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

Hubert Loisel[1], Daniel Schaffer Ferreira Jorge[1], Rick A. Reynolds[2], and Dariusz Stramski[2]

[1]Laboratoire d'Océanologie et de Géosciences, Université du Littoral-Côte-d'Opale, Université Lille, CNRS, IRD, UMR 8187, LOG, 32 avenue Foch, Wimereux, France
[2]Marine Physical Laboratory, Scripps Institution of Oceanography, University of California San Diego, La Jolla, California 92093-0238, USA.

Correspondence: Hubert Loisel (hubert.loisel@univ-littoral.fr)

Abstract. Radiative transfer (RT) simulations have long been used to study relationships between the inherent optical properties (IOPs) of seawater and light fields within and leaving the ocean from which the ocean apparent optical properties (AOPs) can be calculated. For example, inverse models to estimate IOPs from ocean color radiometric measurements have been developed or validated using results of RT simulations. Here we describe the development of a new synthetic optical database based on hyperspectral RT simulations across the spectral range from the near-ultraviolet to near-infrared performed with the HydroLight radiative transfer code. The key component of this development was the generation of the synthetic dataset of seawater IOPs which served as input to RT simulations. Compared to similar developments of optical databases in the past, the present dataset of IOPs is characterized by probability distributions of IOPs that are consistent with global distributions representative of vast areas of open ocean pelagic environments and coastal regions covering a broad range of optical water types. The generation of the synthetic data of IOPs associated with particulate and dissolved constituents of seawater was driven largely by an extensive set of field measurements of the phytoplankton absorption coefficient collected in diverse oceanic environments. Overall, the synthetic IOP dataset consists of 3320 combinations of IOPs. Additionally, the pure seawater IOPs were assumed following recent recommendations. The RT simulations were performed using 3320 combinations of input IOPs assuming vertical homogeneity within an infinitely deep ocean. These input IOPs were used in three simulation scenarios associated with assumptions about inelastic radiative processes in the water column and three simulation scenarios associated with sun zenith angle. Specifically, the simulations were made assuming no inelastic processes, the presence of Raman scattering by water molecules, and the presence of both Raman scattering and fluorescence of chlorophyll-a pigment. Fluorescence of colored dissolved organic matter was omitted from all simulations. For each of these three simulation scenarios, the simulations were made for three sun zenith angles of 0°, 30, and 60° assuming clear skies, standard atmosphere, and wind speed of 5 m s^{-1}. Thus, overall 29880 RT simulations were performed. The output results of these simulations include the radiance distributions, plane and scalar irradiances, and the whole set of AOPs including the remote-sensing reflectance, vertical diffuse attenuation coefficients, and mean cosines where all optical variables are reported in the spectral range from 350 to 750 nm at 5 nm intervals for different depths between the sea surface and 50 m. The consistency of this new synthetic database has been assessed through comparisons with in situ data and previously developed empirical relationships involving the IOPs and AOPs.

1 Introduction

The investigation of the propagation of natural light in the ocean can be addressed experimentally through in situ measurements and theoretically through numerical simulations of radiative transfer. The understanding of the relationships between the radiometric quantities (i.e., radiance and irradiances) that characterize the light fields within and leaving the ocean and the inherent optical properties (IOPs) of the water column, boundary conditions at the sea surface (i.e., surface illumination conditions and sea state) and at the ocean bottom (i.e., bottom depth and albedo) requires comprehensive datasets of multiple variables acquired over a broad range of environmental conditions. For example, of particular interest are the relationships between the spectral remote-sensing reflectance of the ocean, $R_{rs}(\lambda)$, which is an apparent optical property (AOP) derivable from radiometric quantities, and the seawater IOPs that are directly linked to various seawater constituents because these relationships form the cornerstone of various applications of optical (ocean color) remote sensing. Recent technological developments and broader accessibility of optical in situ instrumentation have led to significant increase in optical datasets collected across diverse oceanic environments and efforts have been undertaken to merge data from various sources within publicly available databases (e.g., Werdell and Bailey, 2005; Valente et al., 2019; Casey et al., 2020). Although the importance of field data collection across diverse environments cannot be overstated, the existing database compilations are subject to certain limitations. In addition to typical measurement errors, it is difficult to ensure consistent data quality and characterization of uncertainties across all merged data because individual datasets are often obtained with different instruments as well as measurement and data processing methods. Also, even the large databases such as NASA's SeaBASS cannot ensure the balanced representativeness of collected field data in terms of a broad range of optical conditions across diverse ocean environments. In this context, radiative transfer (RT) simulations provide a useful tool to generate comprehensive synthetic databases and complement the existing datasets of field measurements in support of studies in ocean optics and optical remote sensing.

Over the past decades various radiative transfer models that employ different numerical solution techniques have been developed and used to address a wide range of problems related to optics of natural water bodies (e.g., Mobley et al., 1993; Mobley, 1994; Stamnes et al., 2017). Since the 1990s the HydroLight code based on invariant imbedding technique (Mobley, 1989; Mobley et al., 1993; Mobley, 1994) has been among the most commonly used radiative transfer models in oceanographic optics. The HydroLight code solves the scalar (i.e., polarization of light is not included) time-independent radiative transfer equation for a horizontally homogeneous water body, in which the inherent optical properties can vary with depth, under given boundary conditions at the surface and bottom of the water body. The inelastic radiative processes within the water column that include Raman scattering by water molecules, fluorescence of chlorophyll-a pigment, and fluorescence of colored dissolved organic matter (CDOM) can be included in HydroLight simulations.

The radiative transfer simulations with HydroLight code have proven useful for generating synthetic databases of light field characteristics (i.e., radiance and irradiances) within and leaving the ocean and the AOPs derived from the simulated radiometric quantities for various scenarios of seawater IOPs that provide input to the simulations. In particular, as a result of efforts dedicated to inverse bio-optical algorithms and coordinated under the auspices of the International Ocean Colour Coordinating Group (IOCCG Report, 2006), the publicly available synthetic database was generated within the spectral range 400 - 800 nm with a 10 nm resolution for clear sky conditions with three different sun zenith angles (0°, 30°, and 60°), a sea surface state corresponding to a wind

81   speed of 5 m s^{-1}, and 500 different IOP combinations driven by chlorophyll-a concentration, Chla, within the

82   surface ocean layer. The input IOP data included the spectral absorption coefficients of phytoplankton, $a_{ph}(\lambda)$, non-

83   algal particles (also referred to as depigmented or detrital particles which can include various types of particles

84   such as organic detritus, mineral particles, heterotrophic bacteria, and depigmented phytoplankton cells), $a_d(\lambda)$,

85   colored dissolved organic matter (CDOM), $a_g(\lambda)$, and the spectral backscattering coefficients of phytoplankton,

86    $b_{b-ph}(\lambda)$, and non-algal particles, $b_{b-d}(\lambda)$ (λ represents the wavelength of light in vacuum in units of nm and the IOP

87   coefficients are typically expressed in units of m^{-1}). The output parameters provided by those simulations which

88   are available in the public database included the following AOPs: the spectral remote-sensing reflectance, $R_{rs}(\lambda)$,

89   the remote-sensing reflectance just below the sea surface, $r_{rs}(\lambda)$, the irradiance reflectance just below the sea

90   surface, $R(z=0^-, \lambda)$, and the diffuse attenuation coefficient for downwelling plane irradiance, $K_d(\lambda, z)$, at the depths

91    $z = 0^-$, 5, and 10 m (where 0^- indicates the depth just beneath the sea surface).

92   Another synthetic database that is publicly available was developed as part of the CoastColour Round Robin

93   project (Nechad et al., 2015). This project was focused on coastal waters and IOPs were driven by 5000

94   combinations of Chla, $a_g(\lambda)$, and mass concentration of mineral particles. The HydroLight simulations were run

95   from 350 nm to 900 nm at 5 nm intervals for cloudless sky, three sun zenith angles (0, 40, and 60°), and a wind

96   speed of 5 m s^{-1}. The output parameters included in the publicly available database are the water leaving

97   reflectance, $RL_w(\lambda) = \pi\, R_{rs}(\lambda)$, $K_d(\lambda)$, the photosynthetically available radiation, PAR, and the euphotic depth, z_{eu}.

98   Most recently, a synthetic database was also developed by the first NASA PACE (Plankton, Aerosols, Cloud, and

99   Ocean Ecosystem) Science Team where the ocean contribution to the top of the atmosphere radiances were

100   simulated by HydroLight (Craig et al., 2020). These simulations were performed from 350 to 800 nm with a 5 nm

101   step for a cloudless sky, three sun zenith angles (10°, 30°, and 60°), wind speed of 5 m s^{-1}, and a set of 720 IOP

102   combinations driven by $a_{ph}(\lambda)$. The publicly available output of these HydroLight simulations is $R_{rs}(\lambda)$.

103   While these existing synthetic databases have offered valuable information to the ocean color radiometry

104   (OCR) community, especially for the purpose of algorithm development where the ocean AOPs are linked to IOPs,

105   there are several reasons that have motivated the present study aiming at generating a new optical synthetic

106   database. First, the inelastic Raman scattering and fluorescence processes were ignored in the previous RT

107   simulations. These inelastic radiative processes are known to be important for simulating realistic characteristics

108   of light fields within and leaving the ocean, including $R_{rs}(\lambda)$ that is a primary optical quantity used in ocean color

109   remote sensing. For example, Raman scattering by water molecules may have an important influence on light

110   within and leaving the ocean and AOPs, especially in the green and red parts of the spectrum (e.g., Marshall and

111   Smith, 1990; Stavn, 1993; Sugihara et al., 1984; Westberry et al., 2013). Second, the three synthetic databases

112   described above are based on the use of the spectral pure seawater absorption, $a_w(\lambda)$, and scattering $b_w(\lambda)$,

113   coefficients values as defined by Pope and Fry (1997) and Morel (1974) in the visible part of the spectrum,

114   respectively. However, more recent measurements and theoretical considerations provide new recommendations

115   for spectral values of $a_w(\lambda)$ and $b_w(\lambda)$ (IOCCG Protocol Series, 2018; Zhang and Hu, 2009; Zhang et al., 2009).

116   Third, the suites of IOP data combinations and probability distributions of different IOPs that were used as input

117   to previous RT simulations do not appear to match well with the IOP distributions observed in extensive field

118   datasets or satellite-derived datasets representing the global ocean. This issue may have a biasing effect when the

119   synthetic databases are used to develop the optical algorithms based on the AOP vs. IOP relationships, especially

120   when the underlying goal is to represent a broad range of IOPs encountered within the global ocean, even if the

121   primary interest is in open-ocean pelagic environments. Finally, the previous synthetic databases were developed

122   specifically for OCR-oriented studies and the publicly accessible data generally include only the surface

123   reflectances, $R_{rs}(\lambda)$, $R(\lambda)$, $r_{rs}(\lambda)$, and $K_d(\lambda)$ at selected depths. These databases do not include many of the various

124   output variables obtained from RT simulations, such as the various underwater AOPs, which can be useful in

125   supporting a broader range of studies in ocean optics beyond ocean color remote sensing.

126   In this article, we present a new synthetic optical database generated using RT simulations which addresses

127   some of the limitations of similar databases developed in the past. First, we describe the development of the

128   synthetic IOP dataset that is required to run RT simulations. The key roles in this development are played by the

129   measured data of phytoplankton absorption coefficient and desired consistency between the probability

130   distributions of synthetic IOPs and the global distributions based on satellite observations. Following this, we

131   describe different configurations of RT simulations which were performed with the HydroLight code. The next

132   section is dedicated to consistency between the new optical synthetic database and in situ data, including some

133   previously reported empirical relationships. We provide example illustrations of consistency for both the IOP and

134   AOP data. The closing section summarizes the structure of synthetic database files and provides example

135   illustration of one output radiometric variable, the spectral downwelling plane irradiance, calculated with RT

136   simulations.

## 137   2 Development of synthetic dataset of seawater inherent optical properties

138   The scope of the synthetic database generated with RT simulations and the degree of its representativeness of

139   diverse marine optical environments within the global ocean depend most critically on a dataset of seawater IOPs

140   that are used as input to RT simulations. In the present study, our approach to generate the IOP dataset was driven

141   largely by an underlying goal to obtain the probability distributions of IOPs which are generally consistent with

142   the distributions observed in the global ocean dominated by open-ocean pelagic environments. The key IOPs

143   involved in the creation of our IOP dataset include the spectral absorption and backscattering coefficients

144   associated with the main categories of seawater constituents representing suspended particulate matter and CDOM.

