Consistency analysis of water diffuse attenuation between ICESat-2 and MODIS in Marginal Sea: A case study in China Sea

Zhenhua Zhang, Siqi Zhang, Michael J Behrenfeld, Cédric Jamet, Paolo Di Girolamo, Davide Dionisi, Yongxiang Hu, Xiaomei Lu, Yuliang Pan, