Version of Record: https://www.sciencedirect.com/science/article/pii/S0034425723000330
Manuscript_d4f61f734f892f64b60ca8c1daf91104

Retrieving b_{bp} and POC from CALIOP: A Deep Neural Network Approach

Zhenhua Zhang[1,2], Peng Chen[1,2*], Cédric Jamet[3], Davide Dionisi[4], Yongxiang Hu[5], Xiaomei Lu[5], Delu Pan[1, 2]

[1] Southern Marine Science and Engineering Guangdong Laboratory (Guangzhou), No. 1119, Haibin Rd., Nansha District, Guangzhou 511458, China

[2] State Key Laboratory of Satellite Ocean Environment Dynamics, Second Institute of Oceanography, Ministry of Natural Resources, 36 Bochubeilu, Hangzhou 310012, China

[3] Université Littoral Côte d'Opale, CNRS, Université Lille, IRD, UMR 8187, Laboratoire d'Océanologie et de Géosciences (LOG), 62930 Wimereux, France

[4] Institute of Marine Sciences (ISMAR), Italian National Research Council (CNR), Rome - Tor Vergata, Italy

[5] NASA Langley Research Center, Hampton, VA 23681, USA

*chenp@sio.org.cn

ABSTRACT

The CALIOP instrument has demonstrated tremendous potential in terms of oceanic detection. Nighttime measurements from spaceborne lidar increase knowledge of the diurnal variability of ocean optical properties, while measurements in polar regions fill the gaps that passive measurements of ocean color may leave in these areas. However, due to non-ideal instrumental characteristics, errors may occur during retrieval of the ocean optical properties that are derived directly from CALIOP's measured signals. These errors are caused by the transient response (TR) function of the detectors and the polarization crosstalk (CT) between the polarization beam splitters. However, there is no study that integrates and explicitly describes both the required corrections. In addition, the published physics-based methods for retrieving the particulate backscatter coefficient (b_{bp}) from CALIOP use an empirical linear relationship between b_{bp} and the backscatter coefficient at 180° $b(\pi)$ without considering the spatial and temporal variability associated with the conversion

© 2023 published by Elsevier. This manuscript is made available under the Elsevier user license
https://www.elsevier.com/open-access/userlicense/1.0/

coefficient. In this study, after both effects are eliminated through TR deconvolution and CT correction, an innovative deep neural network (DNN) approach is proposed to retrieve b_{bp} and particulate organic carbon (POC) from CALIOP. The DNN is trained with the CALIOP total ocean column-integrated depolarization ratio and subsurface column-integrated perpendicular backscatter data, which are combined with the collocated daytime MODIS b_{bp} and POC products. Once the DNN has been trained with the daytime MODIS product, measurements can be derived from CALIOP in conditions not covered by MODIS data, such as at night or in polar regions. Compared to published physics-based methods, this approach does not require diffuse attenuation coefficients, nor are additional assumptions regarding the conversion coefficient between b_{bp} and the backscatter coefficient at 180° $b(\pi)$ necessary. A comparison of global total depolarization ratios, which are calculated before and after TR deconvolution and CT correction, indicates the necessity and efficacy of correction, and the preliminary results show that the oceanic properties derived from CALIOP via this DNN-based approach agree better with the MODIS products than those derived with the traditional physics-based methods. The use of lidar measurements to derive b_{bp} and POC without relying on passively acquired data is of great interest to the ocean color community.

Keywords: Lidar; Polarization; CALIOP; DNN; b_{bp}; POC

1. Introduction

The launch of the Coastal Zone Color Scanner (CZCS, 1978-1986) marked a new era of ocean color remote sensing (Sullivan et al. 1993). Since then, with its extraordinary global data acquisition capabilities and with observations at daily to interannual scales, ocean color remote sensing has reshaped our understanding of the upper-ocean biogeochemistry in the global oceans as well as in regional basins (Blondeau-Patissier et al. 2014; Brewin et al. 2017; Dickey et al. 2006; Jackson et al.

2017; McClain 2009). However, previous ocean color observations have heavily relied on passive remote sensing techniques, which are limited to the upper part of the water column and cannot provide information regarding the underlying vertical structure (Hostetler et al. 2018; Jamet et al. 2019). Moreover, due to the need for sunlight (Jamet et al. 2019), passive remote sensing can only operate during the daytime, which results in the inability to observe high-latitude ocean areas during lengthy polar nights. Thus, only limited data are available for these regions.

In the past decade, light detection and ranging (lidar), an active remote sensing technique, has regained widespread attention because of its depth-resolving capabilities and lack of dependence on sunlight, which could address the limitations mentioned above. Lidar is a highly reliable tool that can measure several optical and physical parameters of seawater when investigating marine environments. In practice, shipborne and airborne lidar has been widely employed in a variety of oceanographic studies, such as bathymetric mapping (Richter et al. 2017; Saylam et al. 2017; Xie et al. 2021), optical profiling (Chen et al. 2022; Chen and Pan 2019; Chen et al. 2019; Churnside et al. 2018; Churnside et al. 2017; Churnside and Marchbanks 2019; Collister et al. 2018; Lee et al. 2013a; Montes-Hugo et al. 2010; Montes et al. 2011; Roddewig et al. 2020), and plankton layer detection (Chen 2022; Chen et al. 2021; Chen et al. 2020a; Chen et al. 2020b; Churnside and Marchbanks 2015; Churnside et al. 2021; Churnside et al. 2020; Liu et al. 2018; Zhang and Chen 2022; Zhong et al. 2021).

Originally designed to profile the vertical distribution of clouds and aerosols, the Cloud-Aerosol Lidar with Orthogonal Polarization (CALIOP) instrument onboard the Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observation (CALIPSO) platform has shown remarkable potential in oceanic optical property estimation over the last decade (Behrenfeld et al. 2019; Behrenfeld et al. 2017; Churnside et al. 2013; Lu et al. 2014). Compared with passive ocean color sensors, satellite-based lidar is not obstructed by thin cloud coverage and can operate during both the day and night. These capabilities mean that lidar can acquire data in areas where previous optical systems

were less effective, especially in polar regions. Combining the CALIOP depolarization ratios with the Moderate Resolution Imaging Spectroradiometer (MODIS) diffuse attenuation coefficient (K_d, m^{-1}) data and some assumptions supported by empirical evidence, CALIOP ocean backscattered signals can be translated into ocean optical properties, including the total depolarization ratio of ocean water (Dionisi et al. 2020; Lu et al. 2014), particulate backscatter coefficient (b_{bp}, m^{-1}) (Lacour et al. 2020; Lu et al. 2016; Lu et al. 2021b), particulate organic carbon concentration (POC, mg/m^3) (Behrenfeld et al. 2013), and phytoplankton biomass (Behrenfeld et al. 2017).

However, the non-ideal transient response (TR) of 532 nm detectors (Lu et al. 2014; Lu et al. 2020a) and the non-ideal characteristic of polarization beam splitters (PBSs) (Lu et al. 2021a; Pitts et al. 2018) present considerable challenges to the inversion of ocean subsurface optical properties from CALIOP measurements. Previous studies have corrected either the detector (Lu et al. 2020a) or the non-ideal characteristics of PBSs (Lu et al. 2021a), and some have ignored the effects of the non-ideal PBS characteristics (Dionisi et al. 2020; Lacour et al. 2020). Furthermore, some excellent studies (Behrenfeld et al. 2017; Bisson et al. 2021a) did not mention the preprocessing steps used to correct these effects. In general, there is no study that integrates and explicitly describes the required corrections.

To retrieve b_{bp} from CALIOP, the initial methods required the passive remote color product $K_{d,}$(Behrenfeld et al. 2013). Then, the methods were modified to retrieve b_{bp} directly from CALIOP without continuous reliance on passive ocean color observations by building a linear relationship between CALIOP depolarization ratios and collocated MODIS K_d (Behrenfeld et al. 2019; Behrenfeld et al. 2017). However, the specific ratio between the quantities was not provided at that time, and some subsequent studies still used a third-party K_d derived from float measurements (Lacour et al. 2020) to avoid using passive satellite observations. Unfortunately, float measurements are not always sufficient when retrieving ocean properties from CALIOP. A method totally independent of K_d was provided in (Bisson et al. 2021a), which selected a $b(\pi)/b_{bp}$ value of 0.32, and

the uncertainty of these results is substantially affected by the assumed $b(\pi)$ conversion coefficient used to derive b_{bp} (Behrenfeld et al. 2017; Bisson et al. 2021a). This conversion coefficient has its own uncertainty, which limits the accuracy of the physics-based retrieval approach that is derived based on the physical mechanism between the CALIOP return signals and b_{bp} (Berthon et al. 2007; Chami et al. 2014; Hu et al. 2020; Sullivan and Twardowski 2009b; Zhang et al. 2012). In addition, previous studies used the empirical linear relationships between the depolarization ratio and K_d (the depolarization ratio is approximately $2\times K_d$ (Bisson et al. 2021a)). However, the empirical linear relationships do not account for the spatial and temporal disparities of the conversion factors between b_{bp} and $b(\pi)$ or between K_d and the depolarization ratio.

This study intends to establish an innovative deep neural network (DNN) approach to simultaneously derive subsurface b_{bp} and POC from CALIOP data without the inclusion of passive remote sensing data and additional assumptions. The proposed DNN algorithm can address the variability associated with conversion factors between b_{bp} and $b(\pi)$ and between K_d and the depolarization ratio by automatically adjusting the conversion factors across latitude and time. The remainder of the manuscript is organized as follows: first, a systematic correction is applied, and it includes removing the effects of the detector TR and the PBS polarization crosstalk. Then, several filtering steps are used for data quality control, and the physics-based retrieval approach is used for b_{bp} retrieval. In addition, the DNN approach proposed here is used to derive the subsurface b_{bp} and POC simultaneously. After that, we compare the results of the two methods with biogeochemical-Argo data. Finally, both the global b_{bp} and POC values, as well as those in the polar regions, are retrieved from CALIOP data using this DNN algorithm.

2. Data

The data used here include the CALIOP Level-1B V4.1 data products, the Level-2 Cloud, Aerosol, and Merged Layer V4.20 products, and the MODIS Level-3 9 km monthly averaged data

126 acquired by the Aqua satellite. CALIPSO was launched in April 2006 and flew in the "A-Train"
127 constellation following Aqua by 1-2 minutes (Kim et al. 2013), which facilitated simultaneity
128 between their observations prior to September 13, 2018, after which CALIPSO was lowered from an
129 orbit of 705 km to 688 km and became part of the "C-Train" constellation. The C-Train intersects the
130 A-Train ground track approximately every 20 days, providing regular simultaneous and collocated
131 observations between CALIPSO and Aqua. CALIOP, as the main instrument onboard CALIPSO, is
132 a nadir-pointing three-channel elastic backscatter lidar with two wavelengths (532 nm and 1064 nm).
133 The instrument provides information on the attenuated backscatter coefficients and has a receiver
134 footprint diameter of 90 m, a horizontal resolution of 335 m (Winker et al. 2009) and a vertical
135 resolution of 22.5 m in water (Lu et al. 2014). The calibration accuracy of CALIOP Level-1B V4.1
136 data products is significantly improved over previous versions (Getzewich et al. 2018), and these
137 products are used to invert the ocean parameters in this study. The CALIOP Level-2 Cloud, Aerosol,
138 and Merged Layer V4.20 products (Kim et al. 2018) provide the integrated attenuated backscatter
139 (IAB) and aerosol optical depth (AOD) measurements that are used to filter clear sky data. MODIS
140 is a key instrument aboard the Terra and Aqua satellites, which are part of NASA's Earth
141 Observation System (EOS). Terra passes from north to south across the equator in the morning,
142 while Aqua passes south to north over the equator in the afternoon and flies in the "A-Train". The
143 simultaneous observations used in this study are derived from Aqua/MODIS products only. Several
144 algorithms have been developed to retrieve $b_{bp,}$ including the Generalized Inherent Optical Property
145 algorithm (GIOP) (Werdell et al. 2013), Quasi-Analytical Algorithm (QAA) (Lee et al. 2002), and
146 Garver-Siegel-Maritorena algorithm (GSM) (Maritorena et al. 2002). Because the GIOP algorithm
147 outperformed the other inversion algorithms (Bisson et al. 2019), the MODIS GIOP b_{bp}443 L3 9 km
148 products are used here. The MODIS POC products are derived from an empirical relationship
149 between *in situ* POC measurements and the blue-to-green band ratios of remote sensing reflectance
150 (Stramski et al. 2008). Vertical profiles of b_{bp} (700 nm, m^{-1}) were obtained from biogeochemical

Argo data (Claustre et al. 2019), which are available from the Argo Data Assembly Center (ftp://ftp.ifremer.fr/ifremer/argo/dac/ on May 1, 2022) and used to validate the lidar-derived results.

3. Methodology

As shown in Figure 1, the framework for retrieving b_{bp} and POC from CALIOP mainly includes data correction, data filtering, and the inversion of ocean parameters. First, a systematic correction is applied; this correction removes the effects of the detectors' TR function and the PBS polarization crosstalk. Then, the data are filtered to remove the influence of low-lying clouds, ocean foam, white caps and signal saturation. After that, the physics-based retrieval approach is used for b_{bp} and POC retrieval, and a DNN approach is proposed to derive subsurface b_{bp} and POC. The notations and their definitions as well as the dimensional units used in this study are shown in Table 1.

Table 1. Notations and the definitions and dimensional units used in this study

Notations	Definition	Dimensions
z	Water depth	m
$\beta(z)$	Measured attenuated backscatter signals	$sr^{-1}\ km^{-1}$
$\beta'(z)$	True attenuated backscatter signals	$sr^{-1}\ km^{-1}$
$F(z)$	Transient response function of PMT	Dimensionless
$\beta_{\perp,m}(z)$	Measured perpendicular signals	$sr^{-1}\ km^{-1}$
$\beta_{\perp,t}(z)$	True perpendicular signals	$sr^{-1}\ km^{-1}$
$\beta_{\perp,c}(z)$	Corrected perpendicular signals	$sr^{-1}\ km^{-1}$
$\beta_{\parallel,m}(z)$	Measured parallel signals	$sr^{-1}\ km^{-1}$
$\beta_{\parallel,t}(z)$	True parallel signals	$sr^{-1}\ km^{-1}$
$\beta_{\parallel,c}(z)$	Corrected parallel signals	$sr^{-1}\ km^{-1}$
CT	PBS crosstalk	Dimensionless
δ_T	Total ocean column-integrated depolarization ratio	Dimensionless
W	Wind speed	$m\ s^{-1}$
IAB	Integrated attenuated backscatter	sr^{-1}
AOD	Aerosol optical depth	Dimensionless
δ_w	Column-integrated below-surface depolarization ratio	Dimensionless
β_{w+}	Subsurface column-integrated perpendicular backscatter	sr^{-1}
β_s	Surface lidar backscatter	sr^{-1}
ρ	Fresnel reflectance	Dimensionless
θ	Lidar off-nadir angle	rad
σ^2	Variance of the wave slope distribution	Dimensionless

$b(\pi)$	Particulate backscatter coefficient at 180°	$m^{-1}sr^{-1}$
K_d	Water downwelling diffuse attenuation coefficient	m^{-1}
δ_p	Particulate depolarization ratio	Dimensionless
b_{bp}	Particulate backscatter coefficient	m^{-1}
POC	Particulate organic carbon	$mg\ m^{-3}$
t	Ocean surface transmittance	Dimensionless

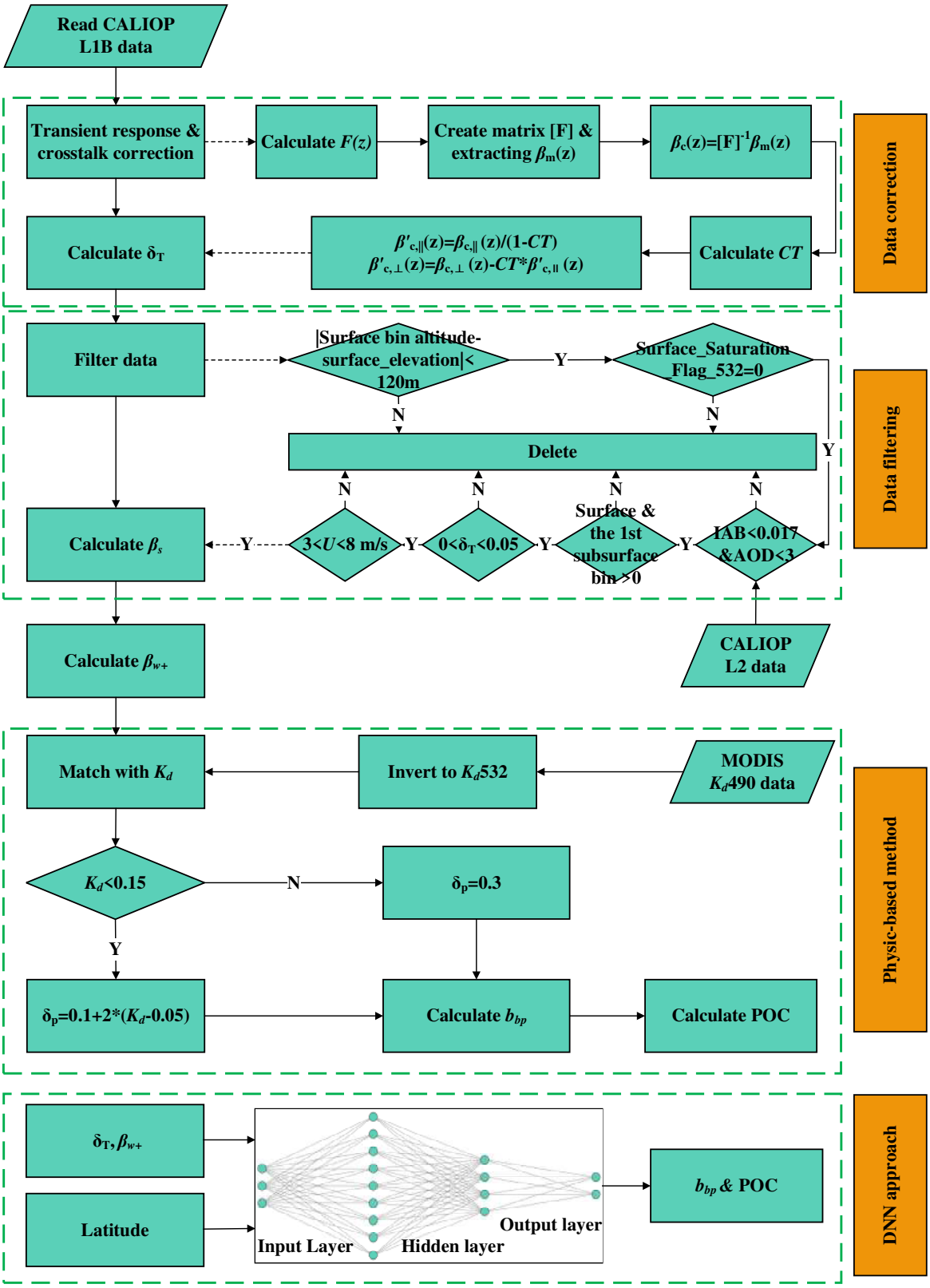


Fig. 1. Framework for retrieving b_{bp} and POC from CALIOP data

3.1 Data correction of instrumental non-ideal characteristics

In the CALIOP receiver subsystem, the backscatter signal at 532 nm is separated into parallel (∥)

and perpendicular (⊥) components by the PBS. Then, these components are detected by photomultiplier tubes (PMTs). However, due to the non-ideal characteristics of these devices, there is a significant difference between the measured signal and the true signal. Thus, it is necessary to correct the measured signal before processing it.

