10 ± 0.34, and exhibits a decreasing trend from oligotrophic (where γ is around -2) to eutrophic waters (where the $b_{bp}(\lambda)$ spectrum is nearly flat). These results are in good agreement with previous studies (Morel and Maritorena, 2001; Loisel et al., 2006; Antoine et al., 2011)."

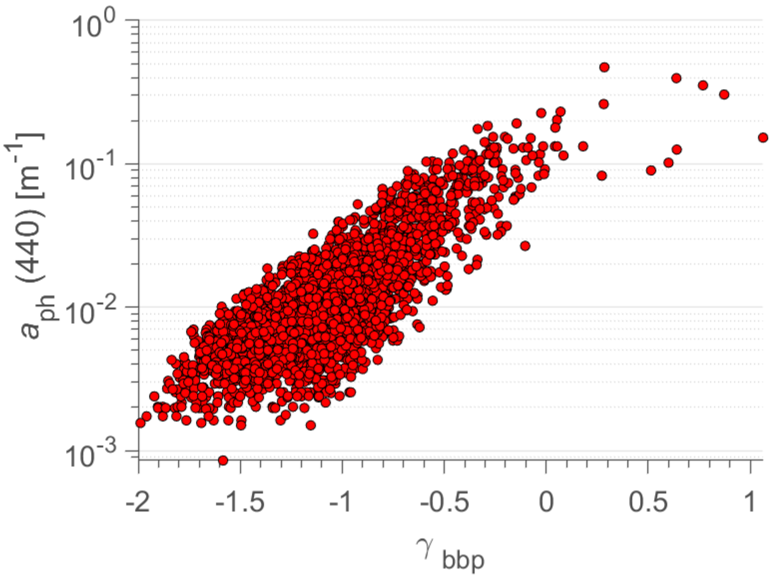

Fig. R2-1. Phytoplankton absorption coefficient, $a_{ph}(440)$, vs. the spectral slope of particulate backscattering coefficient for the new synthetic dataset.

Pages 12-13: not sure I have got the rationale for the P1, 2, 3, 4 parameters and corresponding equations in Table 1. This should be better explained.

Response: The parameters for P_1 and P_2 were chosen to match the relationships between $a_g(443)$ and $a_d(443)$, respectively, vs. $a_{ph}(443)$ that were observed with the in situ dataset used in this study. The P_3 and P_4 parameters were fixed based on the IOCCG (2006) formulations developed for both $b_{b\text{-}ph}$ and $b_{b\text{-}d}$, the two components of b_{bp}. The range of variability for the randomly-generated numbers involved in these P formulations has been chosen to match both the in situ variability observed as well as the satellite distribution. Some explanatory text was added to the revised manuscript and a new column was added to Table 1 which provides references to the sources of these parameters.

Pages 5-9: this is quite long for describing how absorption is modelled. And, there is conversely little about scattering (discussing the phase function only comes at page 14).

Response: As absorption, specifically $a_{ph}(\lambda)$, is the main driver in determining the synthetic IOP dataset and subsequent RT calculations we believe that a detailed description is important for readers. Some additions in text and a new figure (Fig. 8) were added in the revised manuscript to enhance the description of modeled scattering properties in the IOP dataset, i.e., regarding (the spectral slope of b_{bp} (lines 489-492), the b/a and b_w/b ratios (lines 620-627, new Fig. 8), and the particulate phase function (lines 473, 486-488).

Fig. 9 would be advantageously completed by a similar plot of Rrs490/Rrs555 vs. Chl.

Response: Following Reviewer's comment a new figure (Fig. 11) was added in the revised manuscript. This figure shows Chla as a function of the maximum band ratio used in SeaWiFS-specific OC5 algorithm (O'Reilly and Werdell, 2019). A short paragraph was added to describe this figure (lines 690-694).

Another very important parameter to show would be the ratio of CDOM absorption to total non-water absorption vs. Chl. I guess what I am trying to say here is that a more comprehensive assessment of how this data set compares to previous ones and to established Case 1 water bio-optical models would be really useful.

Response: We believe that our responses and several figures provided in response to Reviewer #1, especially Figs. R1-1, R1-4, and R1-5 with absorption data as well as an enhancement of Fig. 7 in the revised manuscript (the addition of panel c), address this comment by the Reviewer.