

Geophysical Research Letters



RESEARCH LETTER

10.1029/2020GL092189

Key Points:

- The model takes into account the vertical change of chlorophyll and improves the estimate of the subsurface photosynthetically available radiation field
- The model is applicable to autonomous platforms, ecosystem and biogeochemical models, and ocean color remote sensing
- Model comparison with independent *in-situ* radiometry confirms that satellite chlorophyll-*a* retrievals are unbiased worldwide

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

X. Xing,
xing@sio.org.cn

Citation:

Xing, X., & Boss, E. (2021). Chlorophyll-based model to estimate underwater photosynthetically available radiation for modeling, *in-situ*, and remote-sensing applications. *Geophysical Research Letters*, 48, e2020GL092189. <https://doi.org/10.1029/2020GL092189>

Received 17 DEC 2020
 Accepted 11 MAR 2021

Chlorophyll-Based Model to Estimate Underwater Photosynthetically Available Radiation for Modeling, *In-Situ*, and Remote-Sensing Applications

Xiaogang Xing¹  and Emmanuel Boss² 

¹State Key Laboratory of Satellite Ocean Environment Dynamics, Second Institute of Oceanography, Ministry of Natural Resources, Hangzhou, China, ²School of Marine Sciences, University of Maine, Orono, ME, USA

Abstract Accurate estimation of the underwater light field associated with photosynthetically available radiation (PAR) is critical to compute phytoplankton growth rate and net primary production (NPP), and to assess photo-physiological response of phytoplankton, such as changes in cellular pigmentation. However, methods to estimate PAR used in many previous studies lack in accuracy, likely resulting in significant bias in light-dependent products such as NPP derived from remote sensing, model simulations, or autonomous platforms. Here we propose and validate a new model for more accurate estimation of the subsurface PAR profile which uses chlorophyll concentration as its input. Validation is performed using 1,744 BGC-Argo profiles of chlorophyll fluorescence that are calibrated with surface satellite-derived chlorophyll concentration over their lifetime. The independent verification with the float's PAR sensors confirms the accuracy of satellite chlorophyll estimate worldwide and in the Southern Ocean in particular.

Plain Language Summary In this study, we propose a new model to compute the underwater light field available for phytoplankton growth. The model uses as its inputs the above water radiation (which can be obtained from public databases) and the subsurface distribution of chlorophyll-*a*, a pigment shared by all phytoplankton that can be estimated from sensors deployed on robotic platforms and that is used by many ocean ecosystem models. We test the model accuracy using data observed by the light sensors on autonomous profiling floats and find it performs well. The comparison also highlights that estimates of surface chlorophyll concentrations from satellites are unbiased worldwide, in contrast to some published accounts.

1. Introduction

The vertical distribution of underwater photosynthetically available radiation (PAR) is a necessary input for models of net primary production (e.g., review by Behrenfeld & Falkowski, 1997) and is key to interpret physiological response of phytoplankton to light, its change in cellular pigmentation (e.g., Behrenfeld et al., 2016) or the associated solar-stimulated fluorescence (Behrenfeld et al., 2009). In the last two decades, carbon-based primary production models (CbPM, Behrenfeld et al., 2005; Westberry et al., 2008), and photoacclimation models (Fox et al., 2020), provide novel approaches to estimate marine net primary production (NPP) from satellite measurements that have also been applied to *in-situ* platforms (Estapa et al., 2019; Yang et al., 2020). These models derive the vertical distribution of PAR by attenuating surface PAR using an estimate of its diffuse attenuation coefficient, $K_d(\text{PAR})$. As we show below, the methods used to determine $K_d(\text{PAR})$ in the above studies could be significantly biased and be improved using the model proposed here.

In some studies, the diffuse attenuation coefficient at 490 nm ($K_d(490)$) was used as a proxy of $K_d(\text{PAR})$ for calculating the mixed-layer median light level (PAR_{ML} , e.g., Behrenfeld et al., 2005 and Yang et al., 2020). However, since the blue light around 490 nm generally penetrates deepest in clear waters (e.g., Morel, 1988), this would lead to an overestimation of light level. In another study, Behrenfeld et al. (2016) used an equation proposed by Morel et al. (2007) to compute PAR_{ML} . This approach, however, has not been thoroughly evaluated for the full range of MLDs in the ocean (e.g., 10–500m).

Lee et al. (2005, hereafter Lee05) proposed a depth-resolved $K_d(\text{PAR})$ algorithm, based on the sea-surface inherent optical properties (IOPs). Xing et al. (2020) have analyzed its performance in the global oceans

© 2021. The Authors. Geophysical Research Letters published by Wiley Periodicals LLC on behalf of American Geophysical Union.
 This is an open access article under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

with an extensive BGC-Argo data set and found it works well near the sea surface, but exhibits a significant underestimation in subsurface oligotrophic waters. This is not surprising as Lee05's algorithm assumes constant IOPs independent of depth, and in oligotrophic waters the presence of a deep chlorophyll maxima (DCM) increases the attenuation of light at depth. In addition, its applicability is limited, given that IOPs are not readily available in observations from most *in-situ* platforms and in ecosystem models.

Here we propose a new chlorophyll-based algorithm to compute the subsurface light field. The algorithm is evaluated by comparing the predicted light field to the observed underwater light field, as well as the associated diffuse attenuation coefficient, $K_d(\text{PAR})$, based on the PAR observations from the global BGC-Argo array. In addition, we evaluate its performance relative to previous approaches in retrieving the mixed-layer median light level from ocean color remote sensing, a necessary input to some NPP models.

2. Materials and Methods

2.1. BGC-Argo Data Set

To provide input data as well as to evaluate the model, we use measurements from 193 BGC-Argo floats, operated from October 2012 to March 2020, with data compiled by the Global Data Assembly Center (GDAC) of the Argo program. These floats are equipped with Satlantic OCR504 radiometers and Wet Labs ECO in-vivo chlorophyll-*a* fluorometers. Each OCR504 has four channels, of which three channels measure downwelling irradiance (E_d) at 380 (or 443), 412, and 490 nm, while the fourth channel measures the instantaneous planar Photosynthetically Available Radiation (iPAR, the downwelling photon flux integrated from 400 to 700 nm). All profiles are classified into 10 regions (see criteria in Table S1), as shown in Figure S1, so that regional biases can be evaluated. All 35,729 raw iPAR profiles are quality controlled following the scheme detailed in Organelli et al. (2016). Finally, only Flag-1 (18,359, 51.4%) and Flag-2 (5,315, 14.9%) profiles are used in this study. The sea-surface iPAR(0^-) is retrieved following Xing et al. (2020), and the layer-averaged diffuse attenuation coefficient at depth z , $K_d(\text{iPAR}, z)$, is defined as:

$$K_d(\text{iPAR}, z) = \frac{1}{z} \ln \left(\frac{\text{iPAR}(0^-)}{\text{iPAR}(z)} \right) \quad (1)$$

The 1% light layer depth (sometimes referred to as the euphotic layer depth, e.g., Ryther, 1956), $z_{1\%}$, is determined as the depth where iPAR reaches 1% of iPAR(0^-). For evaluating $K_d(\text{iPAR})$ within the mixed layer, MLD is calculated based on a threshold potential density difference (0.03 kg m^{-3}) from a near-surface value at 10 m, following De Boyer Montégut et al. (2004).