3.1.1 Removal of the effect of the transient response

Due to the transient response (TR) of the detector, the photomultiplier tubes in the CALIOP receiving system cannot immediately return to their preset level after receiving a strong signal. This phenomenon results in both the noise tail effect and the low-pass filter's broadening effect (Lu et al. 2013; McGill et al. 2007), which causes the energy of the peak signal to leak into several adjacent bins. Therefore, the subsequent signal intensity measurements will be greater than the actual backscatter signal. In fact, the output of the receiver β(z) is the convolution of the TR function F(z) and the real backscattered signal β'(z) (Li et al. 2011) and can be expressed as follows:

$$\beta(z) = F(z)*\beta'(z) \tag{1}$$

where F(z) can be calculated from the signal of the surface and adjacent bins as follows (Lu et al. 2014):

$$\mathrm{F}(\mathrm{z_j}) = \frac{\beta(\mathrm{z_j})}{\sum_{i=p-1}^{i=p+10}\beta(z_i)} (j = 1,2,3,\dots,12) \tag{2}$$

where $\beta(z_j)$ is the attenuated backscatter at range bin z_j, i and j are the range bin numbers, and p is the range bin of the peak surface return.

Then, the corrected signal can be calculated by using deconvolution as follows (Lu et al. 2014):

$$\beta'(z) = [\boldsymbol{F}]^{-1}\beta(z) \tag{3}$$

where [**F**] is the matrix form of the transient function as follows:

$$\boldsymbol{F} = \begin{bmatrix} F(z_2) & F(z_1) & 0 & \cdots & 0 \\ F(z_3) & F(z_2) & F(z_1) & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ F(z_{n+1}) & F(z_n) & F(z_{n-1}) & \cdots & F(z_2) \end{bmatrix} \tag{4}$$

3.1.2 Removal of the effect of the crosstalk

PBSs are used to separate backscatter signals at 532 nm into parallel (co-polarization, ∥) and perpendicular (cross-polarization, ⊥) components in CALIOP's receiver subsystems (Hunt et al. 2009; Winker et al. 2009). However, due to the non-ideal characteristics of the PBS, a fraction of the parallel polarization signal is transferred into the perpendicular channel, which produces crosstalk (CT) (Pitts et al. 2018) that can be calculated as follows:

$$\beta_{\perp,m} = \beta_{\perp,t} + CT \times \beta_{\parallel,t} \quad (5)$$

$$\beta_{\parallel,m} = \beta_{\parallel,t} - CT \times \beta_{\parallel,t} \quad (6)$$

where $\beta_{\perp,m}$ and $\beta_{\perp,t}$ are the measured and true perpendicular signals, respectively, and $\beta_{\parallel,m}$ and $\beta_{\parallel,t}$ are the measured and true parallel signals, respectively.

The effect of the crosstalk can be removed as follows (Lu et al. 2021a; Lu et al. 2021b):

$$\beta_{\parallel,c} = \frac{\beta_{\parallel,m}}{1-CT} \quad (7) \qquad \beta_{\perp,c} =$$

where $\beta_{\parallel,c}$ and $\beta_{\perp,c}$ are the corrected parallel and perpendicular signals, respectively.

Figure 2(a) shows an example of the TR-corrected and measured ocean attenuated backscatter signals of the co-polarization and cross-polarization components; the profile was taken on 1 January 2008 in the Atlantic Ocean (51.61°S, 17.998°W). Figure 2(b) shows the depolarization defined as the ratio of the cross-polarization signal to the co-polarization signal, and the results show that the TR-corrected signals after deconvolution are much smaller than the measured signals and that the TR-corrected depolarization is slightly greater than the measured depolarization. The CT-corrected ocean attenuated backscatter signal of the co-polarization and cross-polarization components shown in Figure 2(a) indicates that CT correction has more influence on the cross-polarization channel for the surface bin. As a result, the depolarization of the CT correction is reduced considerably, as shown in Figure 2(b), which is consistent with previous studies (Lu et al. 2021a). Figure 2(c) shows the total corrected signals and the differences between the total measured signals and total corrected signals for each bin.

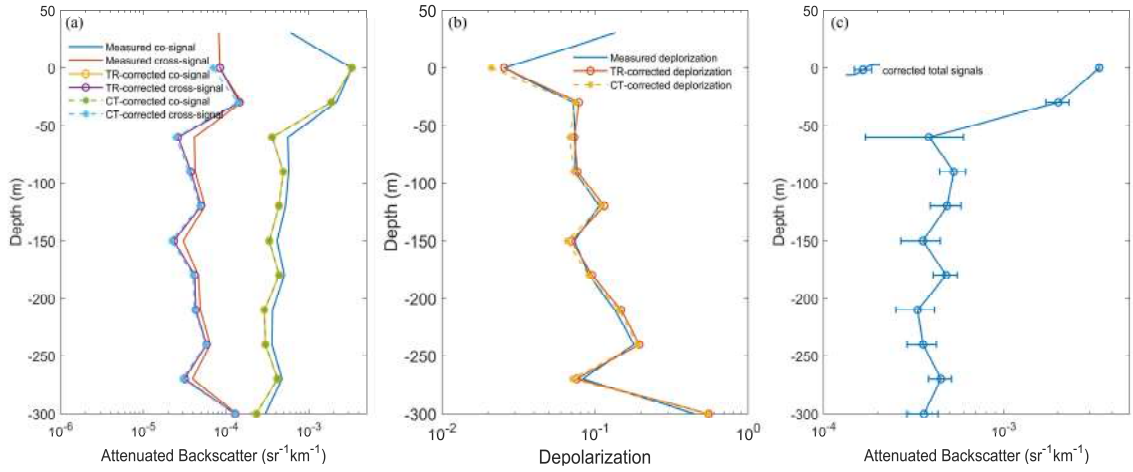


Fig. 2. Example comparison between the measured signals and those after correction. (a) The depth profiles of the measured and corrected ocean attenuated backscatter, (b) the depth profiles of the depolarization calculated from the measured and corrected signals, and (c) the total corrected curve and differences for each bin.

3.2 Data filtering

Considering the influence of low-lying clouds, ice, and large waves on the backscatter signal, the following preliminary screening procedure is used for data quality control (Dionisi et al. 2020):

1. To avoid the peak signals reflected from low-lying atmospheric clouds (Lu et al. 2016), only profiles with surface peaks located within ±120 m of the "Surface Elevation" are used (Dionisi et al. 2020). This threshold allows for small variations in digital elevation model (DEM) accuracy, and most of the dense low-lying atmospheric layers that could be mistaken as surface echoes can be eliminated based on the threshold (Lu et al. 2016).

2. To avoid the influence of signal saturation (Liu et al. 2018), only the unsaturated signals for both $\beta_{\perp}$ and $\beta_{\parallel}$ are considered (i.e., Surface_Saturation_Flag_532Perp = 0 and Surface_Saturation_Flag_532Par = 0).

3. To ensure clear sky conditions (Josset et al. 2011), only data with an IAB less than 0.017 sr^{-1} and AOD less than 3 are used.

4. To ensure the effectiveness of the signal, the value of the data from the peak bin and the first bin below the sea surface should be greater than 0.

5. To avoid effects from ice, the total column-integrated depolarization ratio, including the surface and subsurface backscatter $\delta_{T,}$ should be less than 0.05 (Lu et al. 2014), where $\delta_T = \frac{\sum_{i=p}^{i=p+2} \beta'_{\perp,c}}{\sum_{i=p}^{i=p+2} \beta'_{\parallel,c}}$ (Dionisi et al. 2020).

6. To avoid the effects of ocean foam and white caps during high wind speeds, as well as signal saturation related to low winds (Behrenfeld et al. 2013), a wind speed (W) between 2 m/s and 9 m/s is required. The wind speed is read from CALIOP Level-1B V4.1 data, which are provided by MERRA-2.

Table 2 shows the data profiles corresponding to each filter step for December 2006. The filtered results are shown in Figure 3. After filtering multiple times, the number of data profiles is reduced by 95%.

Table 2. The filtering procedure's steps and the corresponding number of CALIOP profiles

Procedures	Number of profiles
Starting data	19,597,388
Step 1. Surface peaks located within ±120 m of the "Surface Elevation"	8,678,363
Step 2. Surface_Saturation_Flag_532 = 0	5,643,324
Step 3. IAB < 0.017 sr-1 and AOD <3	411,677
Step 4. Peak bin > 0 and first bin after the peak > 0	284,624
Step 5. δ_T< 0.05	272,489
Step 6. 2 m/s <W<9 m/s	230,343

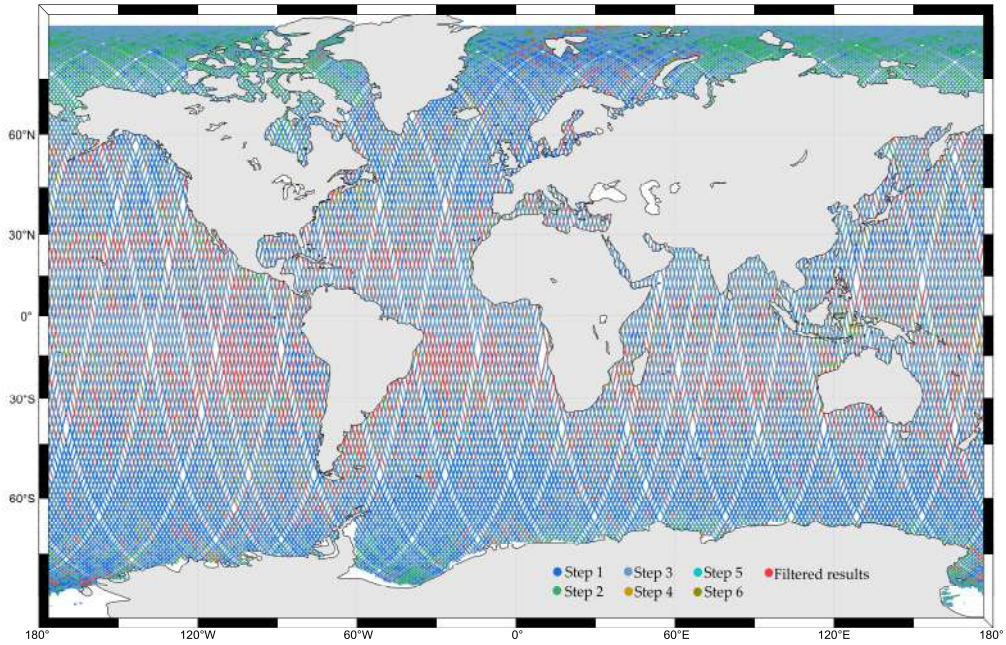


Fig. 3. Profiles filtered by each step and the final results. The red dots represent the remaining profiles; the other colors represent the profiles that were screened out at each step.

Figure 4 shows the number of profiles after each filtering step at different latitudes. The majority of the profiles are located at low latitudes, while the high latitudes have the fewest profiles. The number of points removed in step 1 of the filtering process is greatest in the mid-latitudes, and many of the profiles in the high latitudes are removed in step 2. A large number of profiles across the globe are removed in step 3 to guarantee that only signals under clear sky conditions are selected, and the greatest number of profiles removed in step 4 are located in the low latitudes. Step 5 filters out some high-latitude profiles due to ice coverage, while many high-latitude profiles are removed in step 6 due to the effects of ocean foam and white caps caused by high wind speeds.

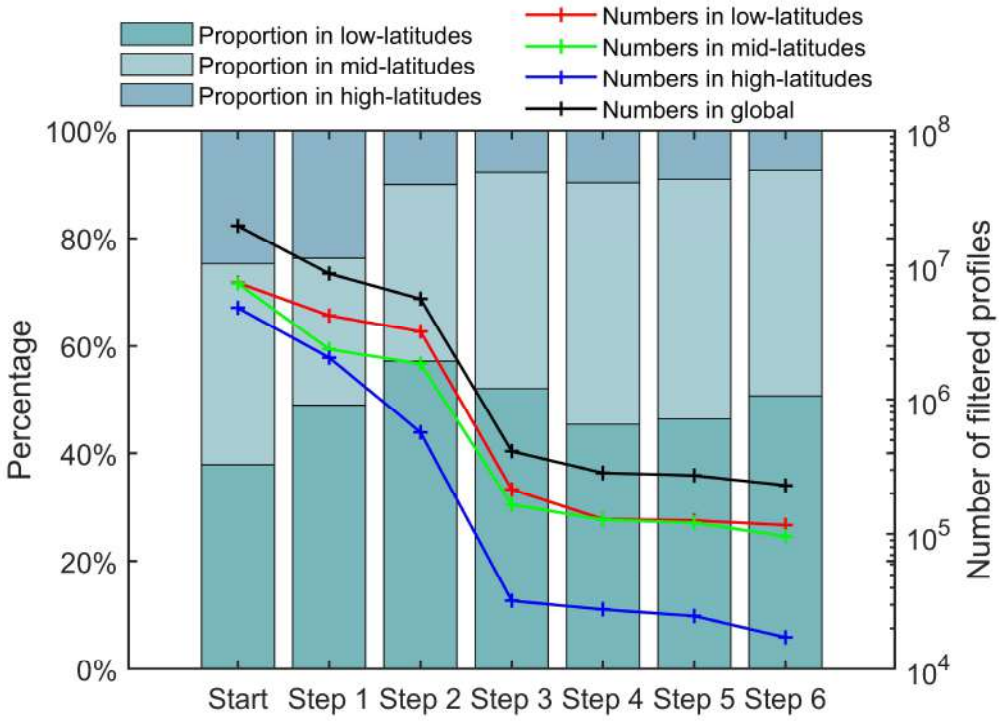


Fig. 4. Number of profiles filtered by each step and the final results. Stacked bars represent the proportion of the filtered profiles after each step in low-, mid- and high-latitudes, and lines represent the changes in the number of profiles.

3.3 Physics-based retrieval method

The physics-based retrieval method is used to estimate the ocean optical properties from CALIOP by following the methodology described in (Behrenfeld et al. 2013) and (Bisson et al. 2021a), and this process has been applied to the production of global data in (Behrenfeld et al. 2019; Behrenfeld et al. 2017) as well. CALIOP receives signals on parallel and perpendicular channels, and it is difficult to extract subsurface signals from parallel channels because of the interference caused by sea surface reflection. However, the perpendicular signal is mainly caused by the depolarization effect of particulates in water, and the influence of the Fresnel reflectance on the air-sea interface and the scattering of light by water molecules is negligible. Therefore, the perpendicular signal is mainly used to characterize particulates. First, considering the influence of the atmosphere, the perpendicular-parallel ratio is used to correct the signal (Behrenfeld et al. 2013)

as follows:

$$\beta_{w+} = \delta_T \frac{\beta_s}{1-\delta_T/\delta_w} \quad (9)$$

where β_{w+} is the subsurface column-integrated backscatter of the perpendicular components, and δ_w=0.1 is the column-integrated below-surface depolarization ratio, which is not retrievable from CALIOP data because of its coarse vertical resolution (22.5 m in water) (Behrenfeld et al. 2013). β_s is the lidar surface backscatter, which is expressed as follows (Hu et al. 2008):

$$\beta_s = \frac{\rho}{4\pi\sigma^2 \cos^4\theta} \exp(-\frac{\tan^2\theta}{2\sigma^2}) \quad (10)$$

where ρ is the Fresnel reflectance, and ρ=0.0209 for seawater at 532 nm (Hu et al. 2008). θ is the off-nadir angle of the lidar, and θ=3.0 degrees since November 28, 2007. σ^2 is the variance of the wave slope distribution, which is a function of the wind speed and is expressed as follows:

$$\begin{aligned} \sigma^2 &= 0.0146\sqrt{W}, \ (W < 7 \ m/s) \\ \sigma^2 &= 0.003 + 0.00512W, \ (7 \ m/s \le W < 13.3 \ m/s) \\ \sigma^2 &= 0.138\, log_{10} W - 0.084, \ (13.3 \ m/s \le W) \end{aligned} \quad (11)$$

Then, the particulate backscatter coefficient at the 180° scattering angle is calculated as follows:

$$b(\pi) = 2K_d \beta_{w+} \frac{1+\delta_p}{\delta_p t^2} \quad (12)$$

where t is the ocean surface transmittance, and t≈0.98 at 532 nm. K_d is the ocean downwelling diffuse attenuation coefficient, and δ_p is the particulate depolarization ratio, which is approximately $2K_d$ (Bisson et al. 2021a). As $1 + \delta_p$ is very close to 1, Equation 12 can be rewritten as follows:

$$b(\pi) = \frac{\beta_{w+}}{0.98^2} \quad (13)$$

Based on the relationship between the backscatter coefficient at 180° b(π) and b_{bp} (Sullivan and Twardowski 2009a) and the particulate backscattering ratio at different wavelengths (Bisson et al. 2021a; Boss et al. 2004; Soja‐Woniak et al. 2019), $b_{bp}443$ can be calculated as follows:

$$b_{\mathrm{bp}}(443) = b_{\mathrm{bp}}(532)\frac{532}{443} \approx \frac{b(\pi)}{0.32}\frac{532}{443} \quad (14)$$

where the final factor is used to transform b_{bp} at 532 nm to b_{bp} at 443 nm based on an assumption regarding the spectral shape of b_{bp} for comparison with MODIS products. Although previous CALIOP data (Behrenfeld et al. 2019) used a processing value of 0.16 for the ratio of $b(\pi)$ to b_{bp}, 0.32 was chosen by considering more recent work (Lacour et al. 2020; Lu et al. 2020a; Lu et al. 2021b).

Then, the POC can be calculated based on the linear relationship with b_{bp} and is provided by (Stramski et al. 2008) as follows:

$$\mathrm{POC} = 70850.7 * b_{bp}(555) - 9.088 \quad (15)$$

Note that the ratio between $b(\pi)$ and b_{bp} exhibits variability. Some studies report that the estimated ratio between these quantities is approximately 1.43 (Sullivan and Twardowski 2009c; Zhang et al. 2014), some report that it is 1.06 (Churnside and Marchbanks 2015; Churnside et al. 2014; Lee et al. 2013b), while others report a value of 0.5 (Boss and Pegau 2001; Chami et al. 2006; Whitmire et al. 2010), which may introduce uncertainties into the physics-based method. Compared with the physics-based method, the data-based approach does not require a diffuse attenuation coefficient or additional assumptions about the relationship between b_{bp} and $b(\pi)$ and thus could constitute an effective alternative method.

3.4 Deep neural network approach

The nonlinear DNN algorithm automatically adjusts the column-integrated below-surface depolarization ratios β_{w+} that minimize b_{bp} and POC retrieval errors without predetermined knowledge. The DNN algorithm does not require an estimation of the exact relationship between $b_{bp,}$ POC and β_{w+} as long as their relationship exists. As shown in Figure 5, the multilayer perceptron, which employs a backpropagation network (MLP BPN), is used for the DNN model, which includes an input layer, 10 hidden layers, and an output layer. The selection of the number of hidden layers and nodes is based on the experience of several tests with the goal of increasing accuracy. The input

layer has 3 nodes that correspond to the CALIOP input variables as follows:

1. The total column-integrated depolarization ratio, including the surface and subsurface backscatter δ_T;

2. The subsurface column-integrated backscatter of the perpendicular components β_{w+};

3. The latitude of the data.

These variables are selected because b_{bp} and POC are derived from δ_T (#1) and β_w (#2), and the algorithm could adjust their relationship in terms of latitude (#3). The output layer has two nodes that correspond to the output variables and are collocated with MODIS products, including b_{bp} and POC. CALIOP variables are averaged 9 km along-track to match the MODIS data. Between the input layer and the hidden layers, a sigmoid function is used as the activation function for each layer to separate the layers and manage the complicated nonlinear functions (Sharma et al. 2020). The sigmoid function transforms the values in the range from 0 to 1. The output layer uses the linear function to generate the final variables, and the DNN model uses a supervised learning algorithm to search for the optimal weight values to solve complex nonlinear problems (Madhiarasan and Deepa 2017). Dropout is used in the hidden layers to avoid overfitting (Srivastava et al. 2014). The experimental platform uses an Intel Core i9-12900K, 3.19 GHz processor, has 64 GB of memory, and has an NVIDIA GeForce GTX 3080 graphics card. In terms of the software environment, the Windows 11 Professional 64-bit operating system was used. The proposed method was programmed using Python 3.9.12, and the deep learning framework is PyTorch 1.11.0. The GPU computing platform is CUDA 11.3 and cuDNN 8.4. The DNN algorithm was trained by using the Adam (Kingma and Ba 2014) stochastic optimizer with an initial learning rate of 10^{-3}. Chlorophyll concentration (Chl) products could also be produced from CALIOP data with the DNN algorithm in the same way that the b_{bp} and POC data are produced except that MODIS Chl products were used as training reference data.

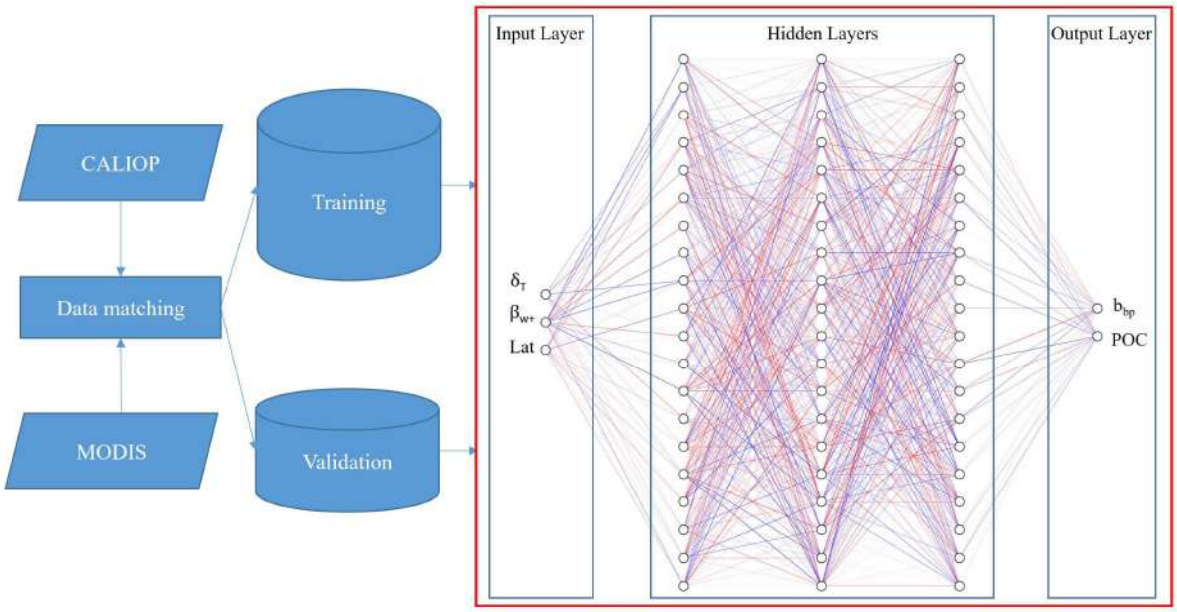


Fig. 5. Flowchart of the deep neural network including its architecture. The input layer has 3 nodes that correspond to the variables, and the output layer has two nodes. The 10 hidden layers have 100 nodes each.

The CALIPSO satellite flew in the "A-Train" constellation following Aqua for 1-2 min (Kim et al. 2013), which allowed for simultaneity between their observations and conveniently permitted subsequent data matching and data training processes. In this study, CALIOP lidar backscatter measurements from 2008 and collocated b_{bp} and POC products from Aqua/MODIS are used for the training model. There are 2,172,528 matched points, and the dataset was randomly divided into 80% for training and 20% for validation. As shown in Figure 6, the low latitudes account for more than half of the matched data, with the available low latitude data peaking in May. The amount of high-latitude data varies sharply from month to month due to changes in sea ice coverage and peaks in August. The DNN algorithm is applied to b_{bp} and POC inversion between 2008 and 2021, and the depolarization ratios of CALIOP's nighttime measurements are increased by 7% due to the polarization gain ratio error in CALIOP's nighttime results (Hu et al. 2021).

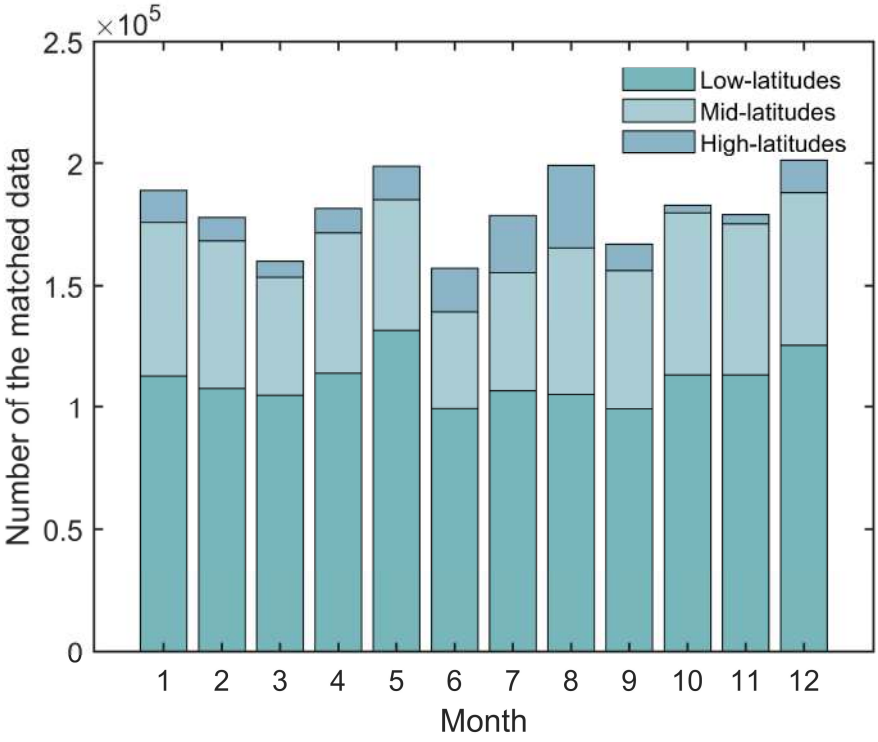


Fig. 6. Number of matched data in each month used for model training in 2018.

3.5 Uncertainty of DNN products

In this study, we identify three sources of errors that contribute to the total uncertainty for DNN products, namely, the uncertainty stemming from the measurement (M), representation (R), and prediction (P) errors. The total uncertainty (E) can thus be expressed as the root of the squared sum of the uncertainties from the three error sources: $E = \sqrt{M^2 + R^2 + P^2}$ (Gregor and Gruber 2021).

The measurement error, M, indicates the combination of potential biases from the measurement of MODIS and CALIOP data. Relying on MODIS data for training, the same bias as could affect MODIS data will be propagated by the DNN approach in the resultant CALIOP products. Therefore, the uncertainty of DNN products includes the uncertainties in the variables, which include the subsurface column-integrated backscatter and corrected total depolarization ratios, and the reference data, i.e., the uncertainty of MODIS products. Summing the uncertainties of the input variables and reference data provides an estimation of the upper bound of the model's uncertainty. First, the uncertainty of the subsurface column-integrated backscatter includes the uncertainties in the

theoretical ocean surface backscatter estimated from MERRA-2 wind speed and the uncertainties in the corrected total depolarization ratios (Lu et al. 2014):

$$\frac{<\Delta^2\beta_{w+}>}{<\beta_{w+}^2>} = \frac{<\Delta^2\beta_s>}{<\beta_s^2>} + \frac{1}{<(1-10\delta_T)^2>}\frac{<\Delta^2\delta_T>}{<\delta_T^2>} \quad (16)$$

As the root mean square error (RMSE) of the MERRA-2 wind speed is 1.68 m/s (Carvalho 2019), the uncertainty of the ocean surface backscatter is assumed to be 10%. The uncertainty of F(z) is less than 1% (Lu et al. 2014), the uncertainty of CT is less than 10% (Lu et al. 2021a), and the uncertainty of the corrected total depolarization ratios due to the uncertainty of F(z) and CT is approximately 2%. As a result, the uncertainty of the subsurface column-integrated backscatter of the perpendicular component is approximately 14%.

Then, the M of the DNN products can be calculated as follows:

$$\frac{<\Delta^2 bbp_{DNN}>}{<bbp_{DNN}^2>} = \frac{<\Delta^2\beta_{w+}>}{<\beta_{w+}^2>} + \frac{<\Delta^2\delta_T>}{<\delta_T^2>} + \frac{<\Delta^2 bbp_{MODIS}>}{<bbp_{MODIS}^2>} \quad (17)$$

The uncertainty of the MODIS b_{bp} is approximately 30% (Bisson et al. 2019), and it is one of the largest contributors to the uncertainty of the CALIOP b_{bp}. Above all, the upper M bound of the CALIOP b_{bp} produced by the DNN is approximately 34.6%. The representation error is a result of the fact that we developed the DNN model on a grid that is in many places coarser in time and space than the typical scales of variability of b_{bp}. The representation error is commonly assumed to be normally distributed with a bias of 0 on a global average (Gregor and Gruber 2021). The prediction error P, which is approximately 16%, is determined by evaluating the statistical model through the DNN model's validation data. Overall, the total uncertainty E is 38.1% in this study.

There is another recently-proposed uncertainty evaluation method based on a random statistical analysis approach (Chau et al. 2022), which calculates the standard deviation and mean value with an ensemble of 100 results by running DNN models 100 times. This method demonstrates a better model uncertainty in this study with a value of approximately 15%. However, this random statistical analysis method does not consider the uncertainties of the input variables and reference data.

4. Results and discussion

4.1 Observed and corrected results

Figure 7 shows the global observed and corrected total depolarization after removing the effects of the detector TR and the PBS polarization crosstalk. Although they have a similar spatial distribution pattern, the lowest values occur over oligotrophic waters (i.e., waters that are nutrient deficient and contain little biomass) between 40° N and 40° S, with the exception of coastal regions and the eastern Pacific upwelling regions, which have higher total depolarization. Figure 7(a) shows that the results prior to correction are much higher than those after the correction was applied, as shown Figure 7(b). In the oligotrophic areas, the depolarization before correction was approximately 0.01 but was reduced to 0.004 after correction, which is consistent with previous studies (Lu et al. 2021a).

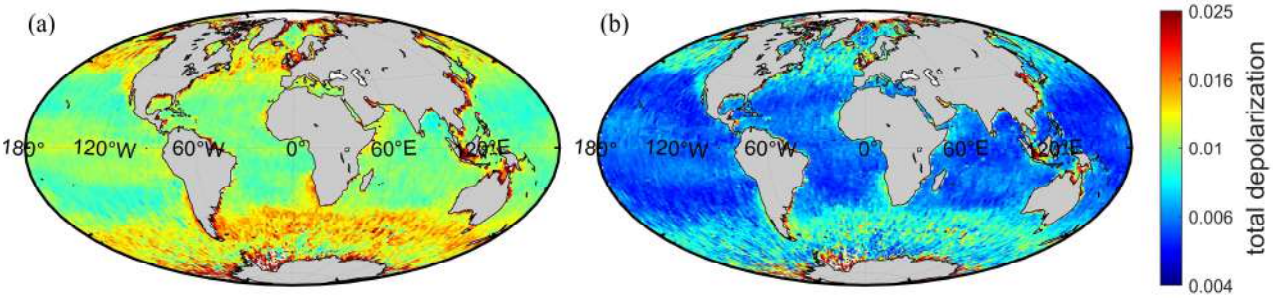


Fig. 7. Global distribution of observed (a) and corrected (b) total depolarization.

Figure 8 (a) shows the frequency histogram of the depolarization ratios both before and after correction, where the frequency values were normalized by the relative probability. The corrected depolarization ratio is much smaller than the uncorrected value. Before correction, the maximum frequency value was approximately 0.01; after correction, the maximum frequency value was reduced to 0.002. Detailed statistics pertaining to the total depolarization ratios are shown in the boxplot (Figure 8(b)), including the minimum, 25th percentile, median, 75th percentile, and maximum values once outliers had been removed. In addition, the average value is represented by

the green asterisk. Before correction, the maximum value (represented by the upper black bar) was 0.032, the 75th percentile (represented by the top edge of the blue box) was 0.016, the median value (marked by the central red line) was 0.01, the 25th percentile (as indicated by the bottom edge of the box) was 0.006, and the average value was 0.012. After the correction was applied, these values became 0.024, 0.011, 0.006, 0.003, and 0.008, respectively. These results indicate a 30% reduction in depolarization after the correction was applied.

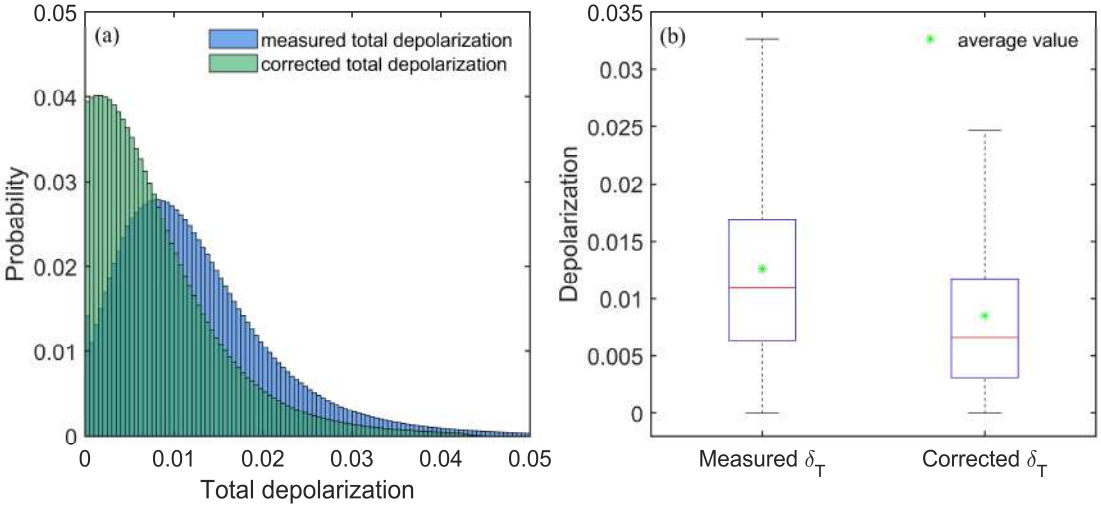


Figure 8. Data statistics of the measured and corrected total depolarization. (a) Frequency histogram, and (b) box plot of the observed and corrected total depolarization

The scatter density diagram shown in Figure 9 indicates that the corrected signal has a greater correlation with the ocean parameters. For consistency with previous studies (Lu et al. 2014), the results are displayed logarithmically, and the correlations are calculated using logarithms. The determination coefficient (R^2) between the CALIOP total depolarization and MODIS b_{bp} is improved from 0.53 to 0.70 after correction, and the determination coefficient between the CALIOP total depolarization and MODIS POC is improved from 0.44 to 0.52 after correction. In addition, the point density, as represented by the color scale, shows that the corrected results are clustered around the best-fit line shown in red. These results demonstrate the necessity of correcting the CALIOP data.

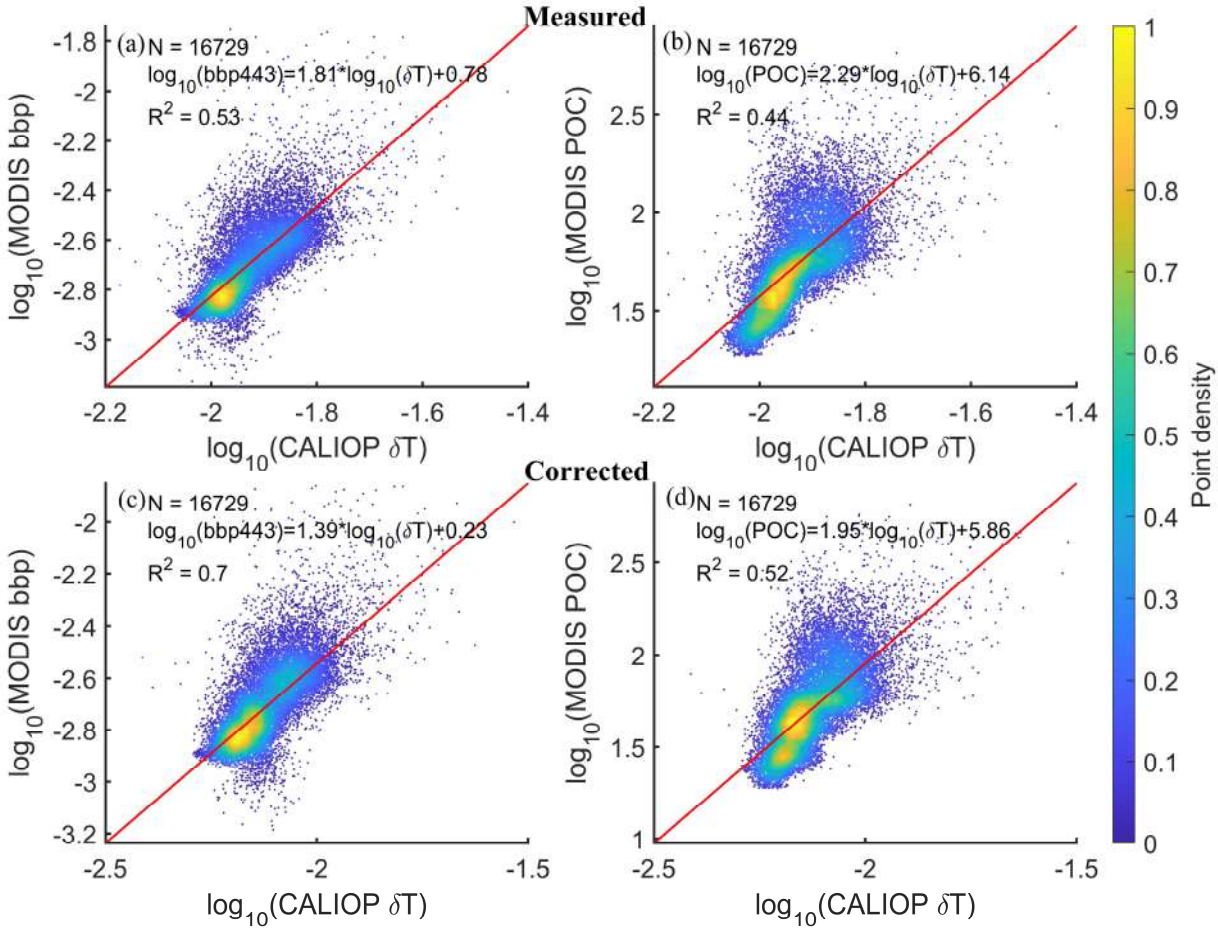
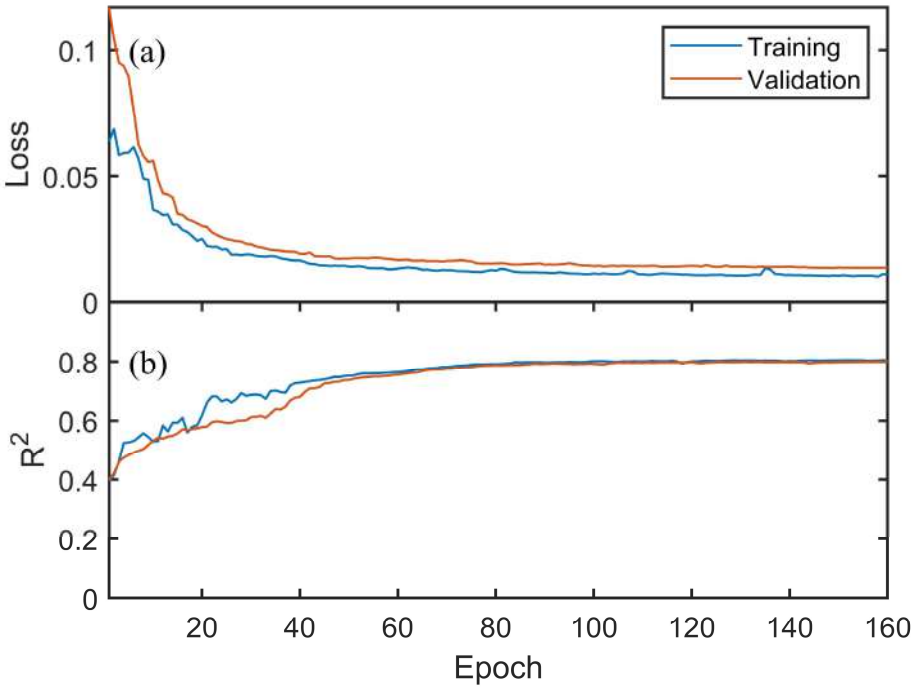


Fig. 9. Log-log density plots of the MODIS b_{bp} and POC as a function of the CALIOP integrated total depolarization. (a) and (b) are the measured results, and (c) and (d) are the corrected results. The red line shows the linear fit, and the color scale represents the density of the occurrences.

4.2 DNN training results

The decrease in losses, which are defined as the mean squared differences between the reference values and the predicted outcome produced by DNN, and the increase in R^2 presented in Figure 10 show the purpose of model training. An early stopping callback was applied provided that the losses were no longer reduced. The losses and R^2 of the training data show the powerful fitting ability of the model, and the generalization capability of the model is indicated by the losses and R^2 of the validation data. Figure 11 shows that the spatial distributions of the annual mean CALIOP b_{bp} and POC agree with the b_{bp} and POC provided by MODIS. CALIOP overestimates b_{bp} in the open sea by as much as 0.001 m^{-1} (red in Figure 11(e)) but underestimates b_{bp} and POC in some coastal areas and

453 at high latitudes.



454

455 Fig. 10. Loss (a) and R^2 (b) versus epoch during model training.

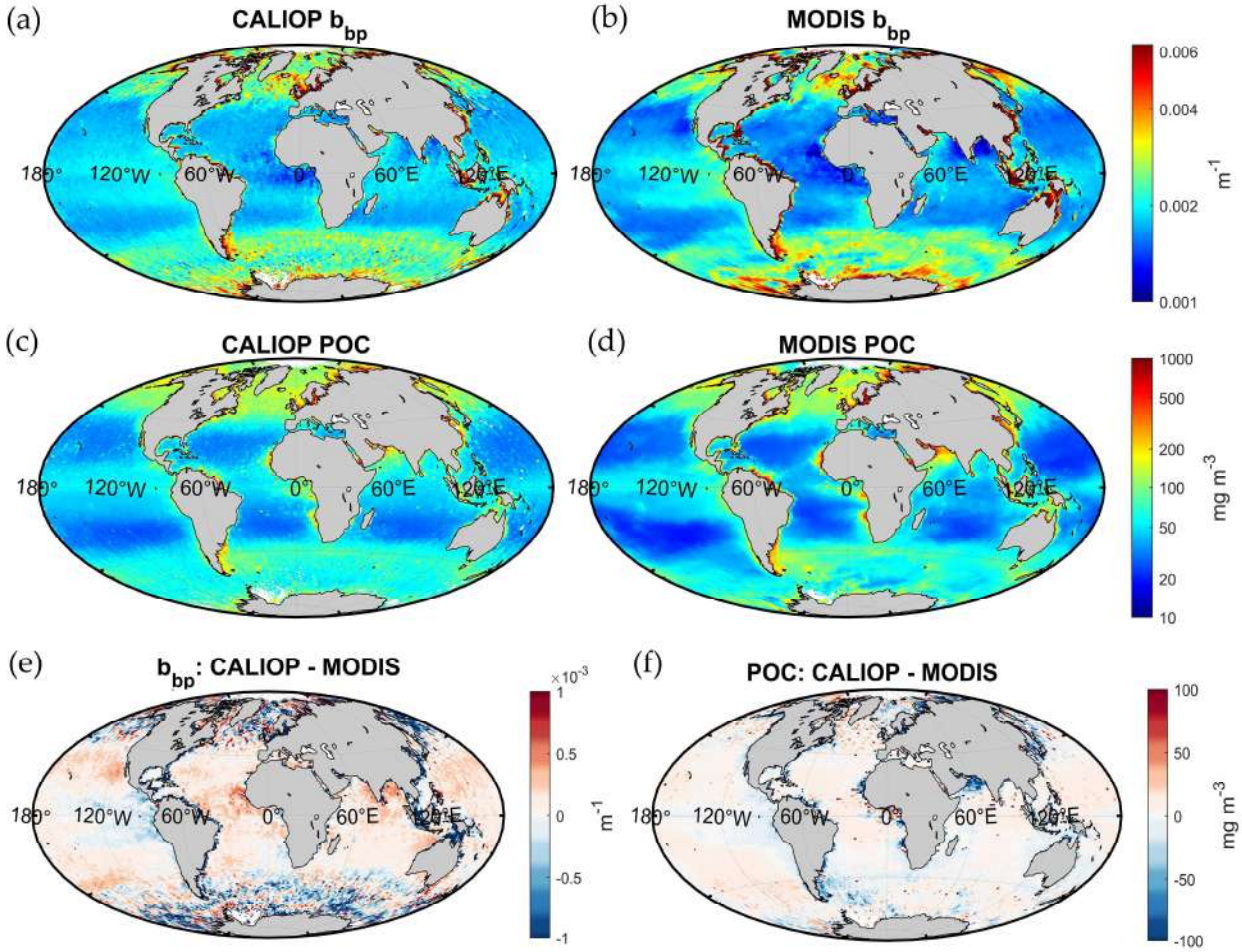


Fig. 11 Comparisons of the annual mean 2008 CALIOP b_{bp} and POC with the MODIS b_{bp} and POC.

4.3 Comparison between the DNN and the physics-based retrieval method

The b_{bp} probability difference between MODIS and CALIOP shown in Figure 12(a) indicates that the DNN retrieval results are unbiased, i.e., the difference is symmetrically distributed around the average b_{bp} difference, which is approximately zero. However, the b_{bp} difference with the physics-based retrieval method, shown in Figure 12(c), is more biased. Additionally, the difference of the DNN method is more concentrated around zero and has a sharper peak, as shown in Figure 12(a). For the DNN method, the normalized frequency of the central bin from -0.0001 m^{-1} to 0 is 13.8%, and the b_{bp} differences from -0.0001 m^{-1} to 0.0001 m^{-1} account for more than 26% of the sample populations. The corresponding values for the physics-based method are 6.7% and 13.1%. The root mean square difference (RMSD) of the DNN method is 0.0009 m^{-1}, which is less than that

of the physics-based retrieval method, which is 0.0022 m^{-1}. The mean absolute difference (MAD) of the DNN method is 0.0001 m^{-1}, while the corresponding value of the physics-based method is 0.001 m^{-1}. In addition, the determination coefficient of the b_{bp} values as determined by MODIS and CALIOP improves remarkably when using the DNN method as opposed to the physics-based retrieval method. This result occurs largely because the DNN method does not need a priori assumptions regarding the relationships between δ_p and K_d and between $b(\pi)$ and b_{bp}. It is also true that the DNN is trained to reproduce MODIS data. These results indicate that the DNN method is effective for b_{bp} inversion.

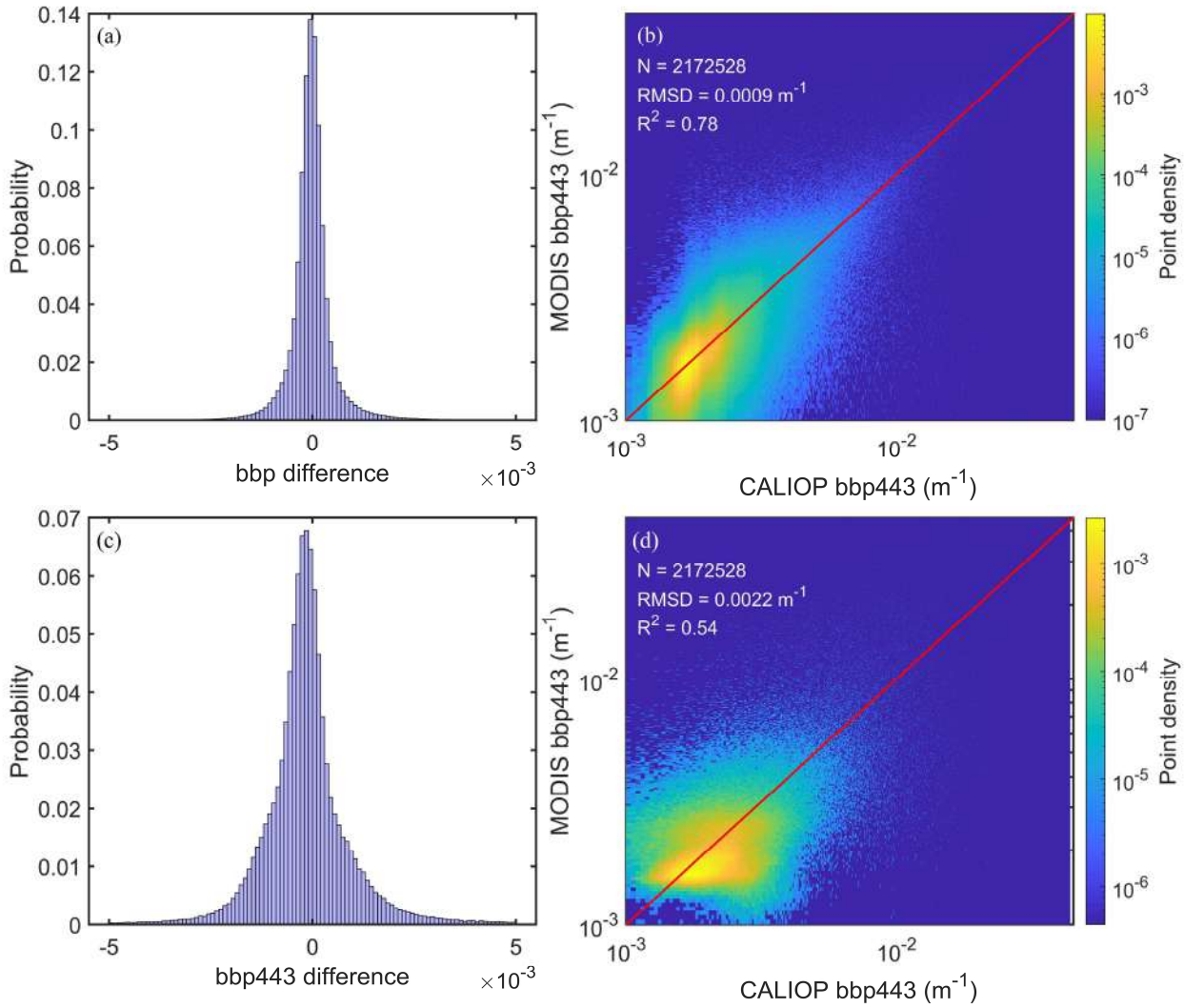


Fig. 12. Statistics of $b_{bp}443$ results derived from CALIOP data. (a) Frequency histogram of the $b_{bp}443$ difference (MODIS $b_{bp}443$ – CALIOP $b_{bp}443$) when using the DNN algorithm, (b) determination coefficient between the MODIS $b_{bp}443$ and CALIOP $b_{bp}443$ when using the DNN algorithm, (c) frequency histogram of the $b_{bp}443$ difference (MODIS $b_{bp}443$ – CALIOP $b_{bp}443$) when using the physics-based method, and (d) determination coefficient between the MODIS $b_{bp}443$ and CALIOP $b_{bp}443$ when using the physics-based method.

Similarly, POC retrieval using the DNN method produces an unbiased result, as shown in Figure 13(a). The normalized frequency of the central bin from -3 mg/m^3 to 0 is 15.3%, and the POC differences from -3 mg/m^3 to +3 mg/m^3 account for more than 26.4% of the sample populations. Relatively, the corresponding values of the alternative POC are 2.9% and 5.4%, where the alternative POC is the result of b_{bp} from the physics-based approach coupled with Eq. 15. Alternative POC values tend to be greater than the MODIS results, and the peak probability difference is located in the -42 mg/m^3 to -39 mg/m^3 bin, as shown in Figure 13(c). Using the DNN method, the RMSD was reduced from 109.8 mg/m^3 of the alternative POC to 81.9 mg/m^3, and the MAD was reduced to 25.5 mg/m^3 from 51.3 mg/m^3. In addition, the determination coefficient improved from 0.44 to 0.80 compared with the alternative POC. Although there is a better agreement when the MODIS POC products are used by the DNN model, it should be noted that there is no physics-based method for POC inversion derived from CALIOP data directly, and no public CALIOP POC data are currently available. The alternative POC is a combination of the physics-based b_{bp} and Equation (15), which converts b_{bp} to POC with an empirical relationship that has been used in previous studies (Behrenfeld et al. 2013). The huge discrepancies between the MODIS POC and alternative POC are in large part due to an inconsistency between POC values computed with the Rrs band ratio and POC values computed from b_{bp} with Equation (15). Previous studies have shown that the Rrs band ratio algorithm used for estimating POC performs better across a broad range of *in situ* datasets (Evers-King et al. 2017). Overall, DNN algorithms can provide POC products derived from CALIOP, which could complement the lack of MODIS POC products in polar regions and at nighttime.

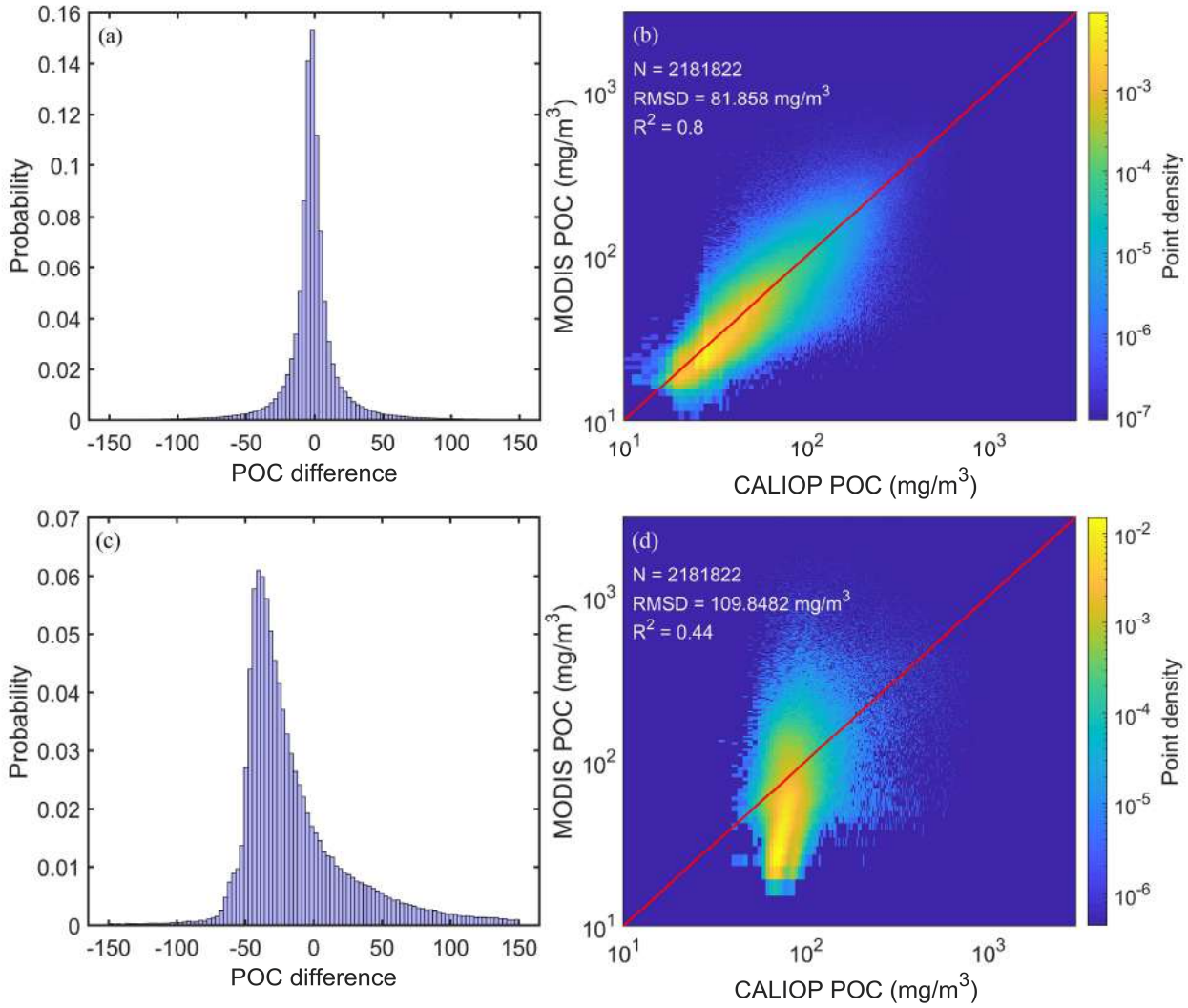


Fig. 13. Statistics of the POC results derived from CALIOP data. (a) Frequency histogram of the POC difference (MODIS POC – CALIOP POC) when using the DNN algorithm, (b) determination coefficient between the MODIS POC and CALIOP POC when using the DNN algorithm, (c) frequency histogram of the POC difference (MODIS POC – CALIOP POC) when using the physics-based b_{bp} and Eq. (15), and (d) determination coefficient between the MODIS POC and CALIOP POC when using the physics-based b_{bp} and Eq. (15).

4.4 Different along-track averages for the DNN approach

Since the DNN algorithm is highly nonlinear, the results may be influenced by noisy data. The data's signal-to-noise ratio could be affected by along-track averaging. Here, we use different along-track averaging to train the DNN algorithm. The MODIS Level-3 4 km monthly averaged data are used here, and the CALIOP variables are averaged 4 km along-track to match the MODIS data. The results are shown in Figure 14. Compared with the 9 km along-track averaged results shown in

Figure 12, the b_{bp} RMSD is slightly reduced to 0.0008 m^{-1} from 0.0009 m^{-1}, but there is no difference between the determination coefficients. The POC RMSD changed to 88.4 mg/m^3 from 81.8 $mg/m^{3,}$ while the determination coefficient was not reduced. The results show that there is no significant difference between different along-track averages, which indicates that the various along-track averages have little influence on the DNN algorithms. This result demonstrates the robustness of the DNN approach.

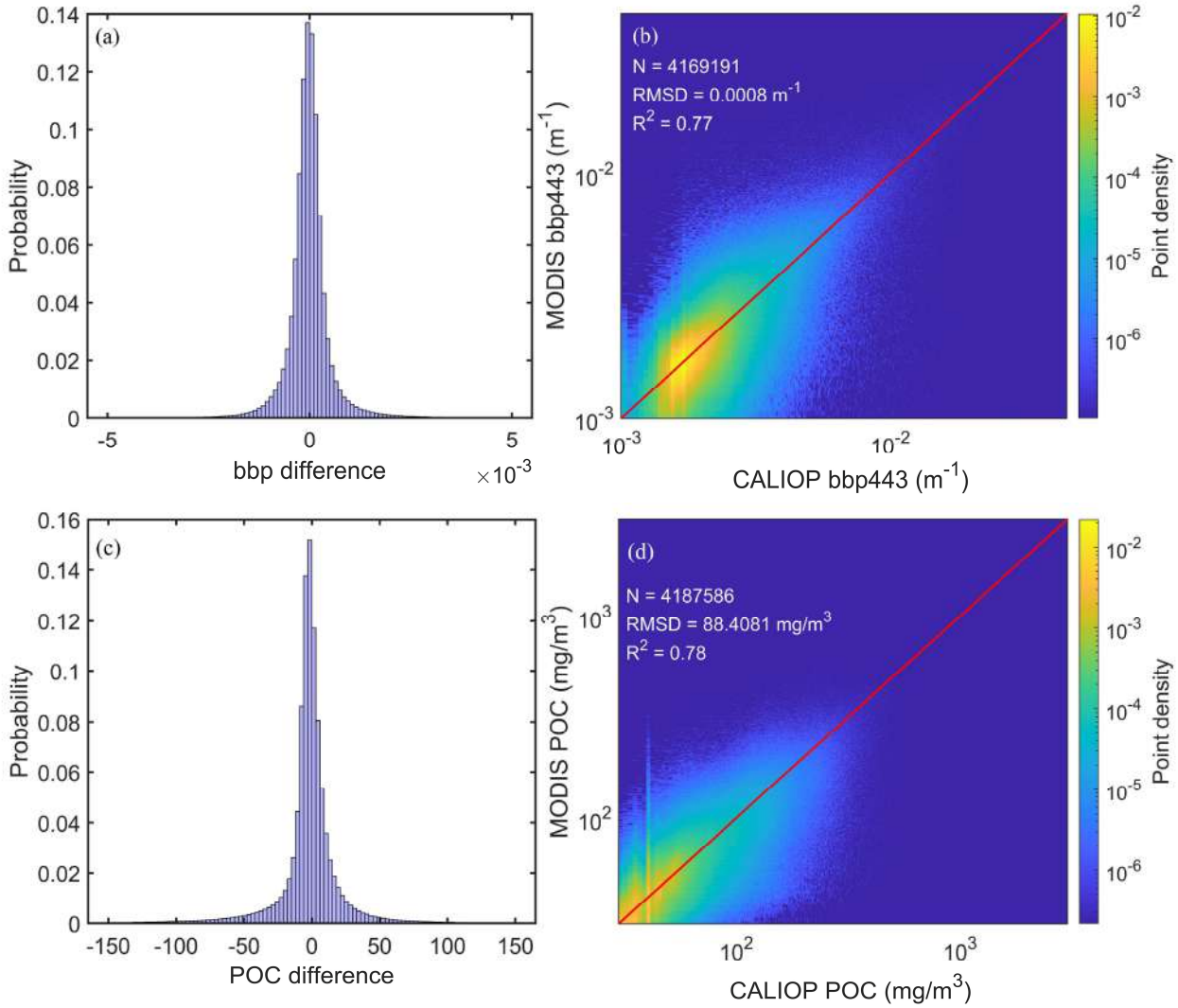


Fig. 14. Statistics of the b_{bp} and POC results determined by the DNN method trained with 4 km along-track averaging. (a) Frequency histogram of the b_{bp} difference (MODIS b_{bp} – CALIOP b_{bp}), (b) determination coefficient between the MODIS b_{bp} and CALIOP b_{bp}, (c) frequency histogram of the POC difference (MODIS POC – CALIOP POC), and (d) determination coefficient between the MODIS POC and CALIOP POC.

4.5 Comparison with previous products and validation using biogeochemical Argo data

To validate the performance of the DNN algorithm and allow comparison with existing products, the Oregon State University (OSU) b_{bp} data produced by the physics-based method described in Section 3.3, which can be downloaded from http://orca.science.oregonstate.edu/lidar_public_v2.php (accessed on 15 June 2022), are used here. The related approach is coherent with that introduced in Section 3.3, except for the value of $b(\pi)/ b_{bp}$ (0.16 for the OSU data). The annual mean 2008 CALIOP b_{bp} produced by OSU (Behrenfeld et al. 2019) is shown in Figure 15(a). Compared with the annual 2008 MODIS b_{bp} shown in Figure 11(b), their spatial distributions are slightly different, especially on the west coast of Africa and the Indian Ocean region, where the MODIS b_{bp} is much lower. This phenomenon is apparent in terms of the difference between them, which is shown in Figure 15(b). In contrast, the difference between the MODIS b_{bp} and the CALIOP b_{bp} produced by the DNN algorithms is much less, as shown in Figure 11.

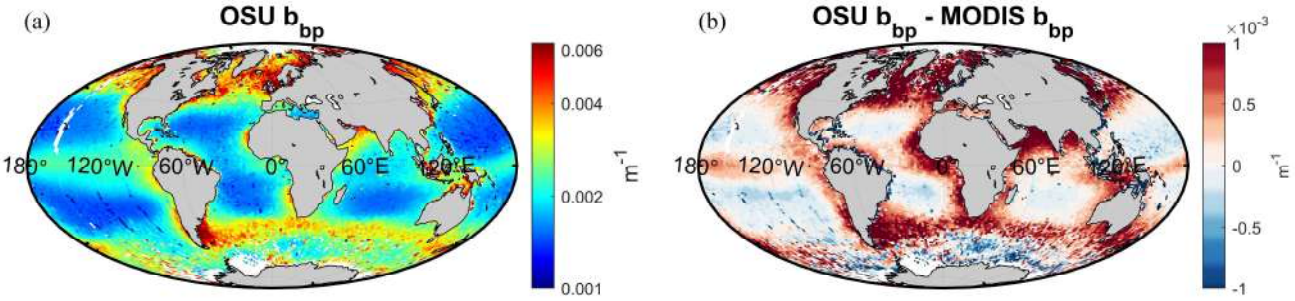


Fig. 15 Comparisons of the annual mean 2008 OSU b_{bp} with the MODIS b_{bp}.

Then, the b_{bp} produced by OSU, the DNN and MODIS are matched with the biogeochemical Argo data. The bias, MAD, and mean absolute percentage difference (MAPD) are calculated as follows:

$$\text{bias} = \frac{1}{n}\sum_n (f - d) \tag{18}$$

$$MAD = \frac{1}{n}\sum_n |f - d| \tag{19}$$

$$MAPD = \frac{1}{n}\sum_n \frac{|f-d|}{d} \tag{20}$$

where d is the b_{bp} produced by the biogeochemical Argo data, and f is the b_{bp} produced by OSU, the DNN or MODIS.

Data points are matched if they fall within 9 km and occur within 12 hours of each other. Figure 16 (a) shows an example of matched data on 6 Jan 2016. The matching criteria are quite strict, and there is only a single match on this day (red asterisk) consistent with the 9 km and ±12 hour thresholds. Figure 16 (b) shows the matched b_{bp} data for both day and night from 2010 to 2017. The data that were separately matched by day and night are shown in Figure 16(c), where the asterisk represents daytime data, and the circle represents the night data. There are 125 matched points in total, including 80 daytime and 46 nighttime data points. In terms of the statistical analysis in Table 3, the MODIS b_{bp} has the highest R^2 for all data, while the OSU b_{bp} (Behrenfeld et al. 2019) has the worse statistical results, particularly in terms of the MAPD, which is approximately twice that of the MODIS b_{bp} and DNN b_{bp}. These large differences may be caused by deviations in the observed conditions from the assumed relationship between $b(\pi)$ and b_{bp} (0.32 in (Bisson et al. 2021a) and 0.16 in (Behrenfeld et al. 2017)). Overall, the DNN b_{bp} and MODIS b_{bp} have similar R^2, RMSD, Bias, MAD and MAPD values. These results indicate that our DNN algorithms have better consistency with Argo data than the physics-based algorithm.

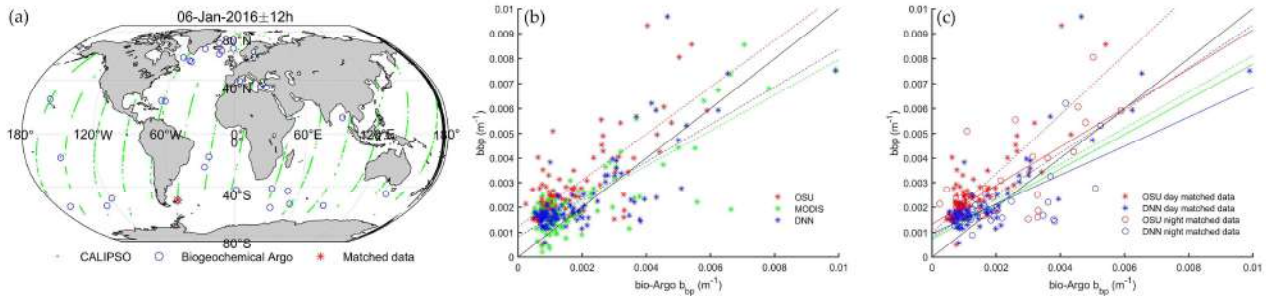


Fig. 16. Map of locations with coincident floats on 6 Jan 2016 ±12 hours (a), matched data both for day and night (b), separate day and night matched data (c). The asterisk represents the matched data, and the dashed lines represent the best-fit function in (b). The asterisk and circles represent daytime data and night data, respectively. Dashed lines and solid lines represent daytime and night best-fit functions, respectively, in (c).

Table 3. Statistical analysis of comparison between the Argo b_{bp} and the b_{bp} produced by OSU,

MODIS and the DNN

		R^2	RMSD (m^{-1})	Bias (m^{-1})	MAD (m^{-1})	MAPD
All data	Argo vs. OSU	0.59	0.0009	0.0011	0.0012	123.5%
	Argo vs. MODIS	0.69	0.0008	0.0003	0.0007	57.8%
	Argo vs. DNN	0.62	0.0008	0.0004	0.0008	62.3%
Daytime data	Argo vs. OSU	0.67	0.0008	0.0012	0.0012	121.3%
	Argo vs. DNN	0.68	0.0008	0.0005	0.0007	63.2%
Nighttime data	Argo vs. OSU	0.56	0.001	0.0009	0.0012	125.8%
	Argo vs. DNN	0.61	0.0008	0.0002	0.0007	56.5%

4.6 Global seasonal climatology b_{bp} and POC derived from CALIOP data

Figure 17 shows the global distribution of b_{bp} during the daytime and nighttime and their difference for each season. All seasonal daytime results have a similar pattern: the oligotrophic waters between 40° N and 40° S have the lowest value, while the regions in high latitudes, coastal regions and the eastern Pacific upwelling regions have higher values. The maximum values are found in the polar regions. In the Southern Ocean, the b_{bp} distribution changes dramatically. In the area south of 40° S, the b_{bp} is quite low in June-August. Then, it increases over time and reaches its maximum in December-February. After that, it decreases in March-May. The results at night have a similar global distribution in general, but the b_{bp} at night in the equatorial Atlantic is greater than the day results represented by blue in Figures 17(c), 17(f), 17(i) and 17(l). In contrast, the results during the daytime are greater than the results at nighttime in some coastal regions, represented by red in Figures 17(c), 17(f), 17(i), and 17(l). This result may be caused by the diurnal migration of plankton (Behrenfeld et al. 2019), but the model uncertainty exerts some subtle influence on the diurnal changes as well. In addition, the differences vary with the seasons. The difference is greatest in September-November (blue in Figure 17(i)). Then, the difference decreases in the Atlantic but increases in the Eastern Pacific in December-February, as shown in Figure 17(l). The difference is much less in March-May and June-August, as shown in Figures 17(c) and 17(f).

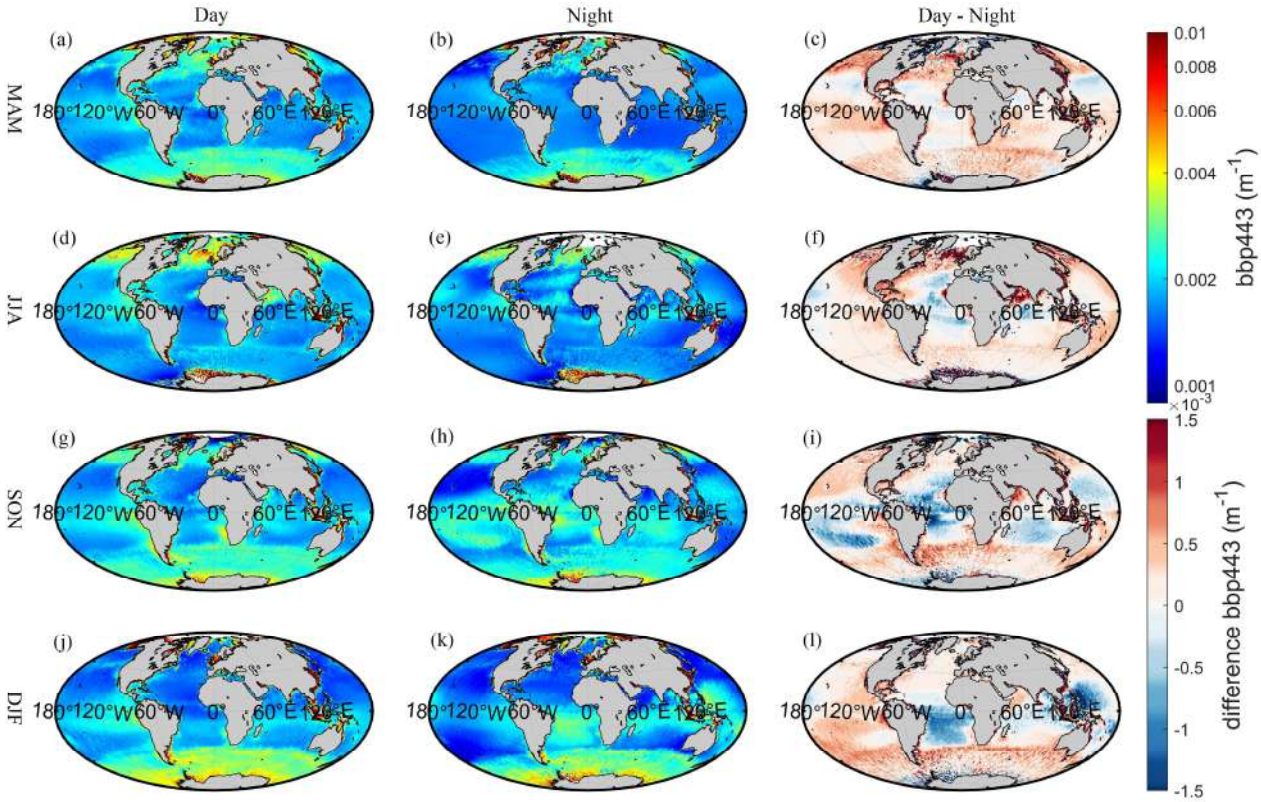


Fig. 17. Seasonal climatology distribution of $b_{bp}443$ derived from CALIOP and the DNN. (a) March-May at daytime; (b) March-May at nighttime; (c) March-May difference between day and night (b_{bp} (daytime)- b_{bp} (nighttime)); (d) June-August at daytime; (e) June-August at nighttime; (f) June-August difference; (g) September-November at daytime; (h) September-November at nighttime; (i) September-November difference; (j) December-February at daytime; (k) December-February at nighttime; and (l) December-February difference. The data are seasonally averaged for the 2008-2021 period and have been gridded to 2° longitude × 1° latitude.

The seasonal POC results are shown in Figure 18, which shows that the global distribution of POC in different seasons is quite similar during both daytime and nighttime. The POC gradually decreases from high latitudes to low latitudes, and it is much greater in coastal areas than in the open ocean. The minimum is located in oligotrophic waters between 40° N and 40° S. However, the Indian Ocean has a higher POC than the Pacific and Atlantic Oceans at low latitudes. The difference between day and night shows that the POC during the daytime is much greater than the POC at night, especially in coastal areas and the equatorial Atlantic region. The difference is much lower in September-November, as shown in Figure 18 (i), than in other seasons. However, the daytime POC is much lower than the results at night in the Pacific in December-February, as shown in Figure

607 18(l).

608 Figure 19 shows the relative percent difference calculated based on the climatology averaged
609 data in $b_{bp}443$ between MODIS ($b_{bp}443M$) and CALIOP ($b_{bp}443C$) produced by OSU and the DNN.
610 Figure 19(a) shows the result between $b_{bp}443M$ and $b_{bp}443C$ by OSU, and the result is basically the
611 same as that in a previous report (Bisson et al. 2021a). Figure 19(b) shows good agreement between
612 $b_{bp}443M$ and $b_{bp}443C$, and Figure 19(c) shows the comparisons between OSU and the DNN in terms
613 of $b_{bp}443C$. The averaged relative percent difference is within 30%; this difference may be present
614 because the physics-based method used empirical linear relationships (a $b(\pi)/b_{bp}$ value of 0.32 and a
615 δ_p/Kd value of 2) without considering the spatial and temporal variability associated with conversion
616 coefficients. This is addressed in the present study since the proposed DNN algorithm automatically
617 adjusts the conversion factors using the ocean surface and subsurface signals across latitude and time.
618 It is noted that MODIS data are being considered in the current study as a 'reference' even though we
619 know that these data also have significant errors. A significant seasonal bias was identified in the
620 MODIS remote sensing reflectance (Bisson et al. 2021b), and the b_{bp} retrieved from the CALIOP
621 data that were corrected for crosstalk and analyzed using off-nadir measurements using a
622 physics-based approach outperformed MODIS data in some cases (Bisson et al. 2021a).

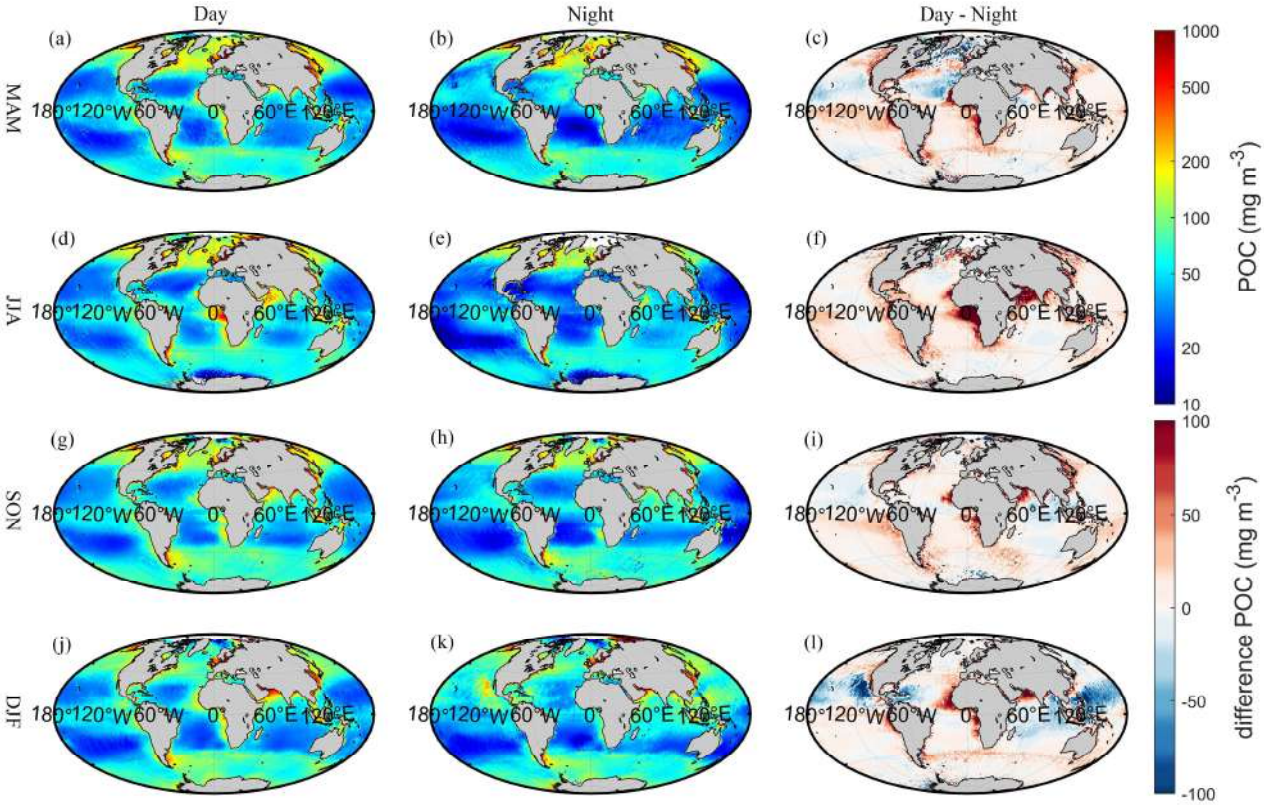
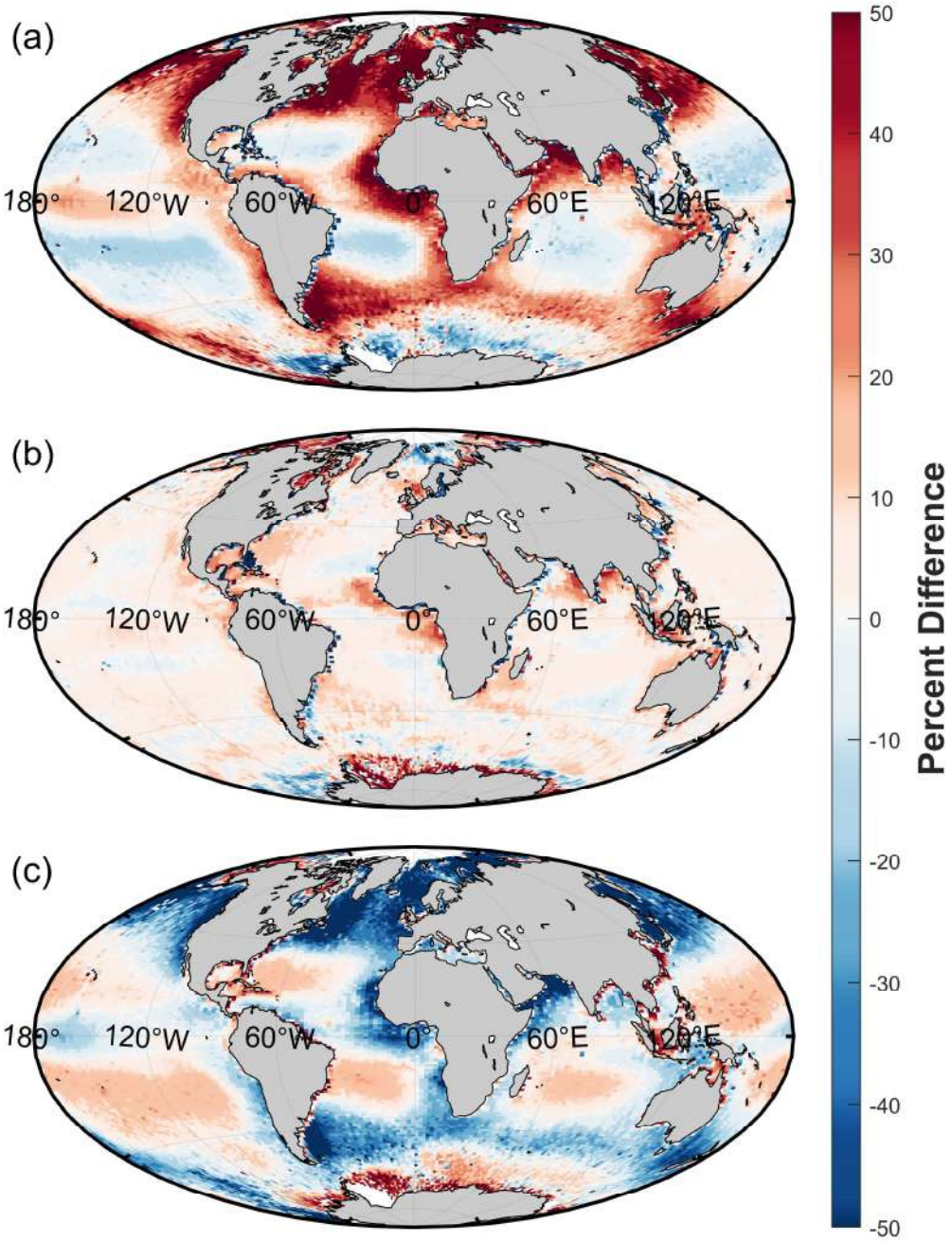


Fig. 18. Seasonal climatology distribution of POC derived from CALIOP and the DNN. (a) March-May at daytime; (b) March-May at nighttime; (c) March-May difference between day and night (b_{bp} (daytime)- b_{bp}(nighttime)); (d) June-August at daytime; (e) June-August at nighttime; (f) June-August difference; (g) September-November at daytime; (h) September-November at nighttime; (i) September-November difference; (j) December-February at daytime; (k) December-February at nighttime; and (l) December-February difference. The data are the seasonal average climatology for the 2008-2021 period and have been gridded to 2° longitude × 1° latitude.



631

632 Fig. 19. Climatology averaged relative percent difference in $b_{bp}443$ between MODIS and CALIOP
633 produced by OSU and the DNN ($100 \times [b_{bp}443C - b_{bp}443M]/ b_{bp}443M$). (a) MODIS $b_{bp}443$ vs. CALIOP
634 $b_{bp}443$ by OSU; (b) MODIS $b_{bp}443$ vs. CALIOP $b_{bp}443$ by the DNN; and (c) CALIOP $b_{bp}443$ by the
635 DNN vs. CALIOP $b_{bp}443$ by OSU.

636

637 **4.7 Seasonal climatology b_{bp} and POC in polar regions derived from CALIOP data**

638 Figure 20 shows the seasonal climatology distribution of $b_{bp}443$ derived from CALIOP in the

polar regions for the period 2008 to 2021, while the corresponding MODIS product is shown in Figure 21. The results indicate that the b_{bp} in polar regions changes considerably with the seasons. The minimum b_{bp} values appear during the boreal winter (December-February) in the Arctic, as shown in Figure 20(d), and the austral winter (June-August) in the Antarctic, as shown in Figure 20(f). Then, b_{bp} begins to increase over time and reaches its peak in the boreal summer (June-August) in the Arctic, as shown in Figure 20(b), and in the austral summer (December-February) in the Antarctic, as shown in Figure 20(h). In the following season, b_{bp} decreases. Because of the influence of continents, the distribution of b_{bp} in the Arctic is more complicated than that in Antarctic regions. However, it is clear that the near-shore b_{bp} values are higher than those offshore. In the Arctic regions, the coverage in winter (December-February) is smaller than that in other seasons, and the b_{bp} is much higher in the coastal area (red in Figure 20 (d)). In summer (June-August), the CALIOP data cover nearly all of the Arctic, as shown in Figure 20(b). In the Antarctic regions, the coverage is at its minimum in June-August because of sea ice. Compared with the seasonal MODIS product shown in Figure 21, the CALIOP b_{bp} coverage is much higher, especially in the Arctic region during autumn and winter (Figure 20(c-d) vs. Figure 21(c-d)). Passive remote sensing data can only be acquired during the daytime in the Antarctic region during the spring and summer (Figure 20(e-f) vs. Figure 21(e-f)), which means that the ability to sample high-latitude ocean areas during polar nights is limited; thus, only a small amount of data are available for these regions at these times of the year. This analysis also suggests that the CALIOP data could yield a wealth of new information about subsurface ice-edge phytoplankton blooms (Lu et al. 2020b).

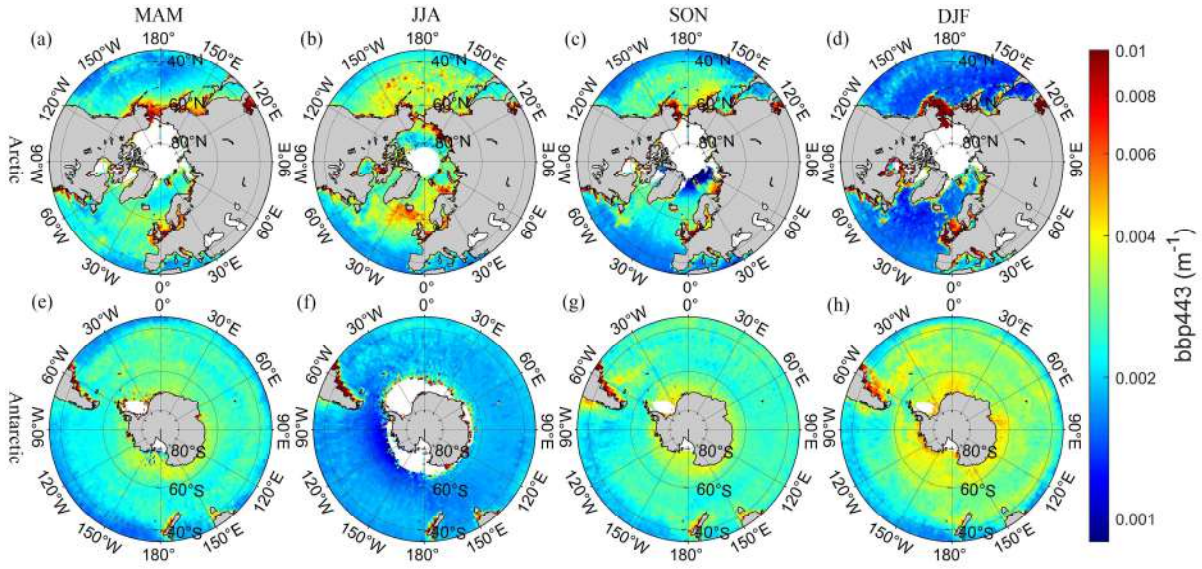


Fig. 20. Seasonal distribution of $b_{bp}443$ derived from CALIOP in polar regions. (a) March-May in the Arctic region; (b) June-August in the Arctic region; (c) September-November in the Arctic region; (d) December-February in the Arctic region; (e) March-May in the Antarctic region; (f) June-August in the Antarctic region; (g) September-November in the Antarctic region; and (h) December-February in the Antarctic region.

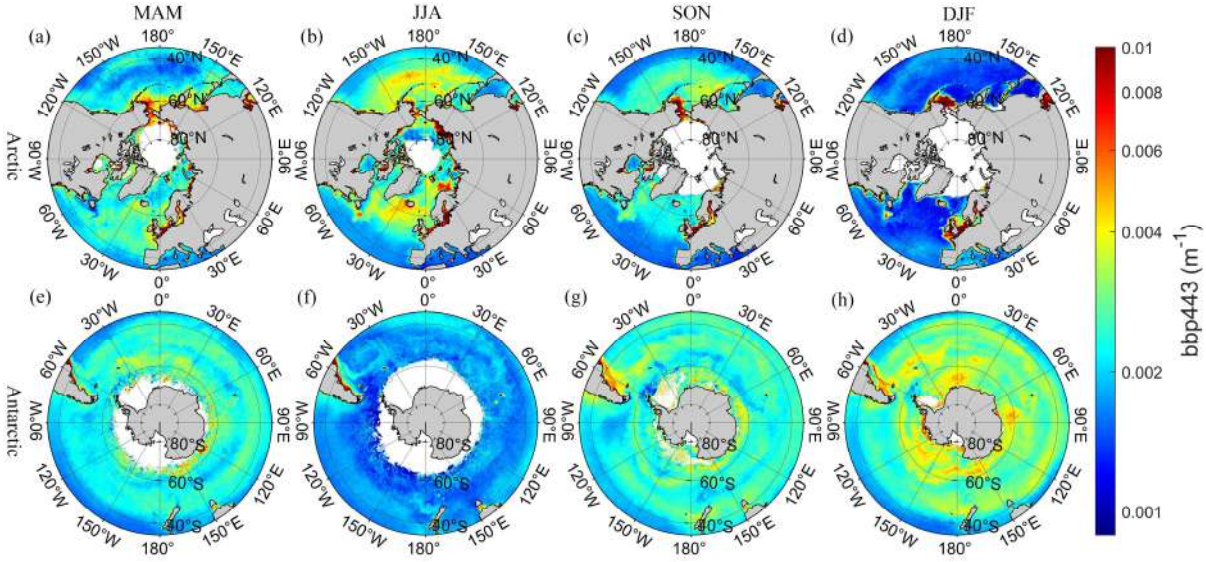


Fig. 21. Seasonal climatology distribution of MODIS $b_{bp}443$ in polar regions. (a) March-May in the Arctic region; (b) June-August in the Arctic region; (c) September-November in the Arctic region; (d) December-February in the Arctic region; (e) March-May in the Antarctic region; (f) June-August in the Antarctic region; (g) September-November in the Antarctic region; and (h) December-February in the Antarctic region.

Figures 22 and 23 show the seasonal distribution of the POC derived from CALIOP and MODIS in the polar regions, respectively. Compared to the b_{bp}, the POC has much less seasonal variability.

The POC in the Arctic regions is much higher than that in the Antarctic regions. Similar to the distribution of b_{bp}, these regions cannot be fully covered in polar winters due to the presence of sea ice. (Arctic: December-February, Antarctic: June-August), as shown in Figures 22(d) and 22(f). The seasonal CALIOP POC distribution during December-February shows that there is a local minimum of POC in the Barents Sea region, which is probably related to convection and deep water formation in the Arctic Ocean-Greenland Sea system (Rudels and Quadfasel 1991). This pattern cannot be observed from MODIS measurements, as evidenced by Figure 23(d). In the water near the North Pole, the POC reaches a minimum in September-November, as shown in Figure 22(c). The MODIS coverage is much smaller than that of CALIOP in the winter months. These results demonstrate CALIOP's ability to acquire data over the polar oceans, which can be used to complement the passive ocean color datasets that lack coverage in these areas.

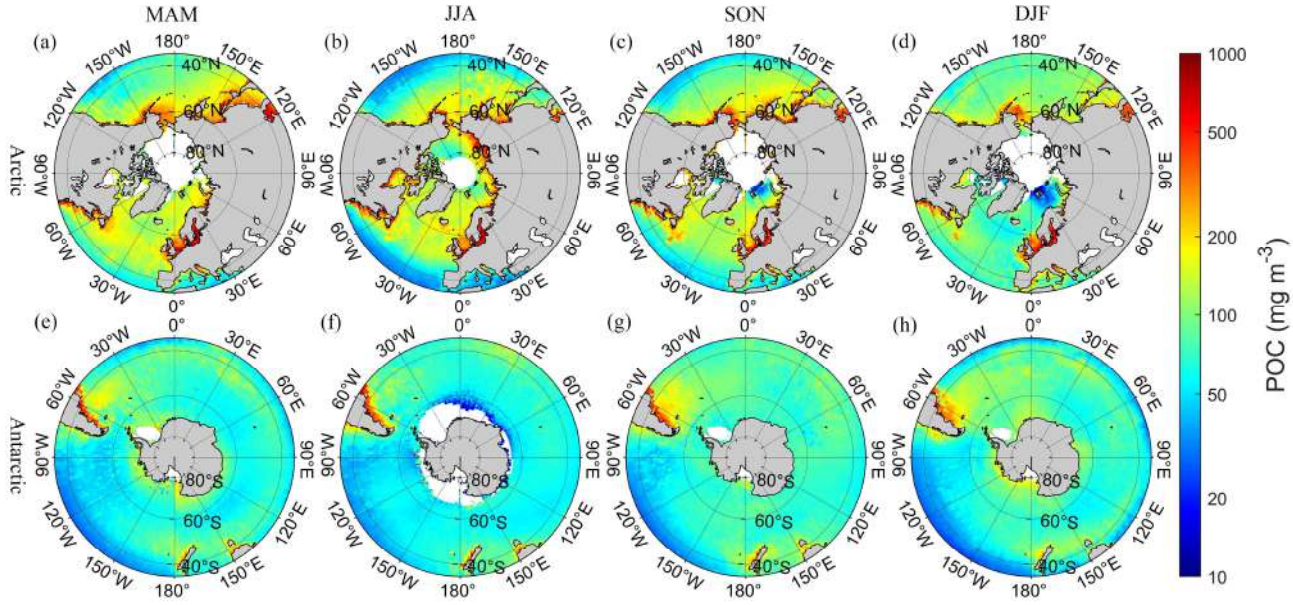


Fig. 22. Seasonal distribution of POC derived from CALIOP in polar regions. (a) March-May in the Arctic region; (b) June-August in the Arctic region; (c) September-November in the Arctic region; (d) December-February in the Arctic region; (e) March-May in the Antarctic region; (f) June-August in the Antarctic region; (g) September-November in the Antarctic region; and (h) December-February in the Antarctic region.

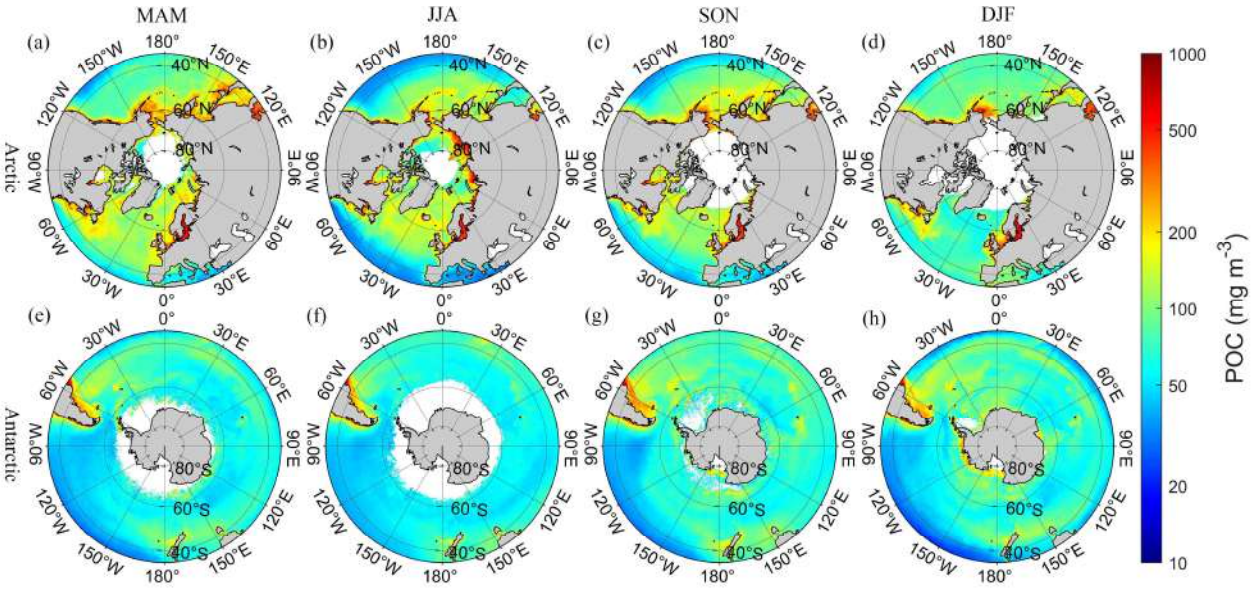


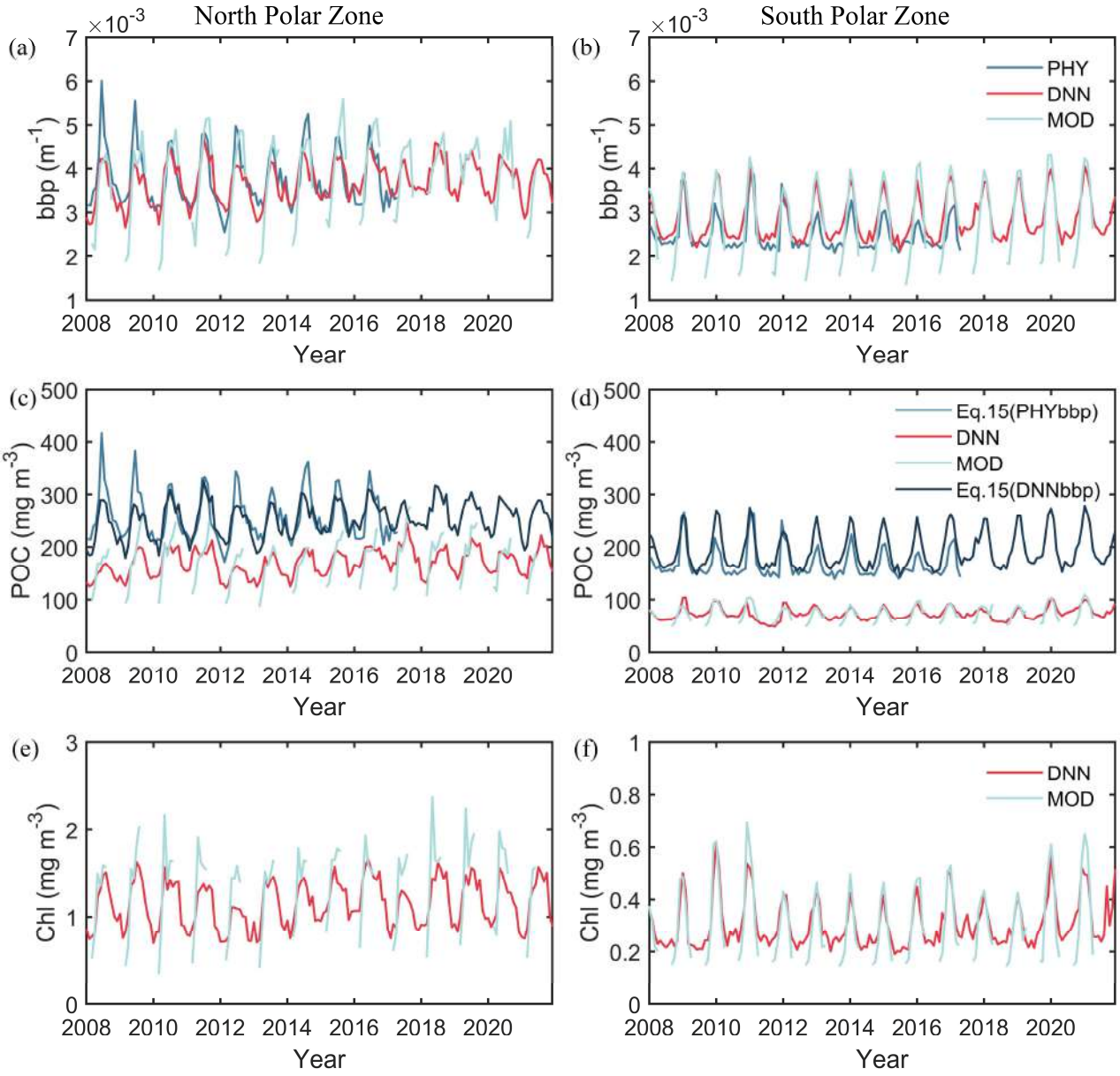
Fig. 23. Seasonal distribution of the MODIS POC in polar regions. (a) March-May in the Arctic region; (b) June-August in the Arctic region; (c) September-November in the Arctic region; (d) December-February in the Arctic region; (e) March-May in the Antarctic region; (f) June-August in the Antarctic region; (g) September-November in the Antarctic region; and (h) December-February in the Antarctic region.

4.8 Comparisons between the products of annual cycles and diurnal changes

Compared with the Oregon SU b_{bp} data shown in Figure 24(a,b), the DNN b_{bp} is slightly smaller than the physics-based b_{bp} in the north polar regions shown in Figure 24 (a) and much greater than the physics-based b_{bp} in the south polar regions shown in Figure 24(b). The DNN estimates are in better agreement with the MODIS products, especially in the south polar regions. Comparisons between the DNN POC, the MODIS POC, the POC computed from the physics-based b_{bp} using Equation (15), and the POC from the DNN b_{bp} using Equation (15) are shown in Figure 24(c,d). The differences are mainly caused by an inconsistency between the POC computed with the Rrs band ratio for MODIS standard products and the POC products based on b_{bp} and Equation (15) (Behrenfeld et al. 2005; Stramski et al. 2008). Previous evaluation shows that the Rrs band ratio algorithm has higher performance than the algorithm based on b_{bp} across a broad range of *in situ* data(Evers-King et al. 2017).

The annual cycles of Chl in polar regions are shown in Figure 24 (e,f). Until now, there have

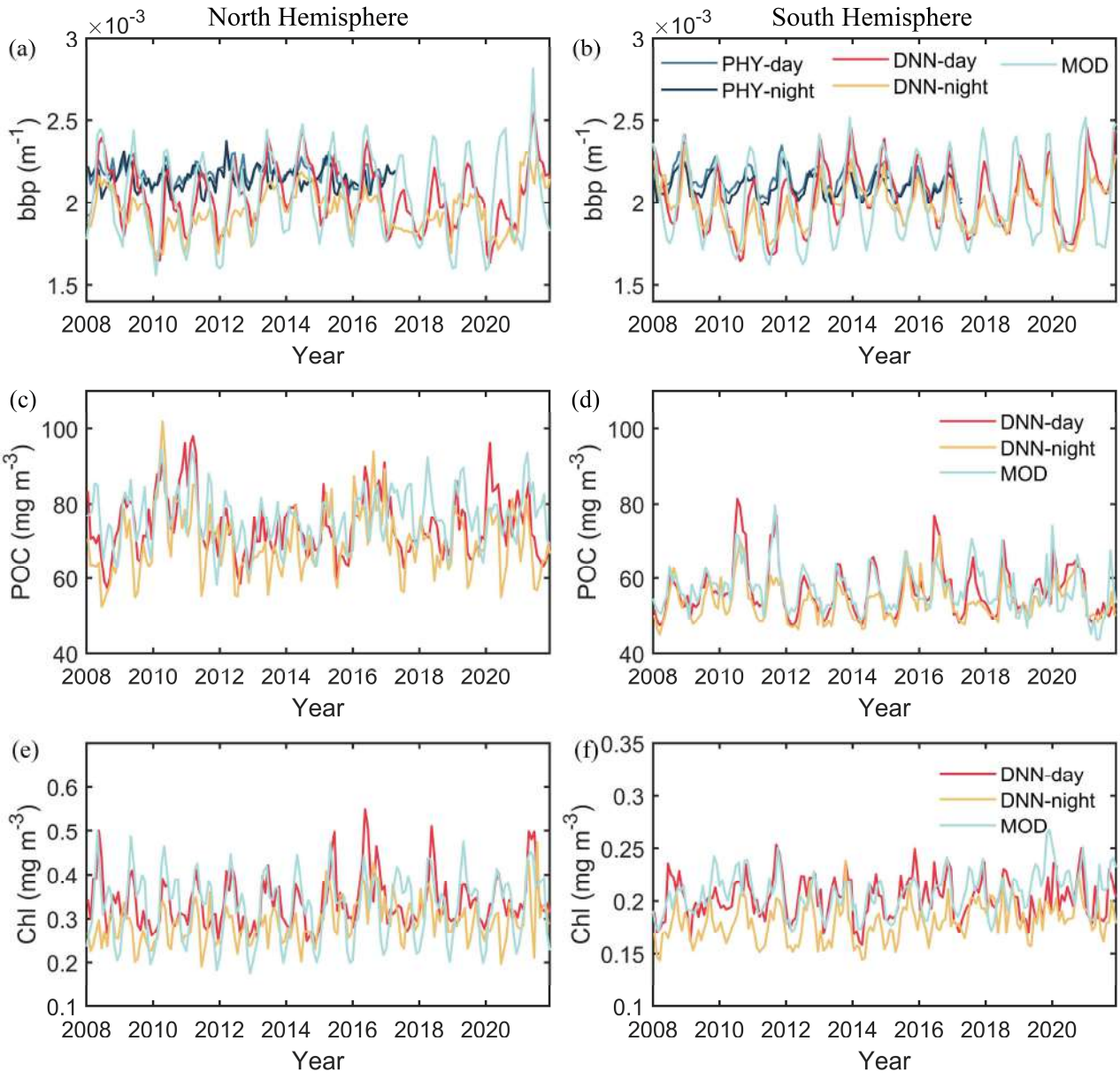
709 been no available CALIOP POC and Chl products based on physics-based methods, while
710 DNN-based methods can produce POC and Chl independently of the CALIOP b_{bp} values, which
711 may be partly explained POC-b_{bp} and Chl-b_{bp} model errors. With these DNN-based data,
712 carbon-to-chlorophyll (C:Chl and POC:Chl) can also be obtained for further analysis of the seasonal
713 and spatial dynamics of phytoplankton relative to nutrient richness in polar regions (Jakobsen and
714 Markager 2016).



715

716 Fig. 24. (Top) b_{bp} spatial average in polar regions with the physics-based (PHY), DNN, and MODIS
717 products. (Middle) POC spatial average in polar regions. (Bottom) Chl spatial average in polar regions.

718 The diurnal annual cycles of b_{bp}, POC and Chl in the global oceans where latitudes are less than
719 60° are shown in Figure 25. For b_{bp}, the values during the daytime are much greater than values
720 at nighttime in summer, but there are no apparent differences in winter in both the Northern
721 and Southern Hemispheres, as shown in Figure 25(a,b). POC has no apparent diurnal changes
722 in annual monthly cycles in the Northern Hemisphere, as shown in Figure 25(c), and Chl
723 during the daytime always has higher values, as shown in Figure 25(e,f). The diurnal
724 climatological monthly changes are shown in Figure 26. Compared with the physics-based b_{bp},
725 the DNN b_{bp} has a much larger seasonal amplitude, i.e., higher values in summer and lower
726 values in winter. Both exhibit significant diurnal changes in summer, and the diurnal changes in
727 the DNN are much greater than those in the physics-based data. The maximum diurnal changes
728 in POC occur in August in the Southern Hemisphere. Even considering the uncertainty of the
729 data (35%), significant diurnal variations can be determined in several months: the Chl during
730 the daytime is 70% larger than that at nighttime in June in the Northern Hemisphere. It should
731 be noted that, there may be some minor problems with using the relationship derived from day-time
732 CALIOP data and MODIS Chl to compute CALIOP-based night-time Chl. Considering different
733 chlorophyll concentration has different scattering properties, it is possible to retrieve Chl by using
734 LIDAR signals (Chen et al. 2021) and some previous attempts have been made to develop the
735 relationship between Chl and CALIOP subsurface backscatter β_{w+} (Lu et al. 2014) as well. Besides,
736 the evaluation with biogeochemical Argo data from 2010 to 2021 shows that R^2, RMSD, Bias, MAD
737 and MAPD for CALIOP-based day-time Chl are 0.52, 0.31, -0.09, 0.23 and 55%, respectively
738 (N=177). The corresponding values for night-time Chl are 0.50, 0.39, 0.02, 0.21 and 68.1%,
739 respectively (N=121). Therefore, DNN is an effective method to retrieve Chl from CALIOP. The
740 relationship between lidar backscatter signal and Chl will be further investigated with in situ data.



741

742 Fig. 25. Diurnal annual cycles of CALIOP b_{bp}, POC and Chl in global oceans. (Top) b_{bp} spatial average
743 in global oceans (where latitudes are less than 60°) during daytime and nighttime. (Middle) POC spatial
744 average in global oceans. (Bottom) Chl spatial average in global oceans. PHY represents the
745 physics-based method.

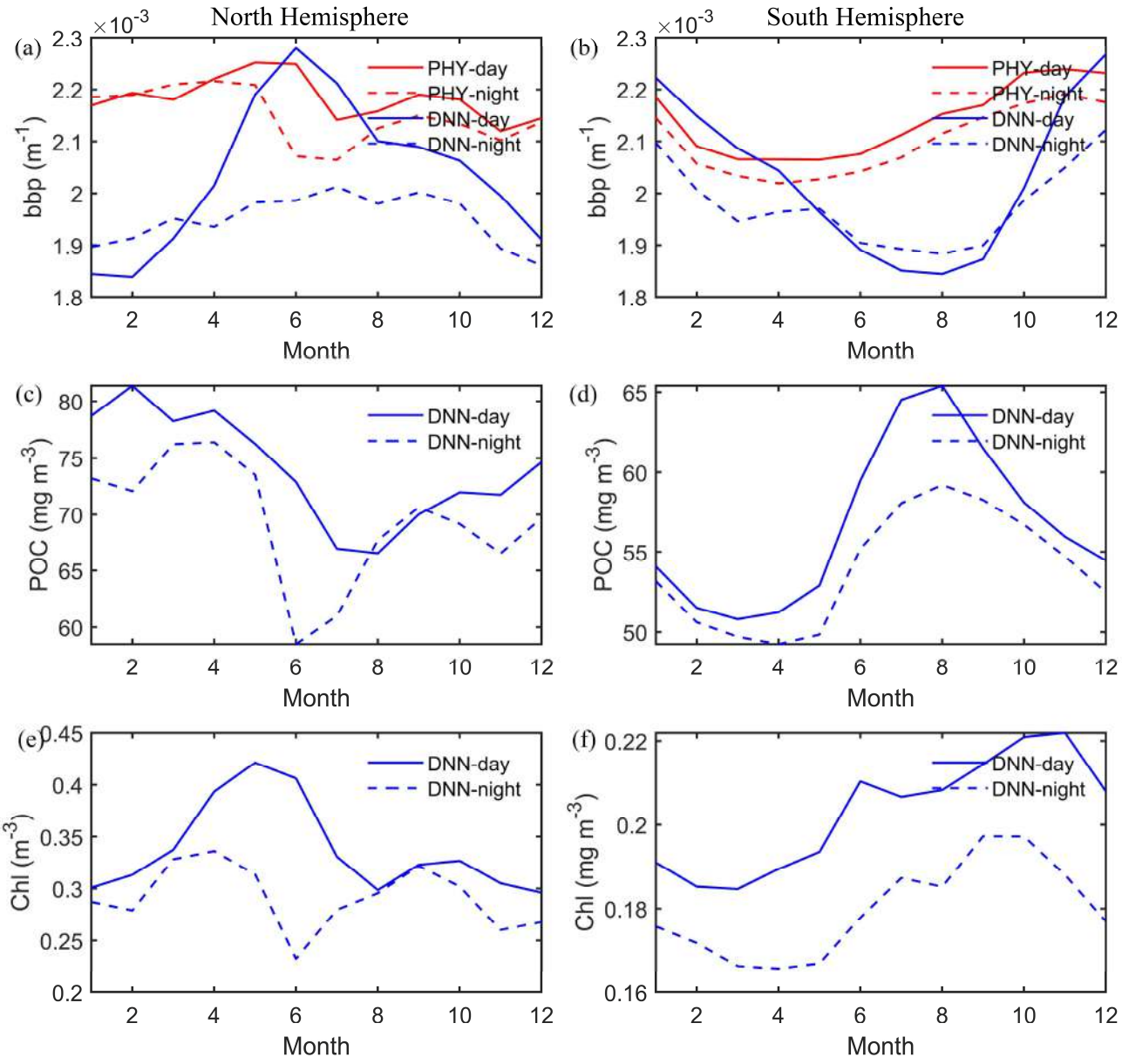


Fig. 26. Monthly climatological average of bbp, POC and Chl during daytime and nighttime in global oceans (where latitudes are less than 60°). PHY represents the physics-based method.

5. Summary and conclusion

In this study, the effects of both the detector TR and the PBS polarization crosstalk have been removed from the data. We found that the backscatter signal attenuation was greatly reduced after TR correction and that the cross-polarization signal was further reduced after CT correction. As a result, the corrected total depolarization ratio is much less than the measured values, which was

reduced by nearly 30%. The higher determination coefficients between the corrected total depolarization ratio and MODIS $b_{bp}443$ and POC indicate the necessity of these corrections. The filtering procedure plays an important role in data processing since filtering can negate the influence of the lower atmosphere, erroneous negative values, white caps from waves, signal saturation, and sea ice, thus ensuring that the data are as accurate as possible.

An innovative DNN approach for retrieving the ocean subsurface b_{bp} and POC from CALIOP lidar measurements was developed in this study to avoid the uncertainties introduced by assuming the nature of the relationship between $b(\pi)$ and b_{bp}. This data-based approach is independent of the diffuse attenuation coefficient and does not require additional assumptions regarding the relationship between b_{bp} and $b(\pi)$ and therefore could constitute an effective alternative to previous physics-based methods. The DNN is trained using the CALIOP total ocean column-integrated depolarization ratio and subsurface column-integrated perpendicular backscatter together with collocated MODIS b_{bp} and POC products. The advantage of this study is in training CALIOP data using the daytime MODIS product and applying it to nighttime and polar regions when MODIS data are unavailable. Statistical analyses comparing Argo b_{bp} and b_{bp} produced by OSU (Behrenfeld et al. 2019), MODIS and the DNN indicate that our DNN algorithms are in better agreement with the biogeochemical Argo data than the physics-based method. It should be noted that, in essence, the DNN approach fits the relationships between the CALIOP retrieved properties and the MODIS properties. Accordingly, if there are errors/biases in the MODIS products, then these biases will be fitted to and imprinted on the resultant CALIOP b_{bp} and POC data. In other words, in this study, the MODIS data are considered 'reference' data even though MODIS data have a significant seasonal bias (Bisson et al. 2021b).

The global seasonal b_{bp} and POC datasets were derived from CALIOP measurements between 2008 and 2021 using the DNN algorithm. The use of lidar measurements to derive b_{bp} and POC without passive data is of great interest to the ocean color community. Nighttime measurements from

780 spaceborne lidar increase knowledge of the diurnal variability of ocean optical properties, while
781 measurements in polar regions fill in the gaps that passive measurements of ocean color may leave in
782 these regions. We find that b_{bp} varies dramatically with season and that CALIOP can provide global
783 coverage throughout the year. Taking advantage of the fact that lidar can work at night and in polar
784 regions, the products derived from CALIOP by the DNN method can compensate for the lack of
785 high-latitude data and can be used to study the diurnal variations in ocean optical properties.
786 However, because of the large error in the MODIS data, this method still has some disadvantages
787 and shortcomings. With the increase in *in situ* data (such as Argo), this model will be improved in
788 the future by replacing MODIS data with *in situ* data. Note that the MERRA-2 wind speeds are used
789 here; hence, the retrieval accuracy depends on the accuracy of the MERRA-2 wind speed. Due to the
790 coarse resolution in time and space, this wind speed data may introduce significant error into the
791 CALIOP retrieval. A high spatial resolution ocean surface wind speed from CALIOP (Murphy and
792 Hu 2021) will be used to improve the model in the future. The b_{bp} and POC data from the DNN
793 approach will be available to the community upon request.
794
795 **Declaration of Competing Interest**
796 The authors declare that they have no known competing financial interests or personal
797 relationships that could have appeared to influence the work reported in this paper.
798
799 **Funding**
800 Key Special Project for Introduced Talents Team of Southern Marine Science and Engineering
801 Guangdong Laboratory (Guangzhou) (GML2019ZD0602), the National Natural Science Foundation
802 (41901305; 61991453), and the Zhejiang Natural Science Foundation (LQ19D060003).
803

Acknowledgments

The authors gratefully acknowledge the NASA Langley Research Center for providing the CALIOP data through the Atmospheric Science Data Center (https://asdc.larc.nasa.gov/data/CALIPSO/). We are grateful to the NASA Ocean Biology Processing Group for providing the MODIS products (https://oceandata.sci.gsfc.nasa.gov/) and biogeochemical-Argo data from the Argo Data Assembly Center (ftp://ftp.ifremer.fr/ifremer/argo/dac/). We thank the anonymous reviewers for their suggestions, which significantly improved the presentation of the paper.