145   Specifically, the constituent-specific absorption IOPs are the spectral absorption coefficients of phytoplankton,

146    $a_{ph}(\lambda)$, non-algal particles, $a_d(\lambda)$, and CDOM, $a_g(\lambda)$. Note that the sum $a_{ph}(\lambda) + a_d(\lambda) = a_p(\lambda)$ represents the

147   particulate absorption coefficient with combined contributions of phytoplankton and non-algal particles, and the

148   sum $a_d(\lambda) + a_g(\lambda) = a_{dg}(\lambda)$ represents the non-phytoplankton absorption coefficient with combined contributions

149   of non-algal particles and CDOM. The constituent-specific backscattering IOPs are the spectral backscattering

150   coefficients of phytoplankton, $b_{b-ph}(\lambda)$, and non-algal particles, $b_{b-d}(\lambda)$, such that the sum $b_{b-ph}(\lambda) + b_{b-d}(\lambda) = b_{bp}(\lambda)$

151   is the particulate backscattering coefficient.

152   Among these constituent-specific IOPs, the phytoplankton absorption coefficient, $a_{ph}(\lambda)$, plays the most

153   fundamental role in the creation of the synthetic dataset of IOPs in this study. The $a_{ph}(\lambda)$ spectra in this dataset

154   were derived from actual measurements of phytoplankton absorption made on near-surface samples collected

155   across diverse oceanic environments. Thus, the $a_{ph}(\lambda)$ data are not "synthetic" in a sense that these data were not

156   obtained from a modeling approach although some spectral interpolation or extrapolation was applied to measured

157   data as described in more detail below. In contrast, the remaining four constituent-specific IOPs in the IOP dataset,

158   i.e., $a_d(\lambda)$, $a_g(\lambda)$, $b_{b-ph}(\lambda)$, and $b_{b-d}(\lambda)$, are "synthetic" in a sense that they are entirely based on calculations using a

159   modeling approach with some assumptions about the magnitude and spectral behavior of the modeled IOPs.

160   Importantly, the measured values of $a_{ph}(\lambda)$ were used in the calculations of these IOPs. These calculations are also

161   described in detail below. Thus, each combination of the five constituent-specific IOPs in the synthetic IOP dataset

162   consists of the measured $a_{ph}(\lambda)$ and the calculated $a_d(\lambda)$, $a_g(\lambda)$, $b_{b\text{-}ph}(\lambda)$, and $b_{b\text{-}d}(\lambda)$ where the results of these

163   calculations depend on the measured $a_{ph}(\lambda)$. As a result of this approach, it would seem justifiable to refer to the

164   created IOP dataset as a quasi-synthetic dataset. For simplicity, however, we refer to it as the synthetic IOP dataset

165   while bearing in mind that $a_{ph}(\lambda)$ spectra were derived from measurements.
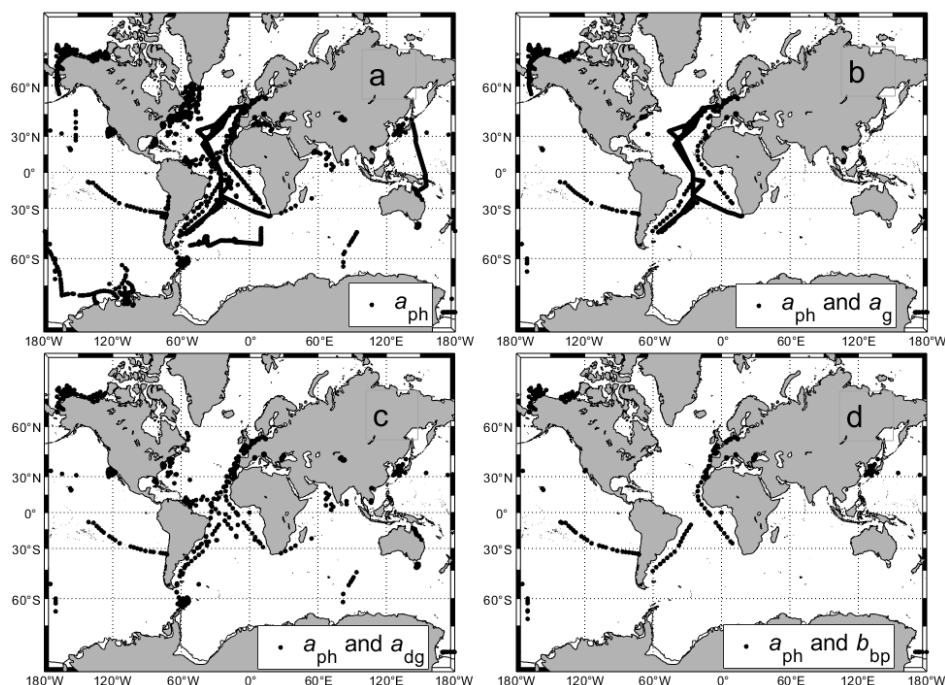
Figure 1. Location of oceanographic stations where in situ measurements were collected for (a) $a_{ph}(\lambda)$, the number of measurements $N = 4382$; (b) $a_{ph}(\lambda)$ and $a_g(\lambda)$, the number of matchup measurements $N = 2206$; (c) $a_{ph}(\lambda)$ and $a_{dg}(\lambda)$, the number of matchup measurements $N = 813$; and (d) $a_{ph}(\lambda)$ and $b_{bp}(\lambda)$, the number of matchup measurements $N = 775$.

199   Figure 1a depicts the location of oceanographic stations where the near-surface measurements of $a_{ph}(\lambda)$ were

200   made. As shown, these measurements were collected across diverse open ocean and coastal environments and their

201   total number is 4382 that constitutes the initial field dataset of $a_{ph}(\lambda)$ considered in this study. Figure 1 also shows

202   the location of stations where coincident measurements are available for the pairs of IOP coefficients, namely

203     $a_{ph}(\lambda)$ and $a_g(\lambda)$ (Fig. 1b), $a_{ph}(\lambda)$ and $a_{dg}(\lambda)$ (Fig. 1c), and $a_{ph}(\lambda)$ and $b_{bp}(\lambda)$ (Fig. 1d). We recall that while the in

204   situ data of $a_g(\lambda)$, $a_{dg}(\lambda)$, and $b_{bp}(\lambda)$ were not used in the development of synthetic IOP dataset, they were assembled

205   for the purpose of comparison with corresponding coefficients that were calculated and included in the synthetic

206   IOP dataset. Many in situ data of IOP coefficients used in the present study were utilized in previous studies (e.g.,

207   Babin et al., 2003; Huot et al., 2008; Bricaud et al., 2010; Antoine et al., 2011; Loisel et al., 2018; Aurin et al.,

208   2018; Reynolds and Stramski, 2019; Stramski et al., 2019; Jorge et al., 2021; Bonelli et al., 2021). Some data are

209   described in publications devoted to compilation of various datasets (Valente et al., 2019; Casey et al., 2020) and

210   are included in several databases (e.g., SeaBASS, CoastlOOC, and GOCAD). As the IOP coefficients in the in situ

211   dataset were measured over a broad range of trophic and environmental conditions, their spectral values span more

212   than 3 or 4 orders of magnitude. This large dynamic range is illustrated in terms of probability distributions at

213   selected light wavelengths, i.e., at 440 nm for the constituent absorption coefficients and 550 nm for the particulate

214   backscattering coefficient (Fig. 2).
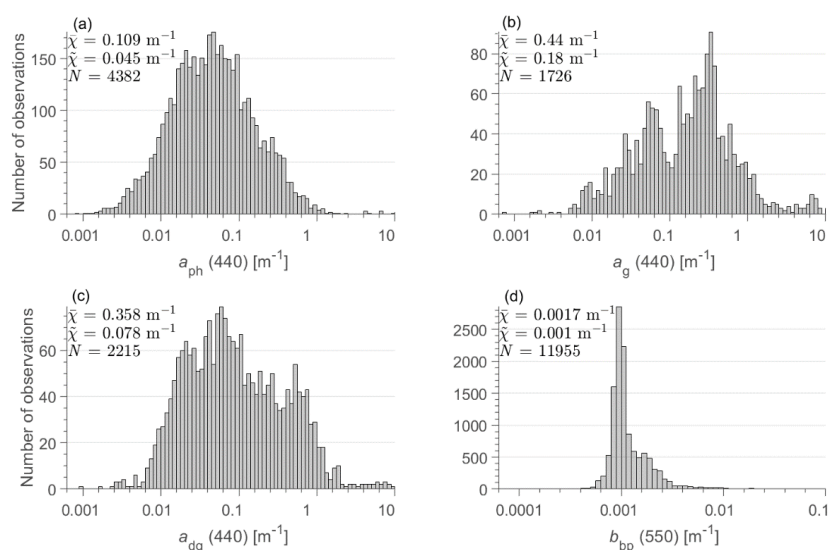
215

216   Figure 2. Histograms and relevant statistical parameters of field measurements of (a) $a_{ph}(440)$, (b) $a_g(440)$, (c) $a_{dg}(440)$, and
217   (d) $b_{bp}(550)$. N is the number of measurements, and $\overline{x}$ and $\tilde{x}$ are the mean and median values of each IOP, respectively.

218

219   The first task necessary for development of the synthetic IOP dataset was to assemble data of hyperspectral

220   absorption coefficient of phytoplankton, $a_{ph}(\lambda)$, from field measurements collected across diverse open ocean and

221   coastal environments (Fig. 1a, Fig. 2a). These $a_{ph}(\lambda)$ data were obtained with the filter-pad spectrophotometric

222   method as a difference between the measurements of $a_p(\lambda)$ and $a_d(\lambda)$ (Kishino et al., 1985; IOCCG Protocol Series,

223   2018). Historically, most of these measurements were acquired with the transmittance configuration of the filter-

224   pad method and such measurements are included in our dataset. However, some data in our dataset were obtained

225   with the inside integrating-sphere configuration of the filter-pad method, which is superior to the transmittance

226   configuration of measurement (Stramski et al., 2015; IOCCG Protocol Series, 2018).

227   A significant portion (23.7%) of the initial dataset of $a_{ph}(\lambda)$ consisting of 4382 measurements covers a spectral

228   range from 400 to 750 nm with high spectral resolution of data reported at 1 nm interval. In some cases, the original

229   measurements extended to near-UV spectral region and/or longer wavelengths in the near-infrared spectral region

230   (800 or 850 nm). The data beyond 750 nm are not used in this study because our RT simulations target the spectral

231   range from 350 to 750 nm. It is notable that the absorption measurements of marine particles and phytoplankton

232   are generally unavailable or are not reported in the UV because of increased methodological challenges and

233   uncertainties in this spectral region (Stramski et al., 2015; IOCCG Protocol Series, 2018; Kostakis et al., 2021).

234   As a result, only a relatively small fraction of $a_{ph}(\lambda)$ measurements in the initial dataset were reported in the near-

235 UV region. In addition, the initial dataset included a relatively large fraction of $a_{ph}(\lambda)$ measurements that were
236 reported at wavelength intervals larger than 1 nm. These lower resolution data (hereafter referred to as
237 multispectral) ranged from a small wavelength interval of 2 nm to data reported at more limited number of
238 wavelengths (as small as <10) within the visible spectral range. It is likely that the multispectral data available
239 from some data sources that we used in this study were originally measured at higher spectral resolution but
240 eventually were reported only for some selected wavelengths, such as those corresponding to spectral bands
241 available on satellite ocean color sensors.

242 The first objective of the analysis of $a_{ph}(\lambda)$ was to consider the initial $a_{ph}(\lambda)$ dataset within the 400–750 nm
243 range and convert the measurements that were reported at lower spectral resolution to uniformly hyperspectral
244 data at 1 nm interval. In this analysis, all measurements originally available at 1 nm interval were considered to
245 provide reference spectral shape functions of $a_{ph}(\lambda)$. The originally multispectral data of $a_{ph}(\lambda)$ were converted to
246 hyperspectral data using a couple of different approaches depending on the spectral features of lower resolution
247 data. One approach utilized the reference spectral shape functions of $a_{ph}(\lambda)$ and was applied to multispectral $a_{ph}(\lambda)$
248 data if they were reported at fewer wavelengths than 100. In this case, a given multispectral spectrum of $a_{ph}(\lambda)$
249 was converted to hyperspectral spectrum using a specific hyperspectral measurement that exhibited the highest
250 correlation with the multispectral measurement under consideration. The correlation coefficient was calculated
251 using the spectral data available at common wavelengths of considered pair of spectra. A necessary condition to
252 proceed with a conversion of a given multispectral spectrum to hyperspectral spectrum was a correlation
253 coefficient of 0.95 or higher. If this condition was satisfied, the multispectral data were converted to hyperspectral
254 data so that the created hyperspectral spectrum maintained the magnitude of multispectral measurement in the
255 range of 440–450 nm and had the spectral shape of the reference hyperspectral measurement. An alternative
256 approach to convert multispectral data to hyperspectral data involved a linear interpolation of multispectral data.
257 This approach was used when the multispectral data were reported at relatively small wavelength intervals (at least
258 100 spectral data available between 400 and 750 nm) or when the correlational analysis described above did not
259 yield the correlation coefficient of 0.95 or higher. The originally multispectral spectra that did not include data
260 below 450 nm or fell into the category of data subject to linear interpolation but had no data above 700 nm were
261 rejected from further analysis. For all hyperspectral spectra that passed the above-described analysis and criteria
262 (i.e., 2204 spectra that included both the original hyperspectral measurements and hyperspectral spectra created
263 from multispectral data), the null-point correction was applied by subtracting the average value of $a_{ph}(\lambda)$ in the
264 745–750 nm range from all spectral values in the 400–750 nm range.

265 The next step of analysis was to extend all null-point corrected spectra of $a_{ph}(\lambda)$ that cover the 400–750 nm
266 range into the UV spectral region. The primary focus was on the 350–400 nm range because our RT simulations
267 were designed to provide output results in the 350–750 nm range. For this purpose we used a separate subset of
268 reference hyperspectral measurements of $a_{ph}(\lambda)$ that includes the near-UV spectral region. This reference subset
269 of data consisted of 233 measurements collected across bio-optically diverse marine environments in the Pacific
270 and Atlantic Oceans and western Arctic seas. The majority of these 233 spectra (170) were collected with the
271 inside integrating-sphere configuration of filter-pad method, while the remaining 63 measurements were done
272 using either the transmittance or transmittance-reflectance filter-pad configuration (Zheng et al., 2014). A
273 correlational analysis was applied to pairs of spectra, each consisting of a spectrum covering the 400–750 nm range
274 and a reference spectrum covering the 350–750 nm range. The correlation coefficient was calculated using data at

275 common wavelengths from the 400–750 nm range. The reference spectrum that yielded the highest correlation

276 with the investigated 400–750 nm spectrum was selected as a basis for extrapolation of the investigated spectrum

277 into the 350–400 nm range. This extrapolation ensured that a given investigated spectrum maintained its magnitude

278 at 400 nm and its extrapolated near-UV portion had the spectral shape of the selected reference spectrum. The final

279 aspect of extrapolation in the UV is related to the spectral range 300–350 nm. The rationale for IOP data extending

280 to 300 nm is to ensure that the results of RT simulations that start at 350 nm account for possible effects of Raman

281 scattering by water molecules in the UV spectral region. Therefore, for the 300–350 nm range we simply assumed

282 that $a_{ph}(\lambda)$ in this range is equal to $a_{ph}(350)$. The limitation associated with this assumption is not considered to be

283 serious given the limited role of the 300–350 nm range in the RT simulations and weak Raman scattering effects

284 in UV spectral region. Example spectra of $a_{ph}(\lambda)$ in the 350–750 nm range from contrasting marine environments

285 are presented in Fig. 3. These examples show significant variation in both the magnitude and spectral shape of

286 $a_{ph}(\lambda)$.

287

288
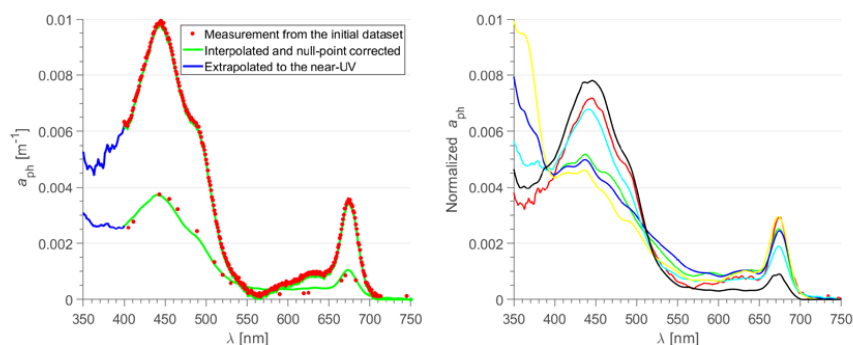
289
290
291 Figure 3. (a) Two example spectra of $a_{ph}(\lambda)$ from contrasting oceanic environments. For each example $a_{ph}(\lambda)$, two spectra are
292 displayed, namely the measurement from the initial $a_{ph}(\lambda)$ dataset shown at the original wavelength intervals (red points) and
293 the spectrum after interpolation to 1 nm intervals (if required) and null-point correction (continuous lines). The UV portion of
294 the latter was obtained by extrapolation based on reference data in the UV (see text for details). (b) Example of normalized
295 $a_{ph}(\lambda)$ spectra illustrating the variability of the spectral shape of the $a_{ph}(\lambda)$ database. These spectra have been normalized to
296 their integral.
297

298   In the next step of analysis, the subset of 2204 $a_{ph}(\lambda)$ spectra that was created from the initial $a_{ph}(\lambda)$ dataset as

299 described above was subject to additional modifications to ensure that the final $a_{ph}(\lambda)$ dataset is characterized by

300 the probability distribution that resembles the distribution representative of the global ocean. This process and

301 background information on the motivation for such adjustments in the probability distribution are described below.

302   When the end goal is to achieve a high degree of representativeness of global ocean like in this study, the

303 process of assembling in situ datasets of IOPs is unavoidably subject to limitations, even if relatively large amount

304 of data from many field experiments and cruises are considered. This is mainly because the global ocean is

305 dominated by vast areas of open-ocean pelagic environments and the amount of IOP data collected in these

306 environments is disproportionally limited compared to amount of data collected in coastal regions that represent a

307 relatively small portion of the global ocean. Thus, the probability distributions based on in situ datasets, such as

308 those presented in Fig. 2, are expected to deviate from the probability distributions representative of the global

309   ocean. In particular, the maxima of probability distributions and the measures of central tendency, such as the

310   median and mean values, obtained from compilations of relatively large amount of in situ IOP data (such as in Fig.

311   2) are expected to be shifted to larger values compared to actual global distributions because the IOPs exhibit a

312   general tendency of higher values in coastal regions compared to open ocean environments. While this issue has

313   been recognized, it has not been addressed or resolved in various studies that focus on global ocean color

314   applications. For example, the current global ocean color algorithms for estimating chlorophyll-a concentration

315   (Chla) are based on relatively large amount of in situ data whose probability distribution is shifted significantly to

316   higher Chla compared with the global Chla distribution (O'Reilly and Werdell, 2019). Similarly, in the

317   development of previous synthetic optical databases with RT simulations (e.g., IOCCG Report, 2006), no special

318   attempt was made to ensure consistency between the probability distributions of input IOP data and the

319   distributions expected for global ocean. In the recent development of refined global ocean color algorithms for

320   estimating the concentration of particulate organic carbon (POC), the in situ dataset was assembled with a goal to

321   achieve reasonable consistency with a global POC distribution (Stramski et al., 2022). This goal was, however,

322   achieved at the expense of significant reduction in the amount of accepted in situ data compared to the size of

323   overall pool of available in situ data.

324   In this study our goal was to create a relatively large synthetic IOP dataset based on the initial dataset of

325   several thousand measurements of spectral $a_{ph}(\lambda)$, so that the probability distributions of IOPs in the final synthetic

326   dataset are reasonably consistent with the expected distributions representative of the global ocean. As described

327   above, the initial field dataset in support of this process consisted of 4382 spectra of $a_{ph}(\lambda)$ and this number was

328   further reduced to 2204 spectra that were accepted as a result of analysis and some criteria applied to the initial

329   dataset. This reduced dataset of accepted $a_{ph}(\lambda)$ spectra was then further modified to ensure that the final

330   probability distribution of $a_{ph}(440)$ resembles the global distribution of $a_{ph}(440)$. The global probability

331   distribution of $a_{ph}(440)$ was estimated using retrievals of $a_{ph}(440)$ from satellite ocean color data. Specifically, we

332   used global satellite observations made with the ocean color sensor OLCI (Ocean and Land Colour Instrument)

333   deployed on the Sentinel-3 mission (Donlon et al., 2012) from the period December 1, 2020 through November

334   30, 2021. The weekly data product of remote-sensing reflectance $R_{rs}(\lambda)$ at 4 km^2 spatial resolution was used as

335   input to the 3-step semi-analytical algorithm (3SAA) to derive $a_{ph}(443)$ as described in Jorge et al (2021). The

336   $a_{dg}(443)$ and $b_{bp}(\lambda)$ coefficients were also derived from this algorithm. In general, the 3SAA first derives the diffuse

337   attenuation coefficient for downwelling plane irradiance averaged within the surface layer down to the first

338   attenuation depth, $<K_d(\lambda)>_1$, from $R_{rs}(\lambda)$, and then utilizes the inverse model LS2 (Loisel et al., 2018) to derive the

339   total absorption, $a(\lambda)$, and backscattering, $b_b(\lambda)$, coefficients from $R_{rs}(\lambda)$ and $<K_d(\lambda)>_1$. After subtracting the pure

340   seawater contributions, the non-water absorption, $a_{nw}(\lambda)$, and the particulate backscattering, $b_{bp}(\lambda)$, coefficients

341   are obtained. Finally, $a_{ph}(\lambda)$ and $a_{dg}(\lambda)$ are derived from $a_{nw}(\lambda)$ using an optimization algorithm of Zhang et al.

342   (2015) with modifications that account for differences in optical water types defined in terms of different spectral

343   shapes of $R_{rs}(\lambda)$ (Mélin and Vantrepotte, 2015). While the original classification of Mélin and Vantrepotte (2015)

344   includes 16 optical water classes (OWC), the derivation of $a_{ph}(\lambda)$ and $a_{dg}(\lambda)$ from the 3SAA additionally included

345   a 17th OWC to improve the representation of ultraoligotrophic waters such as those found in the South Pacific

346   Gyre (Morel et al., 2007; Claustre et al., 2008; Stramski et al., 2008) and the Mediterranean Sea (Loisel et al.,

347   2011). This 17th OWC is described in Jorge et al. (2021).

348   The 3SAA does not yield the separate contributions of CDOM, $a_g(\lambda)$, and non-algal particles, $a_d(\lambda)$, to the

349   overall non-phytoplankton absorption coefficient, $a_{dg}(\lambda)$. Therefore, we also used another semi-analytical model

350   (CDOM-KD2) described in Bonelli et al. (2021) to estimate $a_g(443)$ from OLCI-derived $R_{rs}(\lambda)$. Having $a_{dg}(\lambda)$ from

351   the 3SAA and $a_g(\lambda)$ from the CDOM-KD2, the non-algal particulate absorption, $a_d(\lambda)$, was obtained as a difference

352     $a_{dg}(\lambda) - a_g(\lambda)$. As a result of this analysis, we obtained a dataset of satellite-derived constituent absorption

353   coefficients, $a_{ph}(443)$, $a_g(443)$, $a_d(443)$, and $a_{dg}(443)$, as well as the particulate backscattering coefficient, $b_{bp}(550)$,

354   where we focused on the spectral band near 440 nm for absorption and 550 nm for backscattering.
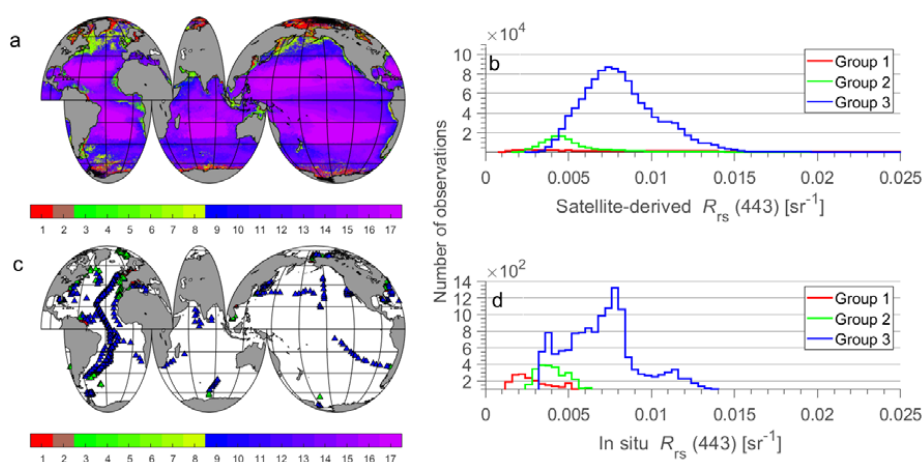
355

369 Figure 4. (a) Global map illustrating the distribution of seventeen optical water classes estimated from monthly $R_{rs}(\lambda)$ values
370 derived from satellite observations with ocean color sensor OLCI from December 2020 through November 2021 (weekly
371 products at 4km^2). The color bar scale refers to optical water classes. (b) Histogram of OLCI-derived $R_{rs}(443)$ for the three
372 optical water groups (see text for details). (c) Location of oceanographic stations where in situ measurements of $R_{rs}(\lambda)$ were
373 collected and used to analyze the consistency of the synthetic dataset with field measurements. (d) Histograms of in situ
374 measurements of $R_{rs}(443)$ for the three optical water groups.