2.2. Satellite Data

Daily 4-km-resolution remote-sensing reflectance (R_{rs}), and standard chlorophyll-*a* concentration (Chl- $_{\text{sat}}$), provided by the MODerate resolution Imaging Spectroradiometer (MODIS) on the satellite "EOS PM (Aqua)," are used. The matchup criteria between float and satellite data is similar to Bailey and Werdell (2006): The median value within a 3×3 pixel box centered on each profile's position on the same date, is extracted from satellite data as the "matchup" value; Furthermore, we require that (1) the corresponding BGC-Argo observation time is within 3 h of satellite overpass; (2) there are at least 50% valid values in the box; and (3) for R_{rs} at any band, the coefficient of variance (CV) within the box must be lower than 0.15. We obtain 1,744 valid satellite-float match-up profiles (Table S1).

2.3. Existing PAR Models

The daily PAR profile, $\text{PAR}(z)$, is calculated from Equation 2, for a given sea surface PAR_{Sat} , and estimated layer-averaged $K_d(\text{PAR}, z)$ profile.

$$\text{PAR}(z) = (\text{PAR}_{\text{Sat}} \cdot \alpha) \cdot \exp[-K_d(\text{PAR}, z) \cdot z] \quad (2)$$

Here, α represents the transmission coefficient of sun light through the air-sea interface provided by Mobley and Boss (2012). Note that here $K_d(\text{PAR}, z)$ represents the diffuse attenuation coefficient for the daily PAR which is well approximated by $K_d(\text{iPAR}, z)$ at local noon (Wei & Lee, 2013).

Given our data set, this study only evaluates the performance of our algorithm and other three published methods/models to estimate $K_d(\text{iPAR}, z)$ and $\text{iPAR}(z)$. The first one, proposed by Behrenfeld et al. (2005) for satellite data and followed by Yang et al. (2020) for NPP estimation with BGC-Argo, used the $K_d(490)$ satellite product as the proxy for $K_d(\text{iPAR})$. These studies implicitly assumed, (1) $K_d(490)$ at the first optical depth ($z_{\text{pd}} = 1/K_d(490)$) is equal to $K_d(\text{iPAR})$, and (2) $K_d(\text{iPAR})$ is depth-independent. Here we use $K_d(490)$ satellite product derived from the semi-analytical algorithm of Lee et al. (2013, from here on “Lee13-490”).

The second one, called as “Morel07,” employs the $K_d(490)$ - $K_d(\text{iPAR})$ empirical relationship (Equation 3) established by Morel et al. (2007) for a layer with a thickness of twice optical depth ($2z_{\text{pd}}$). Here we used $K_d(490)$ computed from Lee13-490 and Equation 3, for deriving $K_d(\text{iPAR})$ based on Morel07:

$$K_d(\text{iPAR}) = 0.0665 + 0.874K_d(490) - 0.00121 / K_d(490). \quad (3)$$

The third model is “Lee05,” a depth-dependent $K_d(\text{iPAR})$ model which assumes depth independent IOPs in its calculation. Its inputs are $a(490)$ and $b_b(490)$, which could be retrieved, for satellite-based applications from the Quasi-Analytical Algorithm (QAA) of Lee et al. (2002). In this study, we used its latest version QAA (QAAv6, updated in 2014), with the pure-seawater absorption coefficients from Lee et al. (2015). $K_d(\text{iPAR}, z)$ is expressed as the function of depth (z), as well as two diffuse attenuation coefficients K_1 and K_2 :

$$K_d(\text{iPAR}, z) = K_1 + \frac{K_2}{\sqrt{1+z}} \quad (4)$$

Both K_1 and K_2 are expressed as the functions of $a(490)$, $b_b(490)$, and solar zenith angle (θ):

$$K_1 = \left[\chi_0 + \chi_1 \cdot a(490)^{0.5} + \chi_2 \cdot b_b(490) \right] (1 + \alpha_0 \sin \theta) \quad (5)$$

$$K_2 = \left[\zeta_0 + \zeta_1 \cdot a(490) + \zeta_2 \cdot b_b(490) \right] (\alpha_1 + \alpha_2 \cos \theta) \quad (6)$$

χ_{0-2} , ζ_{0-2} , and α_{0-2} are model constants tuned by fitting numerical simulations through HydroLight (Mobley, 1994) in Lee et al. (2005).

Once the $K_d(\text{iPAR})$ profile is determined, the instantaneous PAR profile ($\text{iPAR}(z)$) can be estimated with Equation 1, as long as the surface value (i.e., $\text{iPAR}(0^-)$) and appropriate sun angle are known. As for computation of the daily integrated PAR(z) estimates, we recommend that $K_d(\text{iPAR})$ around local noon be converted to $K_d(\text{PAR})$ following Equation 12 in Wei and Lee (2013). Additionally, one needs to be aware that while the technology used to measure PAR (and hence use it to develop algorithms) is typically measuring the planar downwelling irradiance, for NPP computation one should preferentially use scalar PAR. Wei and Lee (2013) compute, based on radiative transfer computations, a mean absolute percentage error of 10.5 between $K_d(\text{PAR})$ and the diffuse attenuation of scalar PAR.

2.4. Chlorophyll-Based PAR Model

In this study, we propose a chlorophyll-based PAR model, which combines the clear-sky spectral irradiance model of Gregg and Carder (1990) (GC90), and the empirical Chla- $K_d(\lambda)$ spectral relationships of Morel and Maritorena (2001) (MM01), with the input of Chla profile. We call the model as “GCMM,” representing a combination of two models (GC90 + MM01). Its procedures include (Figure S2): (1) a clear-sky sea-surface downwelling irradiance $E_d(\lambda, 0^+)$ is obtained from GC90 with wavelength resolution of 1 nm, and then transformed to $E_d(\lambda, 0^-)$ through the transmission coefficient α ; (2) The Chla profile is interpolated to each meter (1, 2, 3 m, etc.); (3) The corresponding $K_d(\lambda, z)$ at each depth z is determined based on MM01 (Equation 7);

$$K_d(\lambda, z) = K_w(\lambda) + \chi(\lambda) \cdot \text{Chla}(z)^{e(\lambda)} \quad (7)$$