References

Behrenfeld, M.J., Boss, E., Siegel, D.A., & Shea, D.M. (2005). Carbon-based ocean productivity and phytoplankton physiology from space - art. no. GB1006. *Global Biogeochemical Cycles*, 19

Behrenfeld, M.J., Gaube, P., Della Penna, A., O'Malley, R.T., Burt, W.J., Hu, Y., Bontempi, P.S., Steinberg, D.K., Boss, E.S., Siegel, D.A., Hostetler, C.A., Tortell, P.D., & Doney, S.C. (2019). Global satellite-observed daily vertical migrations of ocean animals. *Nature, 576*, 257-261

Behrenfeld, M.J., Hu, Y., Hostetler, C.A., Dall'Olmo, G., Rodier, S.D., Hair, J.W., & Trepte, C.R. (2013). Space-based lidar measurements of global ocean carbon stocks. *Geophysical Research Letters, 40*, 4355-4360

Behrenfeld, M.J., Hu, Y., O'Malley, R.T., Boss, E.S., Hostetler, C.A., Siegel, D.A., Sarmiento, J.L., Schulien, J., Hair, J.W., Lu, X., Rodier, S., & Scarino, A.J. (2017). Annual boom–bust cycles of polar phytoplankton biomass revealed by space-based lidar. *Nature Geoscience, 10*, 118-122

Berthon, J.F., Shybanov, E., Lee, E.G., & Zibordi, G. (2007). Measurements and modeling of the volume scattering function in the coastal northern Adriatic Sea. *Applied optics, 46*, 5189-5203

Bisson, K.M., Boss, E., Werdell, P.J., Ibrahim, A., & Behrenfeld, M.J. (2021a). Particulate Backscattering in the Global Ocean: A Comparison of Independent Assessments. *Geophysical Research Letters, 48*, e2020GL090909

Bisson, K.M., Boss, E., Werdell, P.J., Ibrahim, A., Frouin, R., & Behrenfeld, M.J. (2021b). Seasonal bias in global ocean color observations. *Applied Optics, 60*, 6978-6988

Bisson, K.M., Boss, E., Westberry, T.K., & Behrenfeld, M.J. (2019). Evaluating satellite estimates of particulate backscatter in the global open ocean using autonomous profiling floats. *Opt Express, 27*, 30191-30203

Blondeau-Patissier, D., Gower, J.F.R., Dekker, A.G., Phinn, S.R., & Brando, V.E. (2014). A review of ocean color remote sensing methods and statistical techniques for the detection, mapping and analysis of phytoplankton blooms in coastal and open oceans. *Progress in Oceanography, 123*, 123-144

Boss, E., Pegau, W.S., Lee, M., Twardowski, M., Shybanov, E., And, G.K., & Baratange, F. (2004). Particulate backscattering ratio at LEO 15 and its use to study particle composition and distribution. *Journal of Geophysical Research: Oceans*

Boss, E., & Pegau, W.S.J.A.O. (2001). Relationship of light scattering at an angle in the backward direction to the backscattering coefficient*, 40*, 5503-5507

Brewin, R.J.W., Ciavatta, S., Sathyendranath, S., Jackson, T., Tilstone, G., Curran, K., Airs, R.L., Cummings, D., Brotas, V., Organelli, E., Dall'Olmo, G., & Raitsos, D.E. (2017). Uncertainty in Ocean-Color Estimates of Chlorophyll for Phytoplankton Groups. *Frontiers in Marine Science, 4*

Carvalho, D. (2019). An Assessment of NASA's GMAO MERRA-2 Reanalysis Surface Winds. *Journal of Climate, 32*, 8261-8281

Chami, M., Marken, E., Stamnes, J., Khomenko, G., & Korotaev, G.J.J.o.G.R.O. (2006). Variability of the relationship between the particulate backscattering coefficient and the volume scattering function measured at fixed angles*, 111*

Chami, M., Thirouard, A., & Harmel, T. (2014). POLVSM (Polarized Volume Scattering Meter) instrument: an innovative device to measure the directional and polarized scattering properties of hydrosols. *Optics Express, 22*, 26403-26428

Chau, T.T.T., Gehlen, M., & Chevallier, F. (2022). A seamless ensemble-based reconstruction of surface ocean pCO2 and air–sea CO2 fluxes over the global coastal and open oceans. *Biogeosciences, 19*, 1087-1109

Chen, P. (2022). Subsurface phytoplankton vertical structure observations using offshore fixed platform-based lidar in the Bohai Sea for offshore responses to Typhoon Bavi. *Optics Express, 30*, 20614-20628

Chen, P., Jamet, C., & Liu, D. (2022). LiDAR Remote Sensing for Vertical Distribution of Seawater Optical Properties and Chlorophyll-a From the East China Sea to the South China Sea. *IEEE Transactions on Geoscience and Remote Sensing, 60*, 1-21

Chen, P., Jamet, C., Zhang, Z., He, Y., Mao, Z., Pan, D., Wang, T., Liu, D., & Yuan, D. (2021). Vertical distribution of subsurface phytoplankton layer in South China Sea using airborne lidar. *Remote Sensing of Environment, 263*, 112567

Chen, P., Mao, Z., Zhang, Z., Liu, H., & Pan, D. (2020a). Detecting subsurface phytoplankton layer in Qiandao Lake using shipborne lidar. *Opt Express, 28*, 558-569

Chen, P., Mao, Z., Zhang, Z., Liu, H., & Pan, D. (2020b). Detecting subsurface phytoplankton layer in Qiandao Lake using shipborne lidar. *Optics Express, 28*, 558-569

Chen, P., & Pan, D. (2019). Ocean Optical Profiling in South China Sea Using Airborne LiDAR. *Remote Sensing, 11*

Chen, P., Pan, D., Mao, Z., & Liu, H. (2019). A Feasible Calibration Method for Type 1 Open Ocean Water LiDAR Data Based on Bio-Optical Models. *Remote Sensing, 11*, 172

Churnside, J., Hair, J., Hostetler, C., & Scarino, A. (2018). Ocean Backscatter Profiling Using High-Spectral-Resolution Lidar and a Perturbation Retrieval. *Remote Sensing, 10*, 2003

Churnside, J., Marchbanks, R., Lembke, C., & Beckler, J. (2017). Optical Backscattering Measured by Airborne Lidar and Underwater Glider. *Remote Sensing, 9*, 379

Churnside, J., McCarty, B., & Lu, X. (2013). Subsurface Ocean Signals from an Orbiting Polarization Lidar. *Remote Sensing, 5*, 3457-3475

Churnside, J.H., & Marchbanks, R.D. (2015). Subsurface plankton layers in the Arctic Ocean. *Geophysical Research Letters, 42*, 4896-4902

Churnside, J.H., & Marchbanks, R.D. (2019). Calibration of an airborne oceanographic lidar using ocean backscattering measurements from space. *Opt Express, 27*, A536-A542

Churnside, J.H., Marchbanks, R.D., & Marshall, N. (2021). Airborne Lidar Observations of a Spring Phytoplankton Bloom in the Western Arctic Ocean. *Remote Sensing, 13*

Churnside, J.H., Marchbanks, R.D., Vagle, S., Bell, S.W., & Stabeno, P.J. (2020). Stratification, plankton layers, and mixing measured by airborne lidar in the Chukchi and Beaufort seas. *Deep Sea Research Part II: Topical Studies in Oceanography, 177*

Churnside, J.H., Sullivan, J.M., & Twardowski, M.S. (2014). Lidar extinction-to-backscatter ratio of the ocean. *Optics Express, 22*, 18698-18706

Claustre, H., Johnson, K.S., & Takeshita, Y. (2019). Observing the Global Ocean with Biogeochemical-Argo. *Annual*

Review of Marine Science, 12, 1-26
Collister, B.L., Zimmerman, R.C., Sukenik, C.I., Hill, V.J., & Balch, W.M. (2018). Remote sensing of optical characteristics and particle distributions of the upper ocean using shipboard lidar. *Remote Sensing of Environment, 215*, 85-96
Dickey, T., Lewis, M., & Chang, G. (2006). Optical oceanography: Recent advances and future directions using global remote sensing and in situ observations. *Reviews of Geophysics, 44*
Dionisi, D., Brando, V.E., Volpe, G., Colella, S., & Santoleri, R. (2020). Seasonal distributions of ocean particulate optical properties from spaceborne lidar measurements in Mediterranean and Black sea. *Remote Sensing of Environment, 247*
Evers-King, H., Martinez-Vicente, V., Brewin, R.J.W., Dall'Olmo, G., Hickman, A.E., Jackson, T., Kostadinov, T.S., Krasemann, H., Loisel, H., Röttgers, R., Roy, S., Stramski, D., Thomalla, S., Platt, T., & Sathyendranath, S. (2017). Validation and Intercomparison of Ocean Color Algorithms for Estimating Particulate Organic Carbon in the Oceans. *Frontiers in Marine Science, 4*
Getzewich, B.J., Vaughan, M.A., Hunt, W.H., Avery, M.A., Powell, K.A., Tackett, J.L., Winker, D.M., Kar, J., Lee, K.-P., & Toth, T.D. (2018). CALIPSO lidar calibration at 532 nm: version 4 daytime algorithm. *Atmos Meas Tech, 11*, 6309-6326
Gregor, L., & Gruber, N. (2021). OceanSODA-ETHZ: a global gridded data set of the surface ocean carbonate system for seasonal to decadal studies of ocean acidification. *Earth System Science Data, 13*, 777-808
Hostetler, C.A., Behrenfeld, M.J., Hu, Y., Hair, J.W., & Schulien, J.A. (2018). Spaceborne Lidar in the Study of Marine Systems. *Ann Rev Mar Sci, 10*, 121-147
Hu, L., Zhang, X., Xiong, Y., Gray, D.J., & He, M.X. (2020). Variability of relationship between the volume scattering function at 180° and the backscattering coefficient for aquatic particles. *Applied optics, 59*, C31-C41
Hu, Y., Lu, X., Zhai, P.-W., Hostetler, C.A., Hair, J.W., Cairns, B., Sun, W., Stamnes, S., Omar, A., Baize, R., Videen, G., Mace, J., McCoy, D.T., McCoy, I.L., & Wood, R. (2021). Liquid Phase Cloud Microphysical Property Estimates From CALIPSO Measurements. *Frontiers in Remote Sensing, 2*
Hu, Y., Stamnes, K., Vaughan, M., Pelon, J., Weimer, C., Wu, D., Cisewski, M., Sun, W., Yang, P., Lin, B., Omar, A., Flittner, D., Hostetler, C., Trepte, C., Winker, D., Gibson, G., & Santa-Maria, M. (2008). Sea surface wind speed estimation from space-based lidar measurements. *Atmospheric Chemistry and Physics Discussions, 8*, 2771-2793
Hunt, W.H., Winker, D.M., Vaughan, M.A., Powell, K.A., Lucker, P.L., & Weimer, C. (2009). CALIPSO Lidar Description and Performance Assessment. *Journal of Atmospheric and Oceanic Technology, 26*, 1214-1228
Jackson, T., Sathyendranath, S., & Mélin, F. (2017). An improved optical classification scheme for the Ocean Colour Essential Climate Variable and its applications. *Remote Sensing of Environment, 203*, 152-161
Jakobsen, H.H., & Markager, S. (2016). Carbon-to-chlorophyll ratio for phytoplankton in temperate coastal waters: Seasonal patterns and relationship to nutrients. *Limnology and Oceanography, 61*, 1853-1868
Jamet, C., Ibrahim, A., Ahmad, Z., Angelini, F., Babin, M., Behrenfeld, M.J., Boss, E., Cairns, B., Churnside, J., Chowdhary, J., Davis, A.B., Dionisi, D., Duforêt-Gaurier, L., Franz, B., Frouin, R., Gao, M., Gray, D., Hasekamp, O., He, X., Hostetler, C., Kalashnikova, O.V., Knobelspiesse, K., Lacour, L., Loisel, H., Martins, V., Rehm, E., Remer, L., Sanhaj, I., Stamnes, K., Stamnes, S., Victori, S., Werdell, J., & Zhai, P.-W. (2019). Going Beyond Standard Ocean Color Observations: Lidar and Polarimetry. *Frontiers in Marine Science, 6*
Josset, D., Rogers, R., Pelon, J., Hu, Y., Liu, Z., Omar, A., & Zhai, P.W. (2011). CALIPSO lidar ratio retrieval over the ocean. *Opt Express, 19*, 18696-18706
Kim, M.-H., Kim, S.-W., Yoon, S.-C., & Omar, A.H. (2013). Comparison of aerosol optical depth between CALIOP and MODIS-Aqua for CALIOP aerosol subtypes over the ocean. *Journal of Geophysical Research: Atmospheres, 118*, 13,241-213,252

Kim, M.-H., Omar, A.H., Tackett, J.L., Vaughan, M.A., Winker, D.M., Trepte, C.R., Hu, Y., Liu, Z., Poole, L.R., Pitts, M.C., Kar, J., & Magill, B.E. (2018). The CALIPSO Automated Aerosol Classification and Lidar Ratio Selection Algorithm. *Atmos Meas Tech, 11*, 6107-6135

Kingma, D., & Ba, J. (2014). Adam: A Method for Stochastic Optimization. *Computer Science*

Lacour, L., Larouche, R., & Babin, M. (2020). In situ evaluation of spaceborne CALIOP lidar measurements of the upper-ocean particle backscattering coefficient. *Optics Express, 28*, 26989-26999

Lee, J.H., Churnside, J.H., Marchbanks, R.D., Donaghay, P.L., & Sullivan, J.M. (2013a). Oceanographic lidar profiles compared with estimates from in situ optical measurements. *Appl Opt, 52*, 786-794

Lee, J.H., Churnside, J.H., Marchbanks, R.D., Donaghay, P.L., & Sullivan, J.M. (2013b). Oceanographic lidar profiles compared with estimates from in situ optical measurements. *Applied optics, 52*, 786-794

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*, 5755-5772

Li, J., Hu, Y., Huang, J., Stamnes, K., Yi, Y., & Stamnes, S. (2011). A new method for retrieval of the extinction coefficient of water clouds by using the tail of the CALIOP signal. *Atmospheric Chemistry and Physics, 11*, 2903-2916

Liu, H., Chen, P., Mao, Z., Pan, D., & He, Y. (2018). Subsurface plankton layers observed from airborne lidar in Sanya Bay, South China Sea. *Opt Express, 26*, 29134-29147

Lu, X., Hu, Y., Liu, Z., Zeng, S., & Trepte, C. (2013). *CALIOP receiver transient response study*. SPIE

Lu, X., Hu, Y., Omar, A., Baize, R., Vaughan, M., Rodier, S., Kar, J., Getzewich, B., Lucker, P., Trepte, C., Hostetler, C., & Winker, D. (2021a). Global Ocean Studies from CALIOP/CALIPSO by Removing Polarization Crosstalk Effects. *Remote Sensing, 13*

Lu, X., Hu, Y., Pelon, J., Trepte, C., Liu, K., Rodier, S., Zeng, S., Lucker, P., Verhappen, R., Wilson, J., Audouy, C., Ferrier, C., Haouchine, S., Hunt, B., & Getzewich, B. (2016). Retrieval of ocean subsurface particulate backscattering coefficient from space-borne CALIOP lidar measurements. *Opt Express, 24*, 29001-29008

Lu, X., Hu, Y., Trepte, C., Zeng, S., & Churnside, J.H. (2014). Ocean subsurface studies with the CALIPSO spaceborne lidar. *Journal of Geophysical Research: Oceans, 119*, 4305-4317

Lu, X., Hu, Y., Vaughan, M., Rodier, S., Trepte, C., Lucker, P., & Omar, A. (2020a). New attenuated backscatter profile by removing the CALIOP receiver's transient response. *Journal of Quantitative Spectroscopy and Radiative Transfer, 255*, 107244

Lu, X., Hu, Y., Yang, Y., Bontempi, P., Omar, A., & Baize, R. (2020b). Antarctic spring ice-edge blooms observed from space by ICESat-2. *Remote sensing of environment*, 111827

Lu, X., Hu, Y., Yang, Y., Neumann, T., Omar, A., Baize, R., Vaughan, M., Rodier, S., Getzewich, B., Lucker, P., Trepte, C., Hostetler, C., & Winker, D. (2021b). New Ocean Subsurface Optical Properties From Space Lidars: CALIOP/CALIPSO and ATLAS/ICESat‐2. *Earth and Space Science, 8*

Madhiarasan, M., & Deepa, S.N. (2017). Comparative analysis on hidden neurons estimation in multi layer perceptron neural networks for wind speed forecasting. *Artificial Intelligence Review, 48*, 449-471

Maritorena, S., Siegel, D.A., & Peterson, A.R. (2002). Optimization of a semianalytical ocean color model for global-scale applications. *Applied Optics, 41*, 2705-2714

McClain, C.R. (2009). A Decade of Satellite Ocean Color Observations. *Annual Review of Marine Science, 1*, 19-42

McGill, M.J., Vaughan, M.A., Trepte, C.R., Hart, W.D., Hlavka, D.L., Winker, D.M., & Kuehn, R. (2007). Airborne validation of spatial properties measured by the CALIPSO lidar. *Journal of Geophysical Research, 112*

Montes-Hugo, M.A., Churnside, J.H., Gould, R.W., Arnone, R.A., & Foy, R. (2010). Spatial coherence between remotely sensed ocean color data and vertical distribution of lidar backscattering in coastal stratified waters. *Remote Sensing of Environment, 114*, 2584-2593

Montes, M.A., Churnside, J., Lee, Z., Gould, R., Arnone, R., & Weidemann, A. (2011). Relationships between water

attenuation coefficients derived from active and passive remote sensing: a case study from two coastal environments. *Appl Opt, 50*, 2990-2999
Murphy, A., & Hu, Y. (2021). Retrieving Aerosol Optical Depth and High Spatial Resolution Ocean Surface Wind Speed From CALIPSO: A Neural Network Approach. *Frontiers in Remote Sensing, 1*
Pitts, M.C., Poole, L.R., & Gonzalez, R. (2018). Polar stratospheric cloud climatology based on CALIPSO spaceborne lidar measurements from 2006 to 2017. *Atmos. Chem. Phys., 18*, 10881-10913
Richter, K., Maas, H.-G., Westfeld, P., & Weiß, R. (2017). An Approach to Determining Turbidity and Correcting for Signal Attenuation in Airborne Lidar Bathymetry. *PFG – Journal of Photogrammetry, Remote Sensing and Geoinformation Science, 85*, 31-40
Roddewig, M.R., Churnside, J.H., & Shaw, J.A. (2020). Lidar measurements of the diffuse attenuation coefficient in Yellowstone Lake. *Appl Opt, 59*, 3097-3101
Rudels, B., & Quadfasel, D. (1991). Convection and deep water formation in the Arctic Ocean-Greenland Sea System. *Journal of Marine Systems, 2*, 435-450
Saylam, K., Brown, R.A., & Hupp, J.R. (2017). Assessment of depth and turbidity with airborne Lidar bathymetry and multiband satellite imagery in shallow water bodies of the Alaskan North Slope. *International Journal of Applied Earth Observation and Geoinformation, 58*, 191-200
Sharma, S., Sharma, S., & Athaiya, A. (2020). Activation Functions in Neural Networks. *International Journal of Engineering Applied Sciences and Technology, 4*, 310-316
Soja‐Woniak, M., Baird, M., Schroeder, T., Qin, Y., Clementson, L., Baker, B., Boadle, D., Brando, V., & Steven, A.D.L. (2019). Particulate Backscattering Ratio as an Indicator of Changing Particle Composition in Coastal Waters: Observations From Great Barrier Reef Waters. *Journal of Geophysical Research: Oceans, 124*
Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research, 15*, 1929-1958
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. (2008). 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
Sullivan, C.W., Arrigo, K.R., McClain, C.R., Comiso, J.C., & Firestone, J. (1993). Distributions of phytoplankton blooms in the southern ocean. *Science, 262*, 1832-1837
Sullivan, J.M., & Twardowski, M.S. (2009a). Angular shape of the oceanic particulate volume scattering function in the backward direction. *Appl Opt, 48*, 6811-6819
Sullivan, J.M., & Twardowski, M.S. (2009b). Angular shape of the oceanic particulate volume scattering function in the backward direction. *Applied optics, 48*, 6811-6819
Sullivan, J.M., & Twardowski, M.S.J.A.O. (2009c). Angular shape of the oceanic particulate volume scattering function in the backward direction*, 48*, 6811-6819
Werdell, P.J., Franz, B.A., Bailey, S.W., Feldman, G.C., Boss, E., Brando, V.E., Dowell, M., Hirata, T., Lavender, S.J., Lee, Z., Loisel, H., Maritorena, S., Mélin, F., Moore, T.S., Smyth, T.J., Antoine, D., Devred, E., d'Andon, O.H.F., & Mangin, A. (2013). Generalized ocean color inversion model for retrieving marine inherent optical properties. *Applied Optics, 52*, 2019-2037
Whitmire, A.L., Pegau, W.S., Karp-Boss, L., Boss, E., & Cowles, T.J.J.O.E. (2010). Spectral backscattering properties of marine phytoplankton cultures*, 18*, 15073-15093
Winker, D.M., Vaughan, M.A., Omar, A., Hu, Y., Powell, K.A., Liu, Z., Hunt, W.H., & Young, S.A. (2009). Overview of the CALIPSO Mission and CALIOP Data Processing Algorithms. *Journal of Atmospheric and Oceanic Technology, 26*, 2310-2323

Xie, C., Chen, P., Pan, D., Zhong, C., & Zhang, Z. (2021). Improved Filtering of ICESat-2 Lidar Data for Nearshore Bathymetry Estimation Using Sentinel-2 Imagery. *Remote Sensing, 13*, 4303

Zhang, S., & Chen, P. (2022). Subsurface phytoplankton vertical structure from lidar observation during SCS summer monsoon onset. *Optics Express, 30*, 17665-17679

Zhang, X., Boss, E., & Gray, D.J.J.O.e. (2014). Significance of scattering by oceanic particles at angles around 120 degree*, 22*, 31329-31336

Zhang, X., Gray, D.J., Huot, Y., You, Y., & Bi, L. (2012). Comparison of optically derived particle size distributions: scattering over the full angular range versus diffraction at near forward angles. *Applied optics, 51*, 5085

Zhong, C., Chen, P., & Pan, D. (2021). An Improved Adaptive Subsurface Phytoplankton Layer Detection Method for Ocean Lidar Data. *Remote Sensing, 13*, 3875