375

376   For illustrative purposes Fig. 4a depicts the spatial distribution of 17 optical water classes (OWCs) over the

377   global ocean obtained from satellite OLCI data following the methodology of Mélin and Vantrepotte (2015). For

378   further illustrative purposes, these 17 OWCs were grouped into 3 optical water groups (OWGs). Group 1 consists

379   of OWC1 and OWC2 which are characterized by high water turbidity such as in coastal areas affected by discharge

380   from large rivers. Although the focus of this study is to create the synthetic datasets representative primarily of

381   open ocean and moderately turbid coastal waters, an explicit identification of Group 1 data that represent very

382   turbid waters is of interest for comparisons with the database developed specifically for coastal waters by Nechad

383   et al. (2015). The second OWG, Group 2, includes 6 OWCs from OWC3 through OWC8. This group represents

384   mainly productive waters in both coastal and open ocean environments, such as those encountered in the North

385   Atlantic during a period of phytoplankton bloom (Levy et al., 2005). Finally, Group 3 included the remaining 9

386   OWCs from OWC9 through OWC17. These water types are observed mainly in mesotrophic and oligotrophic

387   regions of the global ocean. Based on this classification, 79.6% of OLCI water pixels in Fig. 4a belong to Group

388   3, 10.8% to Group 2, and 9.6% to Group 1. The histograms of OLCI-derived $R_{rs}(443)$ associated with these three

389   groups of data are shown in Fig. 4b. For comparative purposes we also assembled a dataset of in situ measurements

390   of $R_{rs}(\lambda)$, which were collected at various locations within the global ocean (Fig. 4c). The histograms of in situ

391   $R_{rs}(443)$ associated with Groups 1, 2, and 3 are depicted in Fig. 4d, which show a similar pattern to that in Fig. 4b.

392   For in situ dataset of $R_{rs}(\lambda)$, 69.2% of data belong to Group 3, 15.7% to Group 2, and 15.1% to Group 1.

393
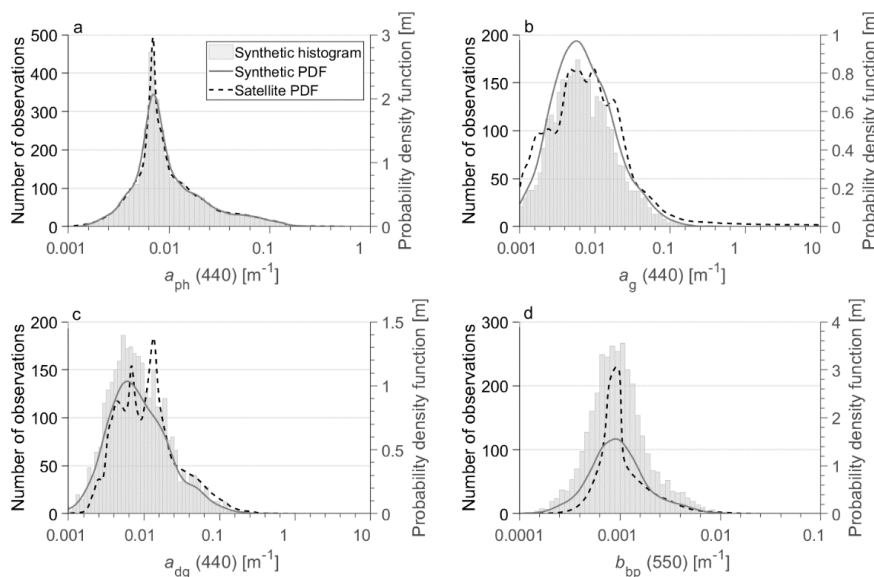
Figure 5. Histograms showing the distribution of the synthetic IOP data used in the present study. The synthetic and satellite-derived probability density functions (PDFs) for each IOP are represented by the solid and dashed curves, respectively.

412

413   The probability density function (PDF) of global satellite-derived $a_{ph}(440)$, $a_g(440)$, $a_{dg}(440)$, and $b_{bp}(550)$ are

414   depicted in Fig. 5. We note that we refer here to satellite-derived absorption coefficients at 440 nm although they

415   were derived from OLCI reflectances at 443 nm, which is a minor difference that is inconsequential for the purpose

416   of this study. The comparison of Fig. 2a and Fig. 5a indicates that the distribution of measured $a_{ph}(440)$ from our

417   initial field dataset (Fig. 2a) is shifted towards higher values compared to the global distribution of satellite-derived

418   $a_{ph}(440)$ (Fig. 5a). The probability distribution of reduced dataset of measured $a_{ph}(440)$ ($N = 2204$) that was created

419   from the initial field dataset of $a_{ph}(\lambda)$ show similar deviations from the global distribution (not shown). Thus, to

420   create the final dataset of $a_{ph}(\lambda)$ that has the probability distribution of $a_{ph}(440)$ consistent with the global satellite-

421   derived distribution, we adjusted the number of $a_{ph}(440)$ measurements in each bin of the histogram of the reduced

422   dataset either by removing the measurements from any given bin or adding the measurements to this bin. The

423   removal or addition of $a_{ph}(440)$ measurements associated with any given bin was done by subjecting all $a_{ph}(440)$

424   measurements originally contained within a given bin to random selection. Specifically, in the case of addition the

425   randomly selected $a_{ph}(440)$ was added as a replicate of $a_{ph}(440)$ to a given bin. In the case of removal, the randomly

426   selected $a_{ph}(440)$ was simply removed from a given bin. As a result of this process we obtained a modified

427   distribution of measured $a_{ph}(440)$ which is fairly consistent with the satellite-derived distribution of $a_{ph}(440)$. Both

428   the modified histogram and the corresponding modified PDF of measured $a_{ph}(440)$ are depicted in Fig. 5a for

429   comparison with the global satellite-derived distribution. In total, this modified distribution consists of 3320

430   measurements of $a_{ph}(440)$ and, obviously, each of these measurements at 440 nm has an associated full spectrum

431   of $a_{ph}(\lambda)$ values between 300 and 750 nm. These 3320 spectra of $a_{ph}(\lambda)$ represent one IOP component of the final

432   synthetic IOP dataset.

433   The full synthetic IOP dataset created in this study consists of 3320 combinations of measured $a_{ph}(\lambda)$ and

434   synthetically-generated $a_d(\lambda)$, $a_g(\lambda)$, $b_{b-ph}(\lambda)$, and $b_{b-d}(\lambda)$. Below is a description of calculations of $a_g(\lambda)$, $a_d(\lambda)$, b_{b-}

435    $_{ph}(\lambda)$, and $b_{b-d}(\lambda)$. We note that all IOP coefficients are expressed in units of [m^{-1}] and the light wavelength is in

436   units of [nm].

437   The four IOP coefficients were calculated using a similar methodology to that applied in previous studies

438   aiming at generation of synthetic ocean optical databases (IOCCG Report, 2006; Craig et al., 2020). Specifically,

439   we used the measured values of $a_{ph}(440)$ as the main driver of calculations of $a_g(\lambda)$, $a_d(\lambda)$, $b_{b-ph}(\lambda)$, and $b_{b-d}(\lambda)$.

440   Thus, the variability in the measured $a_{ph}(440)$, as depicted by the probability distribution of measured $a_{ph}(440)$ in

441   Fig. 5a, is the main source of variability in these four co-existing IOP coefficients. It is notable that the replicate

442   values of $a_{ph}(440)$ present within any given bin of the $a_{ph}(440)$ distribution result in the generation of different

443   values of the four IOP coefficients because the formulas involved in these calculations contain random parameters.

444   The coupling between $a_{ph}(440)$ and CDOM absorption coefficient was defined as:

445    $$a_g(440) = 10^{(P_1 + \gamma)} \tag{1}$$

446   where P_1 is a parameter related to $a_{ph}(440)$ and γ is randomly selected from a predetermined range of values (Table

447   1). The spectral values of $a_g(\lambda)$ are subsequently determined from:

448    $$a_g(\lambda) = a_g(440)\, e^{-S_g(\lambda - 440)} \tag{2}$$

449   where the spectral slope parameter, S_g in units of [nm^{-1}], is randomly selected from a predetermined range of values

450   (Table 1). The absorption coefficient of non-algal particles was modeled in a similar fashion:

451    $$a_d(440) = P_2\, a_{ph}(440) \tag{3}$$

452    $$a_d(\lambda) = a_d(440)\, e^{-S_d(\lambda - 440)} \tag{4}$$

453   where P_2 is a parameter related to $a_{ph}(440)$ and the spectral slope parameter S_d [nm^{-1}] is randomly selected from a

454   predetermined range of values (Table 1).

455   The particulate backscattering is not modeled in terms of the single coefficient, $b_{bp}(\lambda)$, but instead as separate

456   contributions by phytoplankton, $b_{b-ph}(\lambda)$, and non-algal particles, $b_{b-d}(\lambda)$, so that their sum yields $b_{bp}(\lambda)$. In order

457   to calculate $b_{b-ph}(\lambda)$, first the formula that couples $a_{ph}(440)$ with the beam attenuation coefficient of phytoplankton

458   at 550 nm, $c_{ph}(550)$, is used:

459    $$c_{ph}(550) = P_3\, \text{Chla}^{0.57} = P_3 \left[\frac{a_{ph}(440)}{0.05582} \right]^{0.57} \tag{5}$$

460   where Chla is the concentration of chlorophyll-a in units of [mg m^{-3}], 0.05582 [m^2/(mg Chla)] is the value of

461   chlorophyll-specific absorption coefficient of phytoplankton at 440 nm, $a_{ph}^*(440)$, (Maritorena et al., 2002), and

462   P_3 is a parameter with a randomly selected value from a predetermined range (Table 1). The exponent value of

463   0.57 is based on the study of Voss (1992). Subsequently, the spectral values of phytoplankton beam attenuation

464   coefficient are calculated from:

465    $$c_{ph}(\lambda) = c_{ph}(550) \left(\frac{550}{\lambda} \right)^{S_{c-ph}} \tag{6}$$

466   where the spectral slope parameter, S_{c-ph} [dimensionless], is calculated using both $a_{ph}(440)$ and a random number

467   generator (Table 1). Next, the spectral scattering coefficient of phytoplankton is determined:

468 $$b_{ph}(\lambda) = c_{ph}(\lambda) - a_{ph}(\lambda) \tag{7}$$

469 where the spectral values of $a_{ph}(\lambda)$ are from the same measured spectrum as the value of $a_{ph}(440)$ in Eq. (5). Finally,

470 the spectral backscattering coefficient of phytoplankton is calculated from:

471 $$b_{b-ph}(\lambda) = 0.01 \, b_{ph}(\lambda) \tag{8}$$

472 where 0.01 is the value of backscattering ratio of phytoplankton, $\widetilde{b_{ph}}$, assumed to be constant and independent of

473 light wavelength. We note that $b_{b-ph}(\lambda)$ is not required as input to our radiative transfer simulations but $b_{ph}(\lambda)$ is

474 needed.

475   To calculate the backscattering coefficient of non-algal particles, $b_{b-d}(\lambda)$, the phytoplankton absorption at 440

476 nm is first coupled with the scattering coefficient of non-algal particles at 550 nm, $b_d(550)$, using the following

477 relationship:

478 $$b_d(550) = P_4 \, \text{Chla}^{0.766} = P_4 \left[\frac{a_{ph}(440)}{0.05582} \right]^{0.766} \tag{9}$$

479 where the parameter P_4 is randomly selected from a predetermined range (Table 1) and the value of 0.05582 is

480 $a_{ph}^*(440)$ as explained in relation to Eq. (5). The exponent value of 0.766 is based on the study of Loisel et al.

481 (1998). Then, the spectral values of non-algal scattering coefficient are calculated from:

482 $$b_d(\lambda) = b_d(550) \left(\frac{550}{\lambda} \right)^{S_{b-d}} \tag{10}$$

483 where the spectral slope parameter, S_{b-d} [dimensionless], is calculated using both $a_{ph}(440)$ and a random number

484 generator (Table 1). In the final step, the spectral backscattering coefficient of non-algal particles is calculated as:

485 $$b_{b-d}(\lambda) = 0.018 \, b_d(\lambda) \tag{11}$$

486 where the constant 0.018 is the backscattering ratio of non-algal particles, $\widetilde{b_d}$. Again, we note that $b_{b-d}(\lambda)$ is not

487 required as input to radiative transfer simulations but $b_d(\lambda)$ is needed.

488

489 Table 1: Symbols of variables, mathematical expressions, and corresponding equations in the text of the paper.

490 $rng(0,1)$ is a random number between 0 and 1.

Symbols	Mathematical expression	Equation
P_1	$0.79 \log_{10}\left[a_{ph}(440)\right] - 0.37$	(1)
γ	$-0.2 + 0.3 \, rng(0,1)$	(1)
S_g	$(0.02 - 0.01) \, rng(0,1) + 0.01$	(2)
P_2	$0.1 + \dfrac{0.4 \, rng(0,1) \, a_{ph}(440)}{0.05 + a_{ph}(440)}$	(3)
S_d	$(0.015 - 0.007) \, rng(0,1) + 0.007$	(4)
P_3	$(0.3 - 0.03) \, rng(0,1) + 0.03$	(5)
S_{c-ph}	$-0.4 + \dfrac{1.6 + 1.2 \, rng(0,1)}{1 + \left[\frac{a_{ph}(440)}{0.05582} \right]^{0.5}}$	(6)
P_4	$(0.16668 - 0.016668) \, rng(0,1) + 0.016668$	(9)
S_{b-d}	$-0.5 + \dfrac{2 + 1.2 \, rng(0,1)}{1 + \left[\frac{a_{ph}(440)}{0.05582} \right]^{0.5}}$	(10)

491

492

493   The variability of measured $a_{ph}(440)$ illustrated in Fig. 5a along with the dynamic range of parameters P_1, P_2,

494 P_3, P_4, the spectral slopes S_g, S_d, S_{c-ph}, and S_{b-d}, and the degree of randomness in the selection of these parameters

495   for any given value of $a_{\mathrm{ph}}(440)$ that initiates the process of calculating $a_{\mathrm{g}}(\lambda)$, $a_{\mathrm{d}}(\lambda)$, $b_{\mathrm{b\text{-}ph}}(\lambda)$, and $b_{\mathrm{b\text{-}d}}(\lambda)$, resulted in

496   the generation of synthetic dataset of these IOP coefficients that cover a wide dynamic range consistent with in

497   situ and satellite observations over the global ocean. Figure 5b,c,d compares the probability distributions of

498   satellite-derived $a_{\mathrm{g}}(440)$, $a_{\mathrm{dg}}(440)$, and $b_{\mathrm{bp}}(550)$ with the distribution of these coefficients from the final synthetic

499   IOP dataset. This comparison supports the general consistency of the distributions of these IOP coefficients, which

500   is in line with the desired consistency achieved for $a_{\mathrm{ph}}(440)$ (Fig. 5a) as discussed earlier in this section. It is also

501   noteworthy that in contrast to this newly created synthetic IOP dataset, the previous synthetic datasets exhibit

502   significant differences between the probability distributions of synthetic IOPs and global distributions based on

503   satellite observations (Fig. 6).

504
505
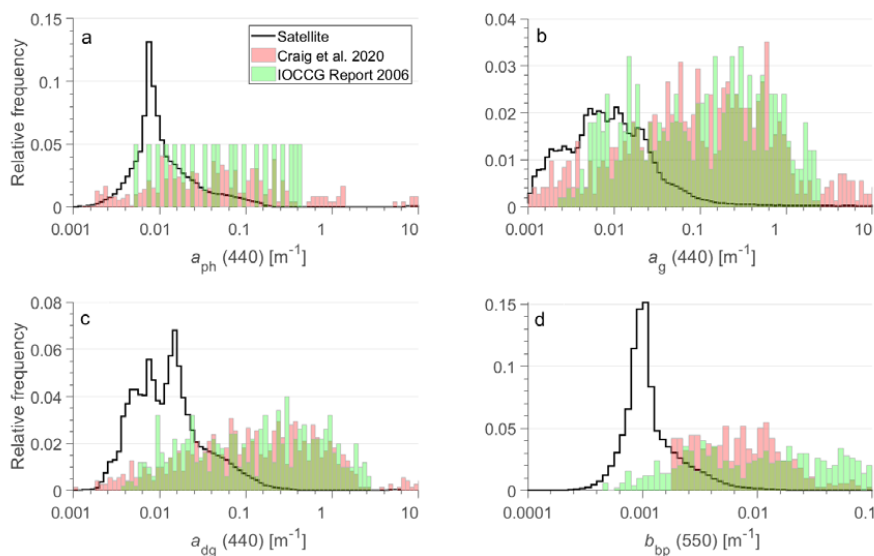
521   Figure 6. Histograms showing the distribution of IOPs from the synthetic datasets of the IOCCG Report (2006) and Craig et
522   al. (2020) in the green and pink, respectively. The IOP distributions estimated from satellite ocean color observations with
523   OLCI sensor over the global ocean are represented by the black line.

524

525   Overall, the above-described synthetic IOP dataset includes 3320 scenarios of non-water IOPs, i.e., IOPs

526   associated with variable contributions of phytoplankton, non-algal particles, and CDOM to optical properties of

527   seawater. In addition to the non-water absorption coefficients, $a_{\mathrm{ph}}(\lambda)$, $a_{\mathrm{d}}(\lambda)$, and $a_{\mathrm{g}}(\lambda)$, as well as the non-water

528   scattering coefficients, $b_{\mathrm{ph}}(\lambda)$ and $b_{\mathrm{d}}(\lambda)$, the radiative transfer simulations required input of scattering phase

529   functions of particles, specifically for phytoplankton and non-algal particles. We assumed the particulate phase

530   functions proposed by Fournier-Forand (1994) with the backscattering ratio $\widetilde{b_{ph}} = 0.01$ for phytoplankton and $\widetilde{b_d}$

531   = 0.018 for non-algal particles. Note that while the backscattering ratios are assumed to be spectrally constant, the

532   phase functions vary with light wavelength because of spectral variations of $b_{\mathrm{ph}}(\lambda)$ and $b_{\mathrm{d}}(\lambda)$. All IOP data in the

533   final synthetic IOP dataset cover the spectral range from 300 to 750 nm with a 5 nm interval. This wavelength

534   interval is consistent with the intended output of our radiative transfer simulations.

535 The radiative transfer simulations also required input of the absorption and scattering properties of pure

536 seawater. For the spectral absorption coefficient of pure seawater, $a_w(\lambda)$, we used the values recommended in

537 IOCCG Protocol Series (2018). This recommendation includes the values from Jonasz and Fournier (2007) in the

538 spectral range 300-330 nm, Morel et al. (2007) in the 340-415 nm range, Pope and Fry (1997) in the 420-725 nm

539 range, and Kou et al. (1993) in the 730-750 nm range. The spectral volume scattering function of pure seawater

540 (from which the spectral scattering coefficient and scattering phase function can be obtained) was calculated

541 following Zhang et al. (2009) assuming water temperature of 18°C and salinity of 35‰. The temperature of 18°C

542 is consistent with the mean sea surface temperature (SST) calculated from the monthly global NOAAv2 SST

543 database at 1° spatial resolution from December 1991 through November 2021 (Jérôme Vialard, personal

544 communication, https://www.psl.noaa.gov/data/gridded/data.noaa.oisst.v2.html). The salinity of 35‰ is also

545 consistent with the global surface average (Durack et al., 2013).

546

547 **3 Radiative transfer simulations**

548 The IOP dataset described in section 2, which includes 3320 combinations of non-water IOPs, provided the

549 key input to radiative transfer (RT) simulations which were performed with the HydroLight v5.0 radiative transfer

550 code (Mobley and Sundman, 2008). All RT simulations were run assuming vertically homogeneous IOPs within

551 the water column and infinitely deep ocean, i.e., no effect of seafloor on light field within the water column. For

552 all simulations the computed radiometric and AOP variables were saved into the output data files at 10 cm depth

553 intervals between the ocean surface and the 1 m depth, and at 1 m intervals between the 1 m and 50 m depth. Thus,

554 the primary focus of our RT simulations is on the ocean surface layer that can potentially contribute to light leaving

555 the ocean with significance to remote sensing with spaceborne or airborne optical instruments. All simulations

556 were carried out in the spectral range from 300 to 750 nm using 5-nm spectral bands and the results were produced

557 for the nominal wavelengths of each of the 81 bands, that is at 350, 355, 360, etc…, 745, 750 nm. The results in

558 the 300–350 nm range were not retained in the output files (that include seawater IOPs, radiometric quantities and

559 AOPs) because this spectral region was included primarily to account for potential effects of inelastic processes at

560 wavelengths longer than 350 nm and, additionally, it is known the uncertainties in the characterization of seawater

561 IOPs can increase significantly at wavelengths shorter than 350 nm.

562 For 3320 scenarios of input IOPs we performed several separate sets of RT simulations which differed in

563 terms of assumed sea-surface boundary conditions and the inclusion or exclusion of inelastic radiative processes

564 within the water column. The assumptions regarding the sea-surface boundary conditions were the same as in the

565 previous RT simulations described in Loisel et al. (2018). Specifically, all simulations were made under the same

566 assumption of wind speed of 5 m s^{-1}, which determines the sea-surface roughness involved in the calculations of

567 transmission and reflection of light at the air-water interface. In all simulations the sky conditions were also

568 assumed to be the same, i.e., clear skies and standard atmosphere. However, three distinct sets of simulations were

569 made for the three values of sun zenith angle, 0°, 30°, and 60°. With regard to consideration of inelastic processes,

570 we also performed three distinct sets of simulations. The first of these sets assumed the absence of inelastic

571 processes in water, that is no Raman scattering by water molecules, no fluorescence by chlorophyll-a, and no

572 fluorescence by CDOM. The second set of these simulations included Raman scattering by water molecules.

573 Finally, the third set included both Raman scattering and chlorophyll-a fluorescence, and this scenario of inelastic

574 processes is expected to generally provide the most realistic simulations of radiative transfer in the ocean surface

575   layer. We note, however, that fluorescence by CDOM was not included in any simulations. The Raman scattering

576   coefficient, phase function, and wavelength distribution function were set to their default values described in

577   HydroLight technical documentation (Mobley, 2012). The quantum efficiency of chlorophyll-a fluorescence was

578   also set to its default value of 0.02 in the HydroLight code. For each scenario of sun zenith angle and inelastic

579   processes, we performed 3320 RT simulations, each for a different combination of seawater IOPs. Thus, given the

580   three sun zenith angles, the three scenarios of inelastic processes, and 3320 combinations of IOPs, overall we

581   performed 29880 simulations. The combination of the synthetic IOP dataset used as input to RT simulations

582   (section 2) and the results for the radiance, other radiometric quantities, and AOPs obtained from these 29880

583   simulations (described in this section) constitute the synthetic ocean optical database developed in this study.

584   **4 Comparisons of the synthetic database with in situ data**

585   In this section we compare the selected spectral IOP coefficients from the synthetic IOP dataset with in situ

586   data of IOPs and the selected spectral AOPs from the synthetic database generated with the RT simulations with

587   in situ data of AOPs. In these comparisons, we also include some empirical relationships between the IOPs or

588   AOPs which were established in previous studies based on the analysis of in situ data.
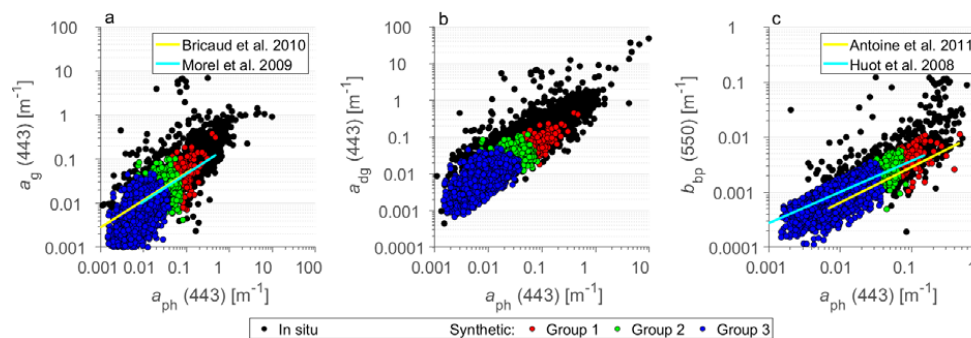
589
590   Figure 7. (a) $a_g(443)$, (b) $a_{dg}(443)$, and (c) $b_{bp}(550)$ as a function of $a_{ph}(443)$ for the in situ dataset (black data points) and the
591   synthetic dataset (colored data points). Each color refers to the optical water group as indicated (139, 262, and 2919 data points
592   for Group 1, 2, and 3, respectively). Empirical relationships previously developed for (a) $a_g(443)$ vs. $a_{ph}(443)$ and (c) $b_{bp}(550)$
593   vs. $a_{ph}(443)$ are also displayed for comparison. The original relationships were formulated as a function of Chla and the
594   presented relationships were obtained by converting Chla to $a_{ph}(443)$ using the chlorophyll-specific phytoplankton absorption
595   at 443 nm from Bricaud et al. (1998).
596
597
598   Figure 7 depicts the scatter plots of IOP coefficients, specifically $a_g(440)$ vs. $a_{ph}(440)$ (Fig. 7a), $a_{dg}(440)$ vs.

599    $a_{ph}(440)$ (Fig. 7b), and $b_{bp}(550)$ vs. $a_{ph}(440)$ (Fig. 7c). The scatter plots include two datasets, the in situ dataset and

600   the synthetic dataset as described in section 2. We recall that in both types of datasets, $a_{ph}(440)$ plotted on the x-

601   axis is the same because the phytoplankton absorption data used in this study were obtained from field

602   measurements with no modeling involved. The scatter plots show a significant degree of overlap which indicates

603   general consistency between the synthetic and in situ datasets. For illustrative purposes, the data from the synthetic

604   IOP dataset are color coded to indicate the partitioning of data into the three OWGs, i.e., Groups 1, 2, and 3 that

605   were defined using the synthetic spectra of $R_{rs}(\lambda)$ generated through RT simulations with input of the synthetic

606   IOP data. As expected, the data with generally lowest values of IOPs belong to Group 3, the data with intermediate

607   values of IOPs to Group 2, and the data with the highest IOPs (most turbid waters) to Group 1. We also note the

608   in situ dataset exhibits somewhat wider dynamic range of variability than the synthetic dataset, especially when

609   the IOP ratios, $a_g(440)/a_{ph}(440)$, $a_{dg}(440)/a_{ph}(440)$, and $b_{bp}(550)/a_{ph}(440)$, are relatively high. While this result can

610   reflect some degree of intrinsic difference in the dynamic range covered by the two datasets, it must also be

611   recognized that some variability in the in situ dataset may be associated with the fact that these data were collected

612   on numerous cruises by different groups of investigators using the methodology (instrumentation, data processing,

613   data quality control, etc.) that unavoidably was not the same across all different data sources.

614   For additional comparative purposes, Fig. 7a,c includes a few empirical relationships between the IOPs in

615   question, which were established in previous studies based on considerable amount of field measurements

616   collected mostly in open ocean environments. As seen, the relationships between $a_g(440)$ and $a_{ph}(440)$ based on

617   the studies of Morel (2009) and Bricaud et al. (2010) agree quite well with the central tendency of variation within

618   our synthetic dataset. We note that Morel (2009) and Bricaud et al. (2010) reported on the relationships between

619    $a_g(440)$ and Chla which were very similar in these studies. For the purpose of illustration in our Fig. 7a, we replaced

620   Chla with $a_{ph}(440)$ using the formula Chla = $a_{ph}(440)/0.05582$. Similarly, the studies of Huot et al. (2008) and

621   Antoine et al. (2011) reported on empirical relationships between $b_{bp}(\lambda)$ and Chla. After converting Chla to

622    $a_{ph}(440)$ as mentioned above, these two relationship are plotted in Fig. 7c. Although these two relationships have

623   different slopes, they are both generally consistent with the average trend of variation in the synthetic dataset.
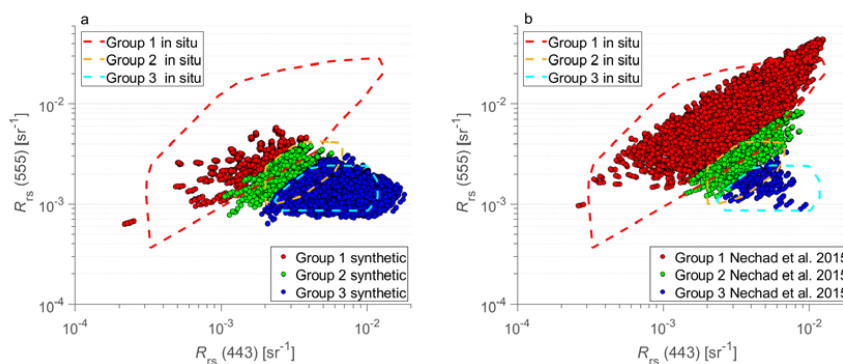
624

625

Figure 8. (a) $R_{rs}(555)$ as a function of $R_{rs}(443)$ for the synthetic dataset (colored data points) and in situ dataset (colored contours). (b) Same as panel (a) but for the synthetic dataset of Nechad et al. (2015) which was developed for coastal waters. The color coding refers to the optical water groups as indicated.

641   For comparing the AOPs from the synthetic database with in situ data, we have chosen two AOPs, the spectral

642   remote-sensing reflectance, $R_{rs}(\lambda)$, and the spectral diffuse attenuation coefficient of downwelling plane irradiance

643   averaged within the water column from the sea surface to the first attenuation depth, $<K_d(\lambda)>_1$. The scatter plot of

644   our synthetic data of $R_{rs}(555)$ vs. $R_{rs}(443)$ are depicted in Fig. 8a. For comparison, the range of in situ data is

645   illustrated by the dashed contour lines. These results are once again illustrated using color coding to represent

646   different optical water types, specifically Groups 1, 2, and 3. As seen, there is a relatively good agreement between

647   the synthetic data and the range of variability of the in situ data for Groups 2 and 3 (Fig. 8a). For Group 1 (very

648   turbid waters), however, the synthetic data exhibit a smaller range of variability compared with in situ data. This

649   result is not unexpected because our primary goal was to generate the synthetic database that is most representative

650 of open ocean pelagic environments as well as coastal areas where water turbidity is low to moderate rather than

651 very high. As described in section 2, turbid waters of Group 1 correspond to Optical Water Classes 1 and 2 as

652 defined in Mélin and Vantrepotte (2015). It is interesting to note that the synthetic optical database that was

653 developed by Nechad et al. (2015) for coastal waters shows a relatively good consistency between the synthetic

654 and in situ data for Group 1 (Fig. 8b). However, in contrast to our synthetic database, the synthetic data of Nechad

655 et al. (2015) exhibit a limited range of variability compared with in situ data for Groups 2 and 3. Thus, the synthetic

656 data of Nechad et al. (2015) for turbid waters in Group 1 can provide useful complementarity to our synthetic

657 database whose main focus is on water types from Groups 2 and 3.

658
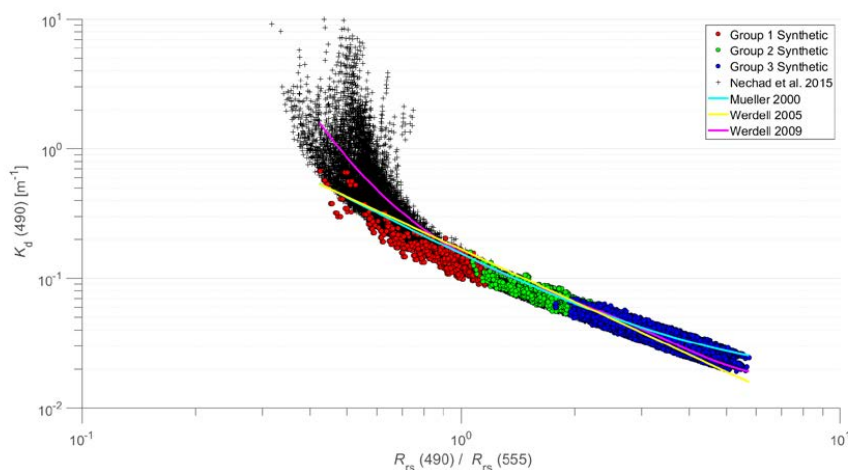
659

660 Figure 9. Scatter plot of $K_d(490)$ vs. the blue-to-green reflectance ratio, $R_{rs}(490)/R_{rs}(555)$, for the synthetic database. The red,
661 green, and blue data points represent the three optical water groups 1, 2, and 3, respectively. The black cross-mark data points
662 are from the Nechad et al. (2015) synthetic dataset. The curves representing the relationships developed by Mueller (2020),
663 Werdell (2005), and Werdell (2009) are also displayed. The $K_d(490)$ data points represent $<K_d(490)>_1$ for the present synthetic
664 database (colored data points), and the near-surface $K_d(490)$ calculated within the top 1 cm layer for the Nechad dataset (black
665 data points).
666
667

668   The scatter plot of the synthetic data of $<K_d(490)>_1$ as a function of blue-to-green band ratio of reflectance,

669 $R_{rs}(490)/R_{rs}(555)$, is shown in Fig. 9. These synthetic data are again color coded according to optical water classes

670 defined in terms of Groups, 1, 2, and 3. For comparison, a few empirical relationships between these AOP variables

671 established in previous analyses of field measurements are also displayed in Fig. 9 (Mueller, 2000; Werdell, 2005;

672 Werdell, 2009). The relationship of Mueller (2000) was formulated during the early phase of SeaWiFS satellite

673 mission to serve as an operational global algorithm for estimating $K_d(490)$ from ocean color observations. Werdell

674 (2005) provided an updated relationship with a primary goal to improve the estimation of $K_d(490)$ at low values

675 of $K_d(490)$ that correspond to high values of reflectance band ratio. Figure 9 shows that these two relationships are

676 generally consistent with our synthetic data across the entire range of variability encompassing data from Groups,

677 1, 2, and 3. This is reassuring given that the main purpose of our synthetic database and these two empirical

678 relationships is similar in a sense of targeting the optical variability within the global ocean dominated by open

679 ocean environments. Figure 9 also includes the relationship of Werdell (2009) which represents the most recent

680 update of global empirical algorithms for estimating $K_d(490)$ from different ocean color satellite sensors.

681   Specifically, the relationship of Werdell (2009) presented in Fig. 9 is referred to as KD2S and is based on SeaWiFS
682   spectral bands. In contrast to relationships of Mueller (2020) and Werdell (2005), the relationship of Werdell
683   (2009) deviates significantly from our synthetic data within the range of relatively high values of $\langle K_d(490) \rangle_1$
684   which correspond to relatively low values of $R_{rs}(490)/R_{rs}(555)$. It is remarkable that this deviation occurs within
685   the range where our synthetic data are classified as Group 1, so these are the optical water types associated with
686   high water turbidity. Another remarkable result illustrated in Fig. 9 is that the relationship of Werdell (2009) in
687   this range is quite consistent with the main trend observed within the synthetic database of Nechad et al. (2015)
688   which was developed for coastal environments. This result further supports the potential complementarity between
689   our synthetic database and database of Nechad et al. (2015).
690
691   **5 Summary**

692   We have generated a new synthetic database that consists of seawater IOPs as well as corresponding
693   radiometric quantities and AOPs within the ocean surface layer down to a depth of 50 m and at the sea surface.
694   The radiometric quantities and AOPs were obtained from radiative transfer (RT) simulations performed with
695   Hydrolight code using the IOPs as input to the calculations. The list of variables included in the database is
696   provided in Table 2. This database is organized following an easy to read netcdf structure and divided into two
697   subsets of data for which the file name identifies the sun zenith angle and the RT simulation scenario related to
698   the presence or absence of inelastic radiative processes within the water column. The first subset of data includes
699   the seawater spectral absorption and backscattering coefficients as well as sea-surface radiometric quantities
700   relevant to ocean color radiometry, $R_{rs}(\lambda)$, $L_w(\lambda)$, $E_d(z=0^+, \lambda)$, and $L_u(z=0^+, \lambda)$ where $z=0^+$ is just above the surface.
701   The surface and depth-profile values of several spectral radiometric quantities and AOPs, as well as *PAR* are
702   included in the second subset of data. The spectra of z_{eu} and z_1 are also provided in the second file. More details
703   on the organization and content of the database are included in readme file that is also provided in the database.

704   In closing, we present an example illustration of one of the radiometric variables included in the output data
705   files generated by RT simulations. We recall that the primary result of HydroLight simulations is the spectral
706   radiance that provides a comprehensive information about the angular distribution of light field, from which
707   different irradiances and AOPs are calculated. However, it is the spectral downwelling plane irradiance, $E_d(z, \lambda)$,
708   which has been the most commonly measured radiometric quantity in ocean optics, so in Fig. 10 we have chosen
709   to illustrate the HydroLight-simulated $E_d(z, \lambda)$ within the ocean surface layer down to a depth of 50 m. These
710   results are presented for three different scenarios of IOPs which are representative of three different optical water
711   types defined in terms of Group 1, Group 2, and Group 3 (see section 2). These RT simulations were performed
712   for the sun zenith angle of 30° in the presence of Raman scattering by water molecules and chlorophyll-a
713   fluorescence in the water column. In addition to significant differences in the variation of the spectral $E_d(\lambda)$ as a
714   function of depth z between the Groups 1, 2, and 3, Fig. 10 also illustrates distinct differences in the magnitude
715   and spectral behavior of the first optical attenuation depth, z_1. This quantity is equivalent to the inverse of the
716   diffuse attenuation coefficient, $\langle K_d(\lambda) \rangle_1$. As expected, the first attenuation depth z_1 is located much closer to the
717   ocean surface for data from Group 1 (Fig. 10a) compared with Group 2 (Fig. 10b) and Group 3 (Fig. 10c),
718   especially across the blue-green region of the spectrum. In the red part of the spectrum where pure water absorption
719   dominates the attenuation of $E_d(\lambda)$, the differences between the three groups are small. It is also notable that the

720 spectral behavior of z_1 for Group 3 (Fig. 10c) that represents relatively clear ocean waters is remarkably similar to

721 the spectral shape of pure water absorption coefficient.
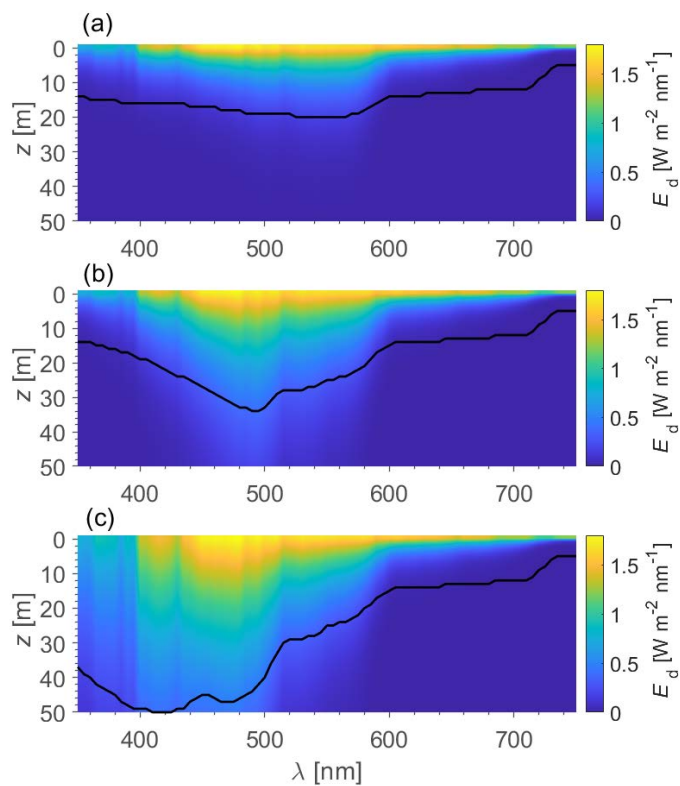
722
723 Figure 10. Examples of depth profiles of $E_d(z, \lambda)$ for a given IOP scenario from (a) the optical water group (OWG) 1, (b)
724 OWG 2, and (c) OWG 3. Radiative transfer simulations were performed for a sun zenith angle of 30° and included Raman
725 scattering by water molecules and chlorophyll-a fluorescence.

726

727   **Table 2:** Symbols, variables, and units for the various quantities included in the final synthetic optical database.

Symbol	Variable*	Units
z	Depth in water	m
λ	Light wavelength in vacuum	nm
a, b, b_b	Total absorption, scattering, and backscattering coefficients of seawater	m^{-1}
a_{nw}	Absorption coefficient of all non-water constituents	m^{-1}
a_{ph}, a_d, a_g	Absorption coefficients of phytoplankton, non-algal particles, and CDOM	m^{-1}
b_{nw}	Backscattering coefficient of all non-water constituents	m^{-1}
$b_{b\text{-}ph}$, $b_{b\text{-}d}$	Backscattering coefficients of phytoplankton and non-algal particles	m^{-1}
b_{nw}	Scattering coefficient of all non-water constituents	m^{-1}
b_{ph}, b_d	Scattering coefficients of phytoplankton and non-algal particles	m^{-1}
E_o, E_{od}, E_{ou}	Total, downwelling, and upwelling scalar irradiances	W m^{-2} nm^{-1}
E_d, E_u	Downwelling and upwelling plane irradiances	W m^{-2} nm^{-1}
L_w, L_u	Water-leaving and upwelling radiances	W m^{-2} sr^{-1} nm^{-1}
PAR	Photosynthetically Available Radiation defined as the total quantum scalar irradiance within the spectral range 400-700 nm	µmol photons s^{-1} m^{-2}
R_{rs}	Remote-sensing reflectance	sr^{-1}
K_x	Diffuse attenuation coefficients for upwelling and downwelling plane irradiances or upwelling radiance (the radiometric quantity is indicated by subscript x)	m^{-1}
μ_d, μ_u	Average cosines of downwelling and upwelling light fields	dimensionless
z_{eu}	Euphotic depth at which PAR is reduced to 1% of its surface value	m
z_1	First optical attenuation depth at which spectral E_d or PAR is reduced to 36.8% of its surface value	m

728 *All optical variables in the database are spectral and provided at different light wavelengths between 350 and
729 750 nm at 5 nm intervals and different depths within the water column between the sea surface and the 50 m
730 depth, except for R_{rs}, and L_w which are defined at the sea surface.

731

752 **References**

753 Antoine, D., Siegel, D. A., Kostadinov, T., Maritorena, S., Nelson, N. B., Gentili, B., Vellucci, V., Guillocheau,
754   N.: Variability in optical particle backscattering in contrasting bio-optical oceanic regimes, Limnol.
755   Oceanogr., 56(3), 955–973, https://doi.org/10.4319/lo.2011.56.3.0955, 2011.
756 Aurin, D., Mannino, A., Lary, D.: Remote sensing of CDOM, CDOM spectral slope, and dissolved organic carbon
757   in the Global Ocean, Appl. Sci., 8, 2687, https://doi.org/10.3390/app8122687, 2018.

758 Babin, M., Stramski, D., Ferrari, G. M., Claustre, H., Bricaud, A., Obolensky, G., Hoepffner, N.: Variations in the
759 light absorption coefficients of phytoplankton, nonalgal particles, and dissolved organic matter in coastal
760 waters around Europe. J. Geophys. Res., 108(C7), 3211, https://doi.org/10.1029/2001JC000882, 2003.
761 Bricaud, A., Babin, M., Claustre, H., Ras, J., Tieche, F.: Light absorption properties and absorption budget of
762 Southeast Pacific waters. J. Geophys. Res., 115, C0800910, https://doi.org/1029/2009JC005517, 2010.
763 Bricaud, A., Morel, A., Babin, M., Allali, K., H. Claustre, H.: Variation of light absorption by suspended particles
764 with chlorophyll a concentration in oceanic (case 1) waters: Analysis and implications for bio-optical models.
765 J. Geophys. Res., 103(C13), 31033–31044, https://doi.org/10.1029/98JC02712, 1998.
766 Bonelli, A. G., Vantrepotte, V., Jorge, D. S. F., Demaria, J., Jamet, C., Dessailly, D., Mangin, A., Fanton d'Andon,
767 O., Kwiatkowska, E., Hubert Loisel, H.: Colored dissolved organic matter absorption at global scale from
768 ocean color radiometry observation: spatio-temporal variability and contribution to the absorption budget,
769 Remote Sens. Environ., 265,112637, https://doi.org/10.1016/j.rse.2021.112637, 2021.
770 Casey, K. A., Rousseaux, C. S., Gregg, W. W., Boss, E., Chase, A. P., Craig, S. E., Mouw, C. B., Reynolds, R. A.,
771 Stramski, D., Ackleson, S. G., Bricaud, A., Schaeffer, B., Lewis, M. R., Maritorena, S.: A global compilation
772 of in situ aquatic high spectral resolution inherent and apparent optical property data for remote sensing
773 applications. Earth Syst. Sci. Data, 12, 1123–1139, https://doi.org/10.5194/essd-12-1123-2020, 2020.
774 Claustre, H., Sciandra, A., Vaulot, D.: Introduction to the special section Bio-optical and biogeochemical
775 conditions in the South East Pacific in late 2004: The BIOSOPE program, Biogeosciences, 5, 679–691,
776 https://doi.org/10.5194/bg-5-679-2008, 2008.
777 Craig, S. E., Lee, Z., Du, K.: Top of Atmosphere, Hyperspectral Synthetic Dataset for PACE (Phytoplankton,
778 Aerosol, and ocean Ecosystem) Ocean Color Algorithm Development, National Aeronautics and Space
779 Administration, PANGAEA, https://doi.org/10.1594/PANGAEA.915747, 2020.
780 Donlon, C., Berruti, B., Buongiorno, A., Ferreira, M.-H., Féménias, P., Frerick, J., Goryl, P., Klein, U., Laur, H.,
781 Mavrocordatos, C., Nieke, J., Rebhan, H., Seitz, B., Stroede, J., Sciarra, R.: The Global Monitoring for
782 Environment and Security (GMES) Sentinel-3 mission, Remote Sens. Environ., 120, 37-57,
783 https://doi.org/10.1016/j.rse.2011.07.024, 2012.
784 Durack, P. J., Wijffels, S. E., Boyer, T. P.: Long-term salinity changes and implications for the global water cycle,
785 in: Ocean Circulation and Climate: A 21st Century Perspective, edited by: Siedler, G., Griffies, S. M., Gould,
786 J., and Church, J. A., International Geophysics, vol. 103, p. 727–757, Academic Press, Elsevier, Oxford, UK,
787 https://doi.org/10.1016/B978-0-12-391851-2.00028-3, 2013.
788 Fournier G. R., Forand, J. L.: Analytic phase function for ocean water, in: Ocean Optics XII, edited by: Jaffe, J.
789 S., Proc. SPIE, Vol. 2258, p. 194–201, https://doi.org/10.1117/12.190063, 1994.
790 Huot, Y., Morel, A., Twardowski, M. S., Stramski, D., Reynolds, R. A.: Particle optical backscattering along a
791 chlorophyll gradient in the upper layer of the eastern South Pacific Ocean, Biogeosciences, 5, 495–507,
792 https://doi.org/10.5194/bg-5-495-2008, 2008.
793 IOCCG Report: Remote Sensing of Inherent Optical Properties: Fundamentals, Tests of Algorithms, and
794 Applications, in: Reports of the International Ocean-Colour Coordinating Group (IOCCG), edited by: Lee,
795 Z.-P., Lee, Z.-P. (ed.), No. 5, 126 pp., IOCCG, Dartmouth, NS, Canada, https://ioccg.org/wp-
796 content/uploads/2015/10/ioccg-report-05.pdf, 2006.
797 IOCCG Protocol Series: Inherent Optical Property Measurements and Protocols: Absorption Coefficient, in:
798 IOCCG Ocean Optics and Biogeochemistry Protocols for Satellite Ocean Colour Sensor Validation, edited
799 by: Neeley, A. R. and Mannino, A., Vol. 1.0, 78 pp., IOCCG, Dartmouth, NS, Canada,
800 https://doi.org/10.25607/OBP-119, 2018.
801 Jonasz, M., Fournier, G. R.: Light Scattering by Particles in Water: Theoretical and Experimental Foundations,
802 Academic Press, Amsterdam, 2007.
803 Jorge, D. S. F., Loisel, H., Jamet, C., Dessailly, D., Demaria, J., Bricaud, A., Maritorena, S., Zhang, X., Antoine,
804 D., Kutser, T., Bélanger, S., Brando, V. O., Werdell, J., Kwiatkowska, E., Mangin, A., Fanton d'Andon, O.: A
805 three-step semi analytical algorithm (3SAA) for estimating inherent optical properties over oceanic, coastal,
806 and inland waters from remote sensing reflectance, Remote Sens. Environ., 263, 112537.
807 https://doi.org/10.1016/j.rse.2021.112537, 2021.
808 Kishino, M., Takahashi, M., Okami, N., Ichimura, S.: Estimation of the spectral absorption coefficient of
809 phytoplankton in the sea, Bull. Mar. Sci., 37, 634–642, 1985.
810 Kostakis, I., Twardowski, M., Roesler, C., Röttgers, R., Stramski, D., McKee, D., Tonizzo, A., Drapeau, S.:
811 Hyperspectral optical absorption closure experiment in complex coastal waters. Limnol. Oceanogr. Methods,
812 19, 589–625, https://doi.org/10.1002/lom3.10447, 2021.
813 Kou, L., Labrie, D., Chylek, P.: Refractive indices of water and ice in the 0.65 to 2.5 μm spectral range, Appl.
814 Opt., 32, 3531–3540, https://doi.org/10.1364/AO.32.003531, 1993.
815 Loisel, H., Morel, A.: Light scattering and chlorophyll concentration in case 1 waters: a re-examination, Limnol.
816 Oceanogr., 43, 847-857, 1998.