(4) For each wavelength, $E_d(\lambda)$ is attenuated with K_d level by level:

$$E_d(\lambda, z) = E_d(\lambda, z-1) \cdot \exp[-K_d(\lambda, z)] \quad (8)$$

For $z = 1$ m, $(z-1)$ represents the sea surface (0^-); finally, all $E_d(\lambda, z)$ at each depth are wavelength-integrated to $\text{iPAR}(z)$, and sea-surface $E_d(\lambda, 0^-)$ values are wavelength-integrated to $\text{iPAR}(0^-)$:

$$\text{iPAR}(z) = \int_{400}^{700} \frac{\lambda}{hc} E_d(\lambda, z) d\lambda, \quad (9)$$

where, h is Plank's constant, and c the speed of light in vacuum; then $K_d(\text{iPAR}, z)$ is determined through Equation 1. $K_d(\text{iPAR}, z)$ computed by the different models evaluated here is then used to estimate instantaneous PAR at any depth for any given $\text{iPAR}(0^-)$. As for the daily PAR estimate, the relationship proposed by Wei and Lee (2013, their Equation 12) should be employed to convert the local-noon $K_d(\text{iPAR})$ to the daily $K_d(\text{PAR})$. To account for reduction in sun light relative to the clear sky model of GC90, we use the local $\text{iPAR}(0^-)$ from float or $\text{PAR}(0^-)$ from satellite and $K_d(\text{iPAR}, z)$ derived from GCMM to compute $\text{iPAR}(z)$ or $\text{PAR}(z)$. This, de-facto, assumes that the differences between the PAR attenuation of cloudy skies and clear skies are similar. Indeed, our results (see below) suggest it does a reasonable job in predicting iPAR even to the depth of 1.5-fold $z_{1\%}$. The R programmes of GCMM are available in <https://github.com/BGC-Argo/GCMM>.

For satellite applications, the Chla profile could be assumed to be homogenous, or a vertical profile is estimated from surface Chla (Uitz et al., 2006; Text S1 and Figure S3). For *in-situ* platforms (e.g., BGC-Argo or glider), the input could be the observed Chla profile. In this study, “GCMM_{surf}” represents the satellite-Chla-based method, with inputs of Chla_{sat} and homogenous assumption; “GCMM_{Uitz}” represents another satellite-Chla-based method, with inputs of Uitz06's estimated Chla profile from Chla_{sat} ; and “GCMM_{prof}” uses the the float-observed Chla profiles, whose quality-control procedure is shown in Text S2.

2.5. Statistical Metrics

The statistical metrics used in this study, include the root mean square error (ε), mean error (δ), unbiased root mean square error (σ), and relative mean error ($\delta\%$), defined as:

$$\varepsilon = \sqrt{\frac{1}{n} \sum_{i=1}^n (E_i - M_i)^2} \quad (10)$$

$$\delta = \frac{1}{n} \sum_{i=1}^n (E_i - M_i) \quad (11)$$

$$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n \left[\left(E_i - \frac{1}{n} \sum_{j=1}^n E_j \right) - \left(M_i - \frac{1}{n} \sum_{k=1}^n M_k \right) \right]^2} \quad (12)$$

$$\delta\% = \frac{1}{n} \sum_{i=1}^n \frac{(E_i - M_i)}{M_i} \cdot 100\% \quad (13)$$

Here, M_i represents the float-measured value, E_i represents the satellite-retrieved or model-estimated value, and n is the number of observations. ε represents the total deviation/difference, δ represents the systematic bias, σ represents the variance, and $\delta\%$ represents the relative system bias.

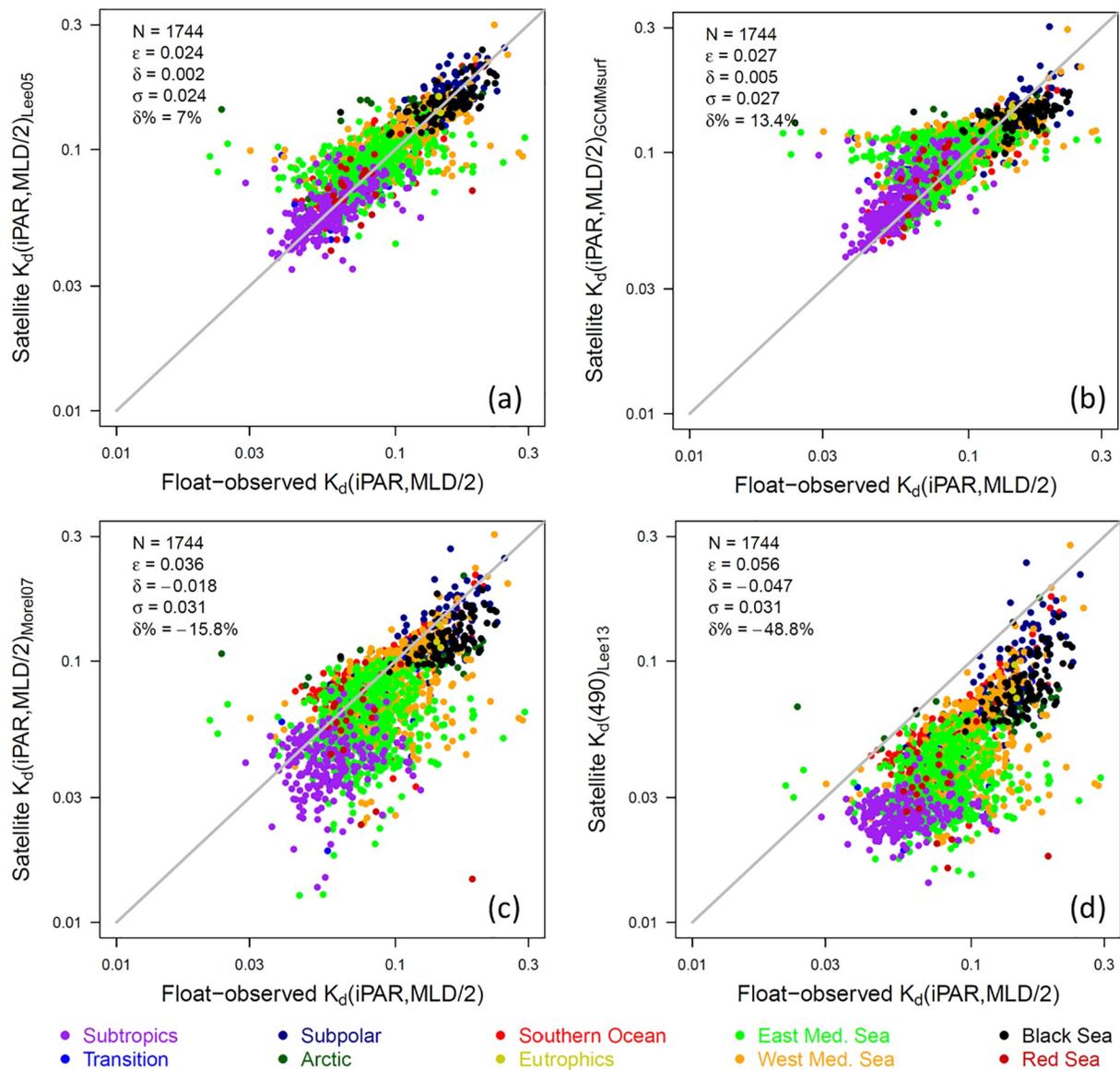


Figure 1. Scatter plot between float-observed $K_d(iPAR, MLD/2)$ versus satellite-estimated ones with different models. (a) Lee05 model, (b) GCMMSurf model, (c) using Morel07's empirical $K_d(490)-K_d(iPAR)$ relationship, and (d) using $K_d(490)$ derived from Lee et al. (2013, Lee13) as the proxy of $K_d(iPAR)$.

3. Results and Discussion

3.1. Retrieval of Mixed-Layer Median iPAR

Mixed-layer median iPAR requires an estimation of a representative $K_d(iPAR)$, that is, $K_d(iPAR, MLD/2)$. Here we compare estimates from different models using the BGC-Argo iPAR data (note that the iPAR and Chla are measured by different sensors on the floats). The IOPs-based model (Lee05) performs best (Figure 1a), with lowest error ($\epsilon = 0.024 \text{ m}^{-1}$) and negligible systematic bias ($\delta = 0.002 \text{ m}^{-1}$). GCMMSurf model also works well (Figure 1b), with a little higher bias, and similar ϵ and σ to Lee05. A consistency test (Figure S4) shows a negative correlation between its retrieval error and the float-observed $K_d(iPAR)$ ($r = -0.66$), indicating a tendency of overprediction in clear waters (lower K_d) and underprediction in turbid waters (high K_d). The main areas with this issue include the Mediterranean Sea and Black Sea, suggesting the cause to be a CDOM influence on $K_d(iPAR)$ retrieval in the GCMMSurf model. It is noteworthy that the other models show a similar issue to different extents (Figure S4), implying that a potential for CDOM-driven bias exists in all models. A test of Lee05 and GCMMSurf sensitivity to solar zenith angle shows that both have

good performance for zenith angles varying between 20° and 60° (Figures S5–S6); GCMM_{surf} has a slight overprediction at low zenith (<20°, Figure S5b), while Lee05 exhibits overprediction at high zenith (>60°, Figure S6c).

Using Morel07's empirical relationship (for $2z_{pd}$, Equation 5 in this study, Equation 9 in Morel et al., 2007) displays an underestimation of about 16% on average K_d (iPAR, MLD/2) (Figure 1c), and with higher variability ($\sigma = 0.031 \text{ m}^{-1}$). This is likely due to the fact that the relationship was determined at $2z_{pd}$, without depth-dependence. We also test Morel07's empirical relationship for z_{pd} (their Equation 9 in Morel et al., 2007), which performs similarly with $\epsilon = 0.032 \text{ m}^{-1}$ (figure not shown). The use of $K_d(490)$ as a proxy for K_d (iPAR, MLD/2) performs worst, underestimating by about 49% (Figure 1d). It follows that studies using $K_d(490)$ for primary production models (e.g., Fox et al., 2020; Yang et al., 2020) underestimated K_d (iPAR, MLD/2) significantly and thus likely overestimated iPAR, PAR, as well as NPP.

The retrieved median light level, iPAR(MLD/2), is consistent with these results (Figure 2). Lee05 outperforms the other models with lowest bias ($\delta\% = -0.3\%$, Figure 2a). GCMM_{surf} works as well, with lowest absolute errors ($\epsilon = 163 \text{ } \mu\text{mol photons m}^{-2} \text{ s}^{-1}$, Figure 2b), while exhibiting some underprediction in the high-light range (e.g., in the Eastern Mediterranean Sea and subtropical gyres), where surface Chla is very low, close to the detection limit of the satellite algorithm. By contrast, using Morel07 (Figure 2c) and $K_d(490)$ (Figure 2d) leads to overestimation of mixed-layer median light level by $\sim 17\%$ and 100% , respectively. We also test Morel07's empirical relationship for z_{pd} (figure not shown), which slightly underestimates iPAR(MLD/2) by -6.2% , but with large errors ($\epsilon = 227 \text{ } \mu\text{mol photons m}^{-2} \text{ s}^{-1}$).

3.2. Retrieval of the Vertical Distribution of iPAR

The good performance of Lee05 and GCMM_{surf} in retrieving K_d (iPAR, MLD/2) is not only due to their depth-resolving characteristics for different MLDs, but also because of the assumption of homogenous IOPs and Chla in the mixed layer (which is consistent with expectation). However, the later assumption is likely wrong below the MLD, particularly in the presence of DCM which is ubiquitous in oligotrophic waters. More Chla below the ML would attenuate light faster than near the surface. Thus, we expect to see an underestimation of K_d (iPAR) and overestimation of iPAR at subsurface in oligotrophic waters, for both surface-based Lee05 and GCMM_{surf} models (as shown in Xing et al., 2020).

An example of a profile observed in the Western Mediterranean Sea is provided in Figure 3. The MLD is at 15 m, $z_{1\%}$ reaches ~ 55 m, and DCM appears at ~ 38 m (Figure 3a). Uitz06-estimated Chla profile has a similar DCM depth to the observed one, but with much lower Chla concentration at DCM (Figure 3a). The observed iPAR profile (Figure 3b) is characterized by two layers with larger attenuation, one near the sea surface and the other around the DCM. The first one is driven by the faster attenuation of the red light due to absorption by water (Lee 2009; Morel et al., 1988) and the second one is due to the higher chlorophyll-*a* concentration, which increases both absorption and backscattering losses. All four methods estimate iPAR well near sea surface (Figure 3b), in other words, all resolve the issue of faster loss of red light near the surface. However, Lee05 and GCMM_{surf} do not resolve the DCM, and thus result in a uniform K_d (iPAR) below the MLD, shown as a straight line on the semi-logarithmic plot (Figure 3b). Consequently, both models underestimate K_d (iPAR) and overestimate iPAR near the DCM. By contrast, GCMM_{prof} calculates the vertical change of K_d (iPAR) based on the float-observed Chla with inhomogeneous distribution, and thus provides an estimated iPAR profile more consistent with the observed one (Figure 3b). Lee05 and GCMM_{surf} overestimate iPAR below about $0.6z_{1\%}$ while GCMM_{prof} performs well throughout the 1.5-fold $z_{1\%}$ layer (Figure 3c).

The relative errors ($\delta\%$) for all profiles are presented in Figure 4. They are first classified into two types, denoted as the "Mixed Type" (maximal Chla within the ML) and "DCM Type" (maximal Chla value below the ML). In Mixed-Type waters (371 profiles in total), the maximal Chla is located within the ML, with relatively shallow $z_{1\%}$, and overall, all four models estimate the iPAR profile well (Figure 4a). Lee05 has a slight underestimation, with lowest relative errors ($\delta\% = 12.8\%$) from surface to $1.5z_{1\%}$; GCMM_{surf} and GCMM_{Uitz} exhibit an overestimation from $0.2z_{1\%}$; GCMM_{prof} performs better than other two GCMM models ($\delta\% = 16.8\%$), but with a slight overestimation at depth from $0.8z_{1\%}$ to $1.5z_{1\%}$. It implies that Lee05's homogenous assumption works down to the 1.5-fold 1% light level depth for a mixed profile.