Loisel, H., Stramski, D., Dessailly, D., Jamet, C., Li, L., Reynolds, R.A.: An inverse model for estimating the optical absorption and backscattering coefficients of seawater from remote-sensing reflectance over a broad range of oceanic and coastal marine environments, J. Geophys. Res. Oceans, 123, 2141–2171, https://doi.org/10.1002/2017JC013632, 2018.

Loisel, H., Vantrepotte, V., Norkvist, K., Mériaux, X., Kheireddine, M., Ras, J., Pujo-Pay, M., Combet, Y., Leblanc, K., Dall'Olmo, G., Mauriac, R., Dessailly, D., Moutin, T.: Characterization of the bio-optical anomaly and diurnal variability of particulate matter, as seen from scattering and backscattering coefficients, in ultra-oligotrophic eddies of the Mediterranean Sea, Biogeosciences, 8, 3295–3317, https://doi.org/10.5194/bg-8-3295-2011, 2011.

Loisel, H., Jorge, D. S. F., Reynolds, R. A., Stramski, D.: A synthetic database of hyperspectral ocean optical properties, Dryad, Dataset, https://doi.org/10.6076/D1630T, 2023.

Maritorena, S., Siegel, D. A., Peterson, A. R.: Optimization of a semianalytical ocean color model for global-scale applications, Appl. Opt., 41, 2705-2714, 2002.

Marshall, B. R., Smith, R. C.: Raman scattering and in-water optical properties, Appl. Opt., 29, 71–84, https://doi.org/10.1364/AO.29.000071, 1990.

Mélin, F., Vantrepotte, V.: How optically diverse is the coastal ocean? Remote Sens. Environ., 160, 235–251, https://doi.org/10.1016/j.rse.2015.01.023, 2015.

Mobley, C.: A numerical model for the computation of radiance distributions in natural waters with wind-roughened surfaces, Limnol. Oceanogr., 34, 1473–1483, https://doi.org/10.4319/lo.1989.34.8.1473, 1989.

Mobley, C. D.: Light and Water. Radiative Transfer in Natural Waters, Academic Press, San Diego, 1994.

Mobley, C. D.: Hydrolight Technical Note 10: Interpretation of Raman Scattering Computations, Sequoia Scientific, Bellevue, WA, 2012.

Mobley, C. D., Gentili, B., Gordon, H. R., Jin, Z., Kattawar, G. W., Morel, A., Reinersman, P., Stamnes, K., Stavn, R.: Comparison of numerical models for the computation of underwater light fields, Appl. Opt., 32(36), 7484–7504, https://doi.org/10.1364/AO.32.007484, 1993.

Mobley, C. D., Sundman, L. K.: HydroLight 5 EcoLight 5 Technical Documentation, Sequoia Scientific, Bellevue, WA, 2008.

Morel, A.: Optical properties of pure water and pure seawater, in: Optical Aspects of Oceanography, edited by: Jerlov, N. G. and Steeman Nielsen, E., eds., Academic Press, London, p. 1–24, 1974.

Morel, A.: Are the empirical relationships describing the bio-optical properties of case 1 waters consistent and internally compatible? J. Geophys. Res., 114, C01016, https://.org/10.1029/2008JC004803, 2009.

Morel, A., Huot, Y., Gentili, B., Werdell, P. J., Hooker, S. B., Franz, B. A.: Examining the consistency of products derived from various ocean color sensors in open ocean (Case 1) waters in the perspective of a multi-sensor approach, Remote Sens. Environ., 111(1), 69–88, https://doi.org/10.1016/j.rse.2007.03.012, 2007.

Mueller, J. L.: SeaWiFS algorithm for the diffuse attenuation coefficient, K(490), using water-leaving radiances at 490 and 555 nm, in: SeaWiFS Postlaunch Calibration and Validation Analyses, Part 3, edited by: Hooker, S. B. and Firestone, E. R., NASA/TM-2000-206892, Vol. 11, p. 24-27, NASA Goddard Space Flight Center, Greenbelt, Maryland, 2000.

Nechad, B., Ruddick, K., Schroeder, T., Oubelkheir, K., Blondeau-Patissier, D., Cherukuru, N., Brando, V., Dekker, A., Clementson, L., Banks, A. C., Maritorena, S., Werdell, P. J., Sá, C., Brotas, V., Caballero de Frutos, I., Ahn, Y.-H., Salama, S., Tilstone, G., Martinez-Vicente, M., Foley, D., McKibben, M., Nahorniak, J., Peterson, T., Siliò-Calzada, A., Röttgers, R., Lee, Z.,Peters, M.,Brockmann, C.: CoastColour Round Robin data sets: a database to evaluate the performance of algorithms for the retrieval of water quality parameters in coastal waters, Earth Syst. Sci. Data, 7, 319–348, https://doi.org/10.5194/essd-7-319-2015, 2015.

O'Reilly, J. E., Werdell, P. J.: Chlorophyll algorithms for ocean color sensors – OC4, OC5 & OC6, Remote Sens. Environ., 229, 32–47, https://doi.org/10.1016/j.rse.2019.04.021, 2019.

Pope, R. M., Fry, E. S.: Absorption spectrum (380-700 nm) of pure water. II. Integrating cavity measurements, Appl. Opt., 36(33), 8710–8723, https://doi.org/10.1364/AO.36.008710, 1997.

Reynolds, R. A., Stramski, D.: Optical characterization of marine phytoplankton assemblages within surface waters of the western Arctic Ocean, Limnol. Oceanogr., 64, 2478–2496, https://doi.org/10.1002/lno.11199, 2019.

Stavn, R. H.: Effects of Raman scattering across the visible spectrum in clear ocean water: A Monte Carlo study, Appl. Opt., 32(33), 6853–6863, https://doi.org/10.1364/AO.32.006853, 1993.

Stamnes, K., Thomas, G. E., J. J. Stamnes, J. J.: Radiative Transfer in the Atmosphere and Ocean, Second Edition, University Cambridge Press, 2017.

Stramski, D., Joshi, I., Reynolds, R. A.: Ocean color algorithms to estimate the concentration of particulate organic carbon in surface waters of the global ocean in support of a long-term data record from multiple satellite missions, Remote Sens. Environ., 269, 112776, https://doi.org/10.1016/j.rse.2021.112776, 2022.

Stramski, D., Li, L., Reynolds, R. A.: Model for separating the contributions of non-algal particles and colored dissolved organic matter to light absorption by seawater, Appl. Opt., 58, 3790–3806, https://doi.org/10.1364/AO.58.003790, 2019.

Stramski, D., Reynolds, R. A., Babin, M., Kaczmarek, S., Lewis, M. R., Röttgers, R., Sciandra, A., Stramska, M., Twardowski, M. S., Franz, B. A., Claustre, H.: Relationships between the surface concentration of particulate organic carbon and optical properties in the eastern South Pacific and eastern Atlantic Oceans, Biogeosciences, 5, 171–201, https://doi.org/10.5194/bg-5-171-2008, 2018.

Stramski, D., Reynolds, R. A., Kaczmarek, S., Uitz, J., Zheng, G.: Correction of pathlength amplification in the filter-pad technique for measurements of particulate absorption coefficient in the visible spectral region, Appl. Opt., 54, 6763–6782, https://doi.org/10.1364/AO.54.006763, 2015.

Sugihara, S., Kishino, M., Okami, M.: Contribution of Raman scattering to upward irradiance in the sea, J. Oceanogr. Soc. Japan, 40, 397–404, 1984.

Valente, A., Sathyendranath, S., Brotas, V., Groom, S., Grant, M., Taberner, M., Antoine, D., Arnone, R., Balch, W. M., Barker, K., Barlow, R., Bélanger, S., Berthon, J.-F., Besiktepe, S., Borsheim, Y., Bracher, A., Brando, V., Canuti, E., Chavez, F., Cianca, A., Claustre, H., Clementson, L., Crout, R., Frouin, R., García-Soto, C., Gibb, S. W., Gould, R., Hooker, S. B., Kahru, M., Kampel, M., Klein, H., Kratzer, S., Kudela, R., Ledesma, J., Loisel, H., Matrai, P., McKee, D., Mitchell, B. G., Moisan, T., Muller-Karger, F., O'Dowd, L., Ondrusek, M., Platt, T., Poulton, A. J., Repecaud, M., Schroeder, T., Smyth, T., Smythe-Wright, D., Sosik, H. M., Twardowski, M., Vellucci, V., Voss, K., Werdell, J., Wernand, M., Wright, S., and Zibordi, G.: A compilation of global bio-optical in situ data for ocean-colour satellite applications – version two, Earth Syst. Sci. Data, 11, 1037–1068, https://doi.org/10.5194/essd-11-1037-2019, 2019.

Voss, K. J.: A spectral model of the beam attenuation coefficient in the ocean and coastal areas, Limnol. Oceanogr., 37, 501-509, 1992.

Werdell, P. J.: OceanColor K490 algorithm evaluation, NASA Ocean Color Web, https://oceancolor.gsfc.nasa.gov/reprocessing/r2005.1/seawifs/k490_update/, 2005.

Werdell, P. J.: Diffuse attenuation coefficient (KD) for downwelling irradiance at 490-nm, NASA Ocean Color Web, https://oceancolor.gsfc.nasa.gov/reprocessing/r2009/kdv4/, 2009.

Werdell, P. J., Bailey, S. W.: An improved in situ bio-optical data set for ocean color algorithm development and satellite data product validation, Remote Sens. Environ., 98, 122–140, https://doi.org/10.1016/j.rse.2005.07.001, 2005.

Westberry, T. K., Boss, E., Lee, Z.-P.: Influence of Raman scattering on ocean color inversion models, Appl. Opt., 52, 5552–5561, https://doi.org/10.1364/AO.52.005552, 2013.

Zhang X., Hu, L.: Estimating scattering of pure water from density fluctuation of the refractive index, Opt. Express, 17, 1671–1678, https://doi.org/10.1364/OE.17.001671, 2009.

Zhang, X., Hu, L., He, M.-X.: Scattering by pure seawater: effect of salinity. Opt. Express 17(7), 5698–5710, https://doi.org/10.1364/OE.17.012685, 2009.

Zhang, X., Huot, Y., Bricaud, A., Sosik, H. M: Inversion of spectral absorption coefficients to infer phytoplankton size classes, chlorophyll concentration, and detrital matter, Appl. Opt. 54(18), 5805–5816, https://doi.org/10.1364/AO.54.005805, 2015.

Zheng, G., Stramski, D., Reynolds, R. A.: Evaluation of the Quasi-Analytical Algorithm for estimating the inherent optical properties of seawater from ocean color: Comparison of Arctic and lower-latitude waters, Remote Sens. Environ., 155, 194–209, https://doi.org/10.1016/j.rse.2014.08.020, 2014.