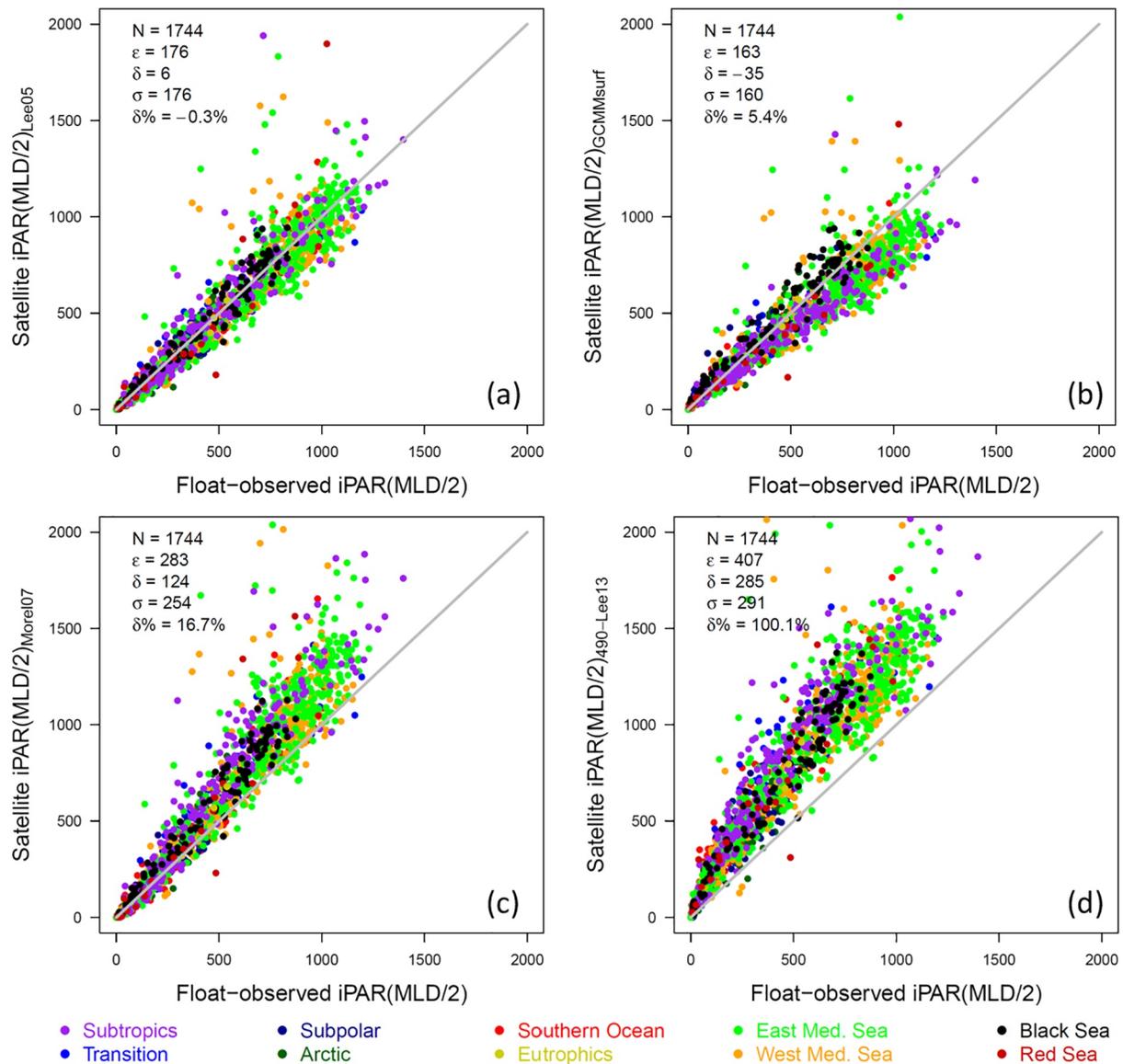


Figure 2. Scatter plot between float-observed mixed-layer median light intensity, iPAR (MLD/2) versus satellite-estimated ones with different models. (a) Lee05 model, (b) GCMM model, (c) using Morel07's empirical $K_d(490)$ - $K_d(iPAR)$ relationship, and (d) using $K_d(490)$ derived from Lee et al. (2013, Lee13) as the proxy of $K_d(iPAR)$.

As for the DCM-Type waters (1,334 profiles in total), Lee05 and GCMM_{surf} overestimate iPAR from half $z_{1\%}$ (Figure 4b), and reach an error of $\sim 100\%$ at $z_{1\%}$, and even more than 300% at $1.5z_{1\%}$ (not shown in Figure 4b). GCMM_{Uitz} demonstrates an improvement relative to GCMM_{surf}, as the former estimates a DCM below the MLD, but the latter assumes no DCM at all. However, because Uitz06 is established based on a statistical analysis of observational data, and thus with an estimated Chla profile representing a global mean for a given sea-surface Chla, it can neither account for regional differences in DCM magnitude and depth, nor generate an abrupt Chla change driven by short-term physical and biological processes (e.g., meso-/sub-mesoscale processes and synoptic-scale events). Nevertheless, GCMM_{Uitz} performs significantly better than GCMM_{surf} ($\delta\% = 30.8\%$, compared to 119.8%), exhibiting an overestimation below $\sim 0.8z_{1\%}$, and $\sim 100\%$ of relative error at 1.5 -fold $z_{1\%}$. This emphasizes the difficulty of obtaining accurate estimates of the light intensity profile from satellites. Although DCM depth and magnitude often display some statistical relationships with surface Chla (e.g., Mignot et al., 2011; Uitz et al., 2006), it is difficult to accurately derive the subsurface Chla distribution from surface observation. Comparatively, GCMM_{prof} performs well down to 1.5 -fold $z_{1\%}$,

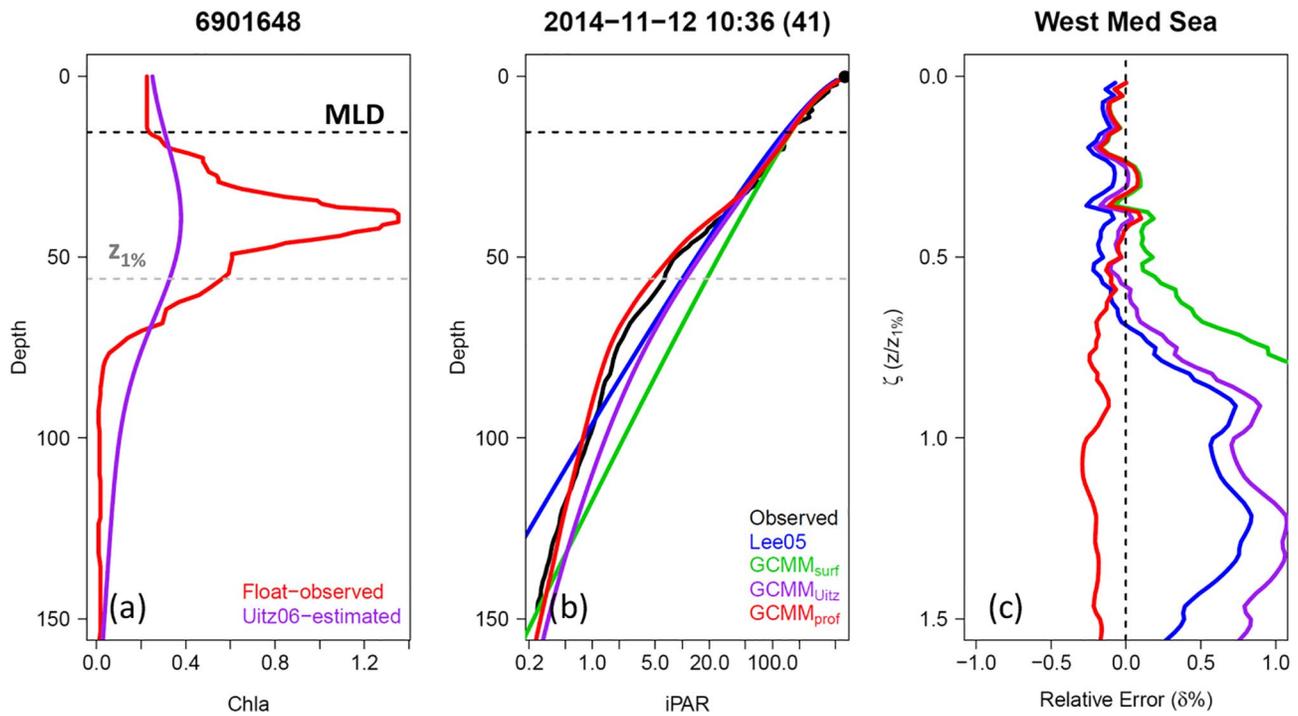


Figure 3. An example of model-estimated iPAR profiles and that observed by Float 6901648 at November 12, 2014 (Cycle 41) in the western Mediterranean Sea. (a) Float-observed Chla after quality control (red) and Uitz06-estimated Chla profile (purple); (b) Observed iPAR (black) and estimated ones from Lee05 (blue), GCMM_{surf} (green), GCMM_{Uitz} (purple), and GCMM_{prof} (red); (c) the relative error ($\delta\%$) of four estimated iPAR, (Est.-Obs.)/Obs., the depth is normalized to $z_{1\%}$. Black dashed line represents the mixed layer depth (MLD), gray dashed line represents the 1% light level depth ($z_{1\%}$).

with a mean relative error of 5.3%, suggesting a good consistency between iPAR and Chla profile, which are independently measured by the BGC-Argo floats.

Regional analysis reveals that GCMM_{prof} outperforms the other algorithms in all oligotrophic waters (mostly with DCM-type profiles), for example, subtropical gyres (Figure S7a), eastern and western Mediterranean Sea (Figures S7b–S7c), as well as the Transition area (Figure S7d). In the subtropical gyres (Figure S7a), although GCMM_{Uitz} has a lowest systematic bias ($\delta\% = 26\%$), GCMM_{prof} works better in the upper 1% light level layer. Lee05 outperforms all three GCMM models at high latitudes (where there are more Mixed-type profiles), for example, the Southern Ocean (Figure S7g) and subpolar (Figure S7h). In the Black Sea (Figure S7f) and Arctic Ocean (Figure S7i), although none works well, GCMM_{prof} displays the best statistical performance.

Besides, the good performance of GCMM_{surf} model suggests that, (1) satellite Chla could be used for the retrieval of mixed-layer median K_d (iPAR) (Figure 1b) and near-surface iPAR (Figure 4) with the GCMM model; and (2) it proves, although indirectly, that no significant regional bias exists for the satellite estimate of Chla retrieved with the OCI algorithm. This is of particular interest in the Southern Ocean (Figure S8a), where the global satellite Chla algorithm have long been challenged (e.g., Dierssen & Smith, 2000; Johnson et al., 2013; Kahru & Mitchell, 2010) though others (Haëntjens et al., 2017; Marrari et al., 2006) suggested they are unbiased. We test the GCMM_{surf} model's sensitivity in the SO by multiplying and dividing the OCI Chla product by a factor of 2, as the input to retrieve K_d (iPAR, MLD/2), and find significant bias (+22.5% and -13.8%; Figures S8b–S8c). This suggests that the model is sufficiently sensitive to the input Chla, and thus useful to evaluate the possible system bias of Chla algorithms. Note that in this analysis, the float-observed and satellite-derived subsurface light fields are mutually independent, except for sharing the same MLDs observed by floats.

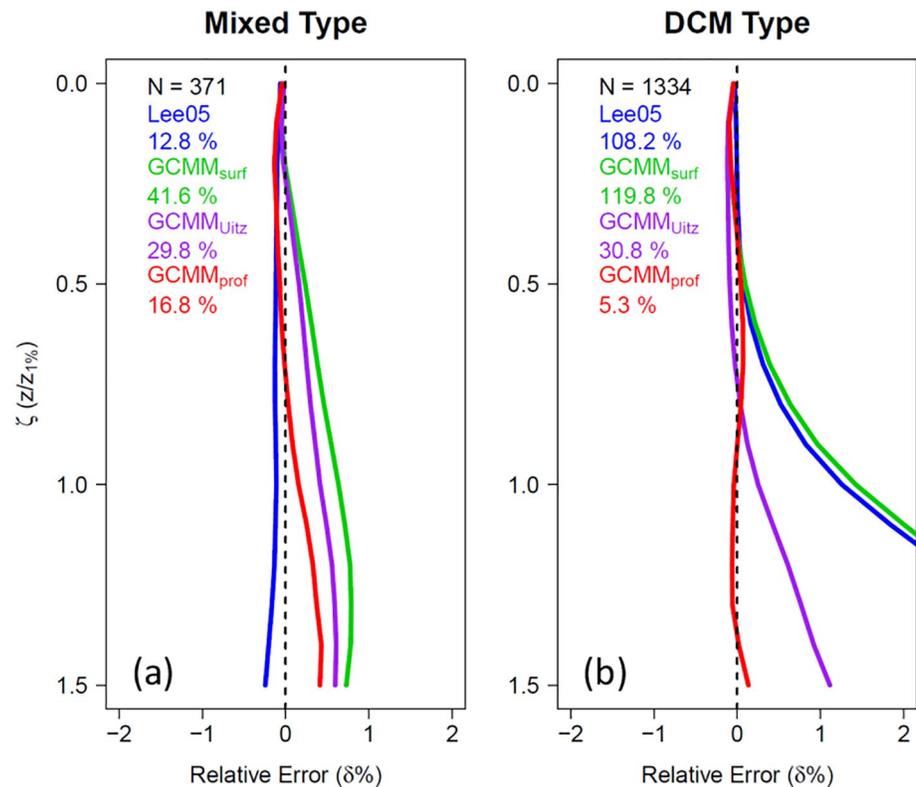


Figure 4. Averaged relative error ($\delta\%$) at each depth normalized by $z_{1\%}$, for four estimated iPAR by the model satellite IOPs-based Lee05 (blue), satellite Chla-based GCMM_{surf} (green), satellite Chla and Uitz06-based GCMM_{Uitz} (purple) and float Chla-based GCMM_{prof} (red), for Mixed Type (maximal FChla locates within the mixed layer) (a) and DCM Type (maximal FChla locates below the mixed layer) (b) waters, respectively. The numbers in legend represent the relative error ($\delta\%$) averaged from surface to the 1.5-fold 1% light level depth ($1.5z_{1\%}$).

4. Summary

Existing models for subsurface light field are evaluated in this study, with the recent BGC-Argo database. We find that the Lee05 model provides the most accurate iPAR(z) retrieval near sea surface and within the mixed layer, and we recommend it be employed for ocean color satellite-based mixed-layer photoacclimation and primary production models. Models using $K_d(490)$ as the proxy for $K_d(\text{iPAR})$ and the formulation of Morel07 significantly underestimate iPAR attenuation resulting in an overestimation of iPAR. Lee05 works well not only near the sea surface, but also within the 1.5-fold 1% light level layer, except for DCM-Type waters. In such waters, for example, the subtropical gyres and Mediterranean Sea, a DCM appears remarkably at subsurface, leading to an obvious overestimation of iPAR from half $z_{1\%}$ to $1.5z_{1\%}$ (relative error of over 300%).

The chlorophyll-based PAR model, GCMM, proposed here, can be used for both remote-sensing (with the homogenous assumption as in Lee05, GCMM_{surf}) and *in-situ* platforms with depth-resolved data (GCMM_{prof}). While Lee05 outperforms GCMM_{surf} in general, it requires inputs seldom available from *in-situ* platforms. GCMM_{prof} performs best for *in-situ* platforms that resolve the vertical structure of chlorophyll and can easily be implemented in ecosystem models where chlorophyll-a concentration is included. We test the Uitz06 model to provide a Chla-profile estimate from the satellite Chla and then merge it with GCMM model (i.e., GCMM_{Uitz}), which shows an improvement of estimated iPAR when compared with GCMM_{surf}. However, the Chla profiles retrieved were often significantly biased. Improvement in estimating vertical profile from surface observations may come from taking more environmental data into consideration (e.g., wind, temperature) and additional *in-situ* Chla data (e.g., BGC-Argo). Besides, the novel machine learning techniques are potentially helpful to better retrieve underwater light fields from surface observations.

While it may seem that using Chla and PAR is a step backward in comparison to using IOPs and spectral radiometry, the reality is that IOPs are scarcely measured globally (except for $b_{bp}(700)$) and are also scarcely parameterized into biogeochemical models (for exceptions see IOCCG, 2020). In addition, both PAR and Chla are considered mature products from Ocean Color (as of today, no ocean color mission provides spectral surface irradiance as a product and IOPs are only distributed as products by a handful of missions). Hence, it is important that, the community that parameterizes biogeochemical processes based on Chla and PAR, uses unbiased and validated relationships such as we provide here. Besides the inclusion of vertical structure, future improvement should include conversions of planar downwelling iPAR to scalar PAR (e.g., Frouin et al., 2018; Mobley & Boss, 2012; Wei & Lee, 2013), and the potential sources of uncertainty in the present GCMM model (including the Chla- K_d relationship variability at different solar zenith angles and under heavy clouds, and the change of average cosine under clouds relative to the clear-sky condition [Frouin et al., 2018]), to provide even less biased inputs to NPP algorithms.

Data Availability Statement

BGC-Argo: <http://doi.org/10.17882/42182> and <ftp://ftp.ifremer.fr/ifremer/argo/>; Satellite Chla and daily PAR: <https://oceancolor.gsfc.nasa.gov/l3/>.

Acknowledgments

This work is supported by the Marine S&T Fund of Shandong Province for Pilot National Laboratory for Marine Science and Technology (Qingdao) (2018SDKJ0206, Xing), and National Natural Science Foundation of China (41876032, 41890805, Xing).

References

- Bailey, S. W., & Werdell, P. J. (2006). A multi-sensor approach for the on-orbit validation of ocean color satellite data products. *Remote Sensing of Environment*, 102, 12–23. <https://doi.org/10.1016/j.rse.2006.01.015>
- Behrenfeld, M. J., Boss, E., Siegel, D. A., & Shea, D. M. (2005). Carbon-based ocean productivity and phytoplankton physiology from space. *Global Biogeochemical Cycles*, 19, GB1006. <https://doi.org/10.1029/2004gb002299>
- Behrenfeld, M. J., & Falkowski, P. G. (1997). A consumer's guide to phytoplankton primary productivity models. *Limnology & Oceanography*, 42, 1479–1491. <https://doi.org/10.4319/lo.1997.42.7.1479>
- Behrenfeld, M. J., O'Malley, R. T., Boss, E. S., Westberry, T. K., Graff, J. R., Halsey, K. H., et al. (2016). Reevaluating ocean warming impacts on global phytoplankton. *Nature Climate Change*, 6, 323–330. <https://doi.org/10.1038/nclimate2838>
- Behrenfeld, M. J., Westberry, T. K., Boss, E. S., O'Malley, R. T., Siegel, D. A., Wiggert, J. D., et al. (2009). Satellite-detected fluorescence reveals global physiology of ocean phytoplankton. *Biogeosciences*, 6, 779–794. <https://doi.org/10.5194/bg-6-779-2009>
- De Boyer Montégut, C., Madec, G., Fischer, A. S., Lazar, A., & Iudicone, D. (2004). Mixed layer depth over the global ocean: An examination of profile data and a profile-based climatology. *Journal of Geophysical Research*, 109, C12003. <https://doi.org/10.1029/2004jc002378>
- Dierssen, H. M., & Smith, R. C. (2000). Bio-optical properties and remote sensing ocean color algorithms for Antarctic Peninsula waters. *Journal of Geophysical Research*, 105, 26301–26312. <https://doi.org/10.1029/1999jc000296>
- Estapa, M. L., Feen, M. L., & Breves, E. (2019). Direct observations of biological carbon export from profiling floats in the subtropical North Atlantic. *Global Biogeochemical Cycles*, 33, 282–300. <https://doi.org/10.1029/2018gb006098>
- Fox, J., Behrenfeld, M. J., Haëntjens, N., Chase, A., Kramer, S. J., Boss, E., et al. (2020). Phytoplankton growth and productivity in the Western North Atlantic: Observations of regional variability from the NAAMES field campaigns. *Frontiers in Marine Science*, 7, 24. <https://doi.org/10.3389/fmars.2020.00024>
- Frouin, R., Ramon, D., Boss, E., Jolivet, D., Compiègne, M., Tan, J., et al. (2018). Satellite radiation products for ocean biology and biogeochemistry: Needs, state-of-the-art, gaps, development priorities, and opportunities. *Frontiers in Marine Science*, 5, 3. <https://doi.org/10.3389/fmars.2018.00003>
- Gregg, W. W., & Carder, K. L. (1990). A simple spectral solar irradiance model for cloudless maritime atmospheres. *Limnology & Oceanography*, 35(8), 1657–1675. <https://doi.org/10.4319/lo.1990.35.8.1657>
- Haëntjens, N., Boss, E., & Talley, L. D. (2017). Revisiting ocean color algorithms for chlorophyll a and particulate organic carbon in the Southern Ocean using biogeochemical floats. *Journal of Geophysical Research: Oceans*, 122, 6583–6593. <https://doi.org/10.1002/2017jc012844>
- IOCCG. (2020). Synergy between ocean colour and biogeochemical/ecosystem models. In S. Dutkiewicz (Ed.), *International Ocean-Colour Coordinating Group (IOCCG)* (Reports of the International Ocean-Colour Coordinating Group No. 19, p. 184). Dartmouth, Canada. <http://dx.doi.org/10.25607/OBP-711>
- Johnson, R., Strutton, P. G., Wright, S. W., McMinn, A., & Meiners, K. M. (2013). Three improved satellite chlorophyll algorithms for the Southern Ocean. *Journal of Geophysical Research: Oceans*, 118, 3694–3703. <https://doi.org/10.1002/jgrc.20270>
- Kahru, M., & Mitchell, B. G. (2010). Blending of ocean colour algorithms applied to the Southern Ocean. *Remote Sensing Letters*, 1, 119–124. <https://doi.org/10.1080/01431160903547940>
- Lee, Z. (2009). KPAR: An optical property associated with ambiguous values. *Journal of Lake Sciences*, 21, 159–164. <https://doi.org/10.18307/2009.0202>
- Lee, Z., Carder, K. L., & Arnone, R. A. (2002). Deriving inherent optical properties from water color: A multiband quasi-analytical algorithm for optically deep waters. *Applied Optics*, 41(27), 5755–5772. <https://doi.org/10.1364/ao.41.005755>
- Lee, Z., Du, K., Arnone, R., Liew, S., & Penta, B. (2005). Penetration of solar radiation in the upper ocean: A numerical model for oceanic and coastal waters. *Journal of Geophysical Research*, 110, C09019. <https://doi.org/10.1029/2004jc002780>
- Lee, Z., Hu, C., Shang, S., Du, K., Lewis, M., Arnone, R., & Brewin, R. (2013). Penetration of UV-visible solar radiation in the global oceans: Insights from ocean color remote sensing. *Journal of Geophysical Research: Oceans*, 118, 4241–4255. <https://doi.org/10.1002/jgrc.20308>
- Lee, Z., Wei, J., Voss, K., Lewis, M., Bricaud, A., & Huot, Y. (2015). Hyperspectral absorption coefficient of “pure” seawater in the range of 350–550 nm inverted from remote sensing reflectance. *Applied Optics*, 54, 546–558. <https://doi.org/10.1364/ao.54.000546>
- Marrari, M., Hu, C., & Daly, K. (2006). Validation of SeaWiFS chlorophyll a concentrations in the Southern Ocean: A revisit. *Remote Sensing of Environment*, 105(4), 367–375. <https://doi.org/10.1016/j.rse.2006.07.008>

- Mignot, A., Claustre, H., D'Ortenzio, F., Xing, X., Poteau, A., & Ras, J. (2011). From the shape of the vertical profile of in vivo fluorescence to chlorophyll-a concentration. *Biogeosciences*, 8, 2391–2406. <https://doi.org/10.5194/bg-8-2391-2011>
- Mobley, C. D. (1994). *Light and water: Radiative transfer in natural waters*. San Diego, CA: Academic Press. https://www.oceanopticsbook.info/packages/iws_12h/conversion/files/LightandWater.zip
- Mobley, C. D., & Boss, E. S. (2012). Improved irradiances for use in ocean heating, primary production, and photo-oxidation calculations. *Applied Optics*, 51, 6549–6560. <https://doi.org/10.1364/ao.51.006549>
- Morel, A. (1988). Optical modeling of the upper ocean in relation to its biogenous matter content (case I waters). *Journal of Geophysical Research*, 93, 10749–10768. <https://doi.org/10.1029/jc093ic09p10749>
- Morel, A., Huot, Y., Gentili, B., Werdell, P. J., Hooker, S. B., & Franz, B. A. (2007). 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 Sensing of Environment*, 111, 69–88. <https://doi.org/10.1016/j.rse.2007.03.012>
- Morel, A., & Maritorena, S. (2001). Bio-optical properties of oceanic waters: A reappraisal. *Journal of Geophysical Research*, 106(C4), 7163–7180. <https://doi.org/10.1029/2000jc000319>
- Organelli, E., Claustre, H., Bricaud, A., Schmechtig, C., Poteau, A., Xing, X., et al. (2016). A novel near-real-time quality-control procedure for radiometric profiles measured by Bio-Argo floats: Protocols and performances. *Journal of Atmospheric and Oceanic Technology*, 33, 937–951. <https://doi.org/10.1175/jtech-d-15-0193.1>
- Ryther, J. H. (1956). Photosynthesis in the ocean as a function of light intensity. *Limnology & Oceanography*, 1, 61–70. <https://doi.org/10.4319/lo.1956.1.1.0061>
- Uitz, J., Claustre, H., Morel, A., & Hooker, S. B. (2006). Vertical distribution of phytoplankton communities in open ocean: An assessment based on surface chlorophyll. *Journal of Geophysical Research*, 111, C08005. <https://doi.org/10.1029/2005jc003207>
- Wei, J., & Lee, Z. (2013). Model of the attenuation coefficient of daily photosynthetically available radiation in the upper ocean. *Methods in Oceanography*, 8, 56–74. <https://doi.org/10.1016/j.mio.2013.12.001>
- Westberry, T., Behrenfeld, M. J., Siegel, D. A., & Boss, E. (2008). Carbon-based primary productivity modeling with vertically resolved photoacclimation. *Global Biogeochemical Cycles*, 22, GB2024. <https://doi.org/10.1029/2007gb003078>
- Xing, X., Boss, E., Zhang, J., & Chai, F. (2020). Evaluation of ocean color remote sensing algorithms for diffuse attenuation coefficients and optical depths with data collected on BGC-Argo floats. *Remote Sensing*, 12, 2367. <https://doi.org/10.3390/rs12152367>
- Yang, B., Boss, E. S., Haëntjens, N., Long, M. C., Behrenfeld, M. J., Eveleth, R., & Doney, S. C. (2020). Phytoplankton phenology in the North Atlantic: Insights from profiling float measurements. *Frontiers in Marine Science*, 7, 139. <https://doi.org/10.3389/fmars.2020.00139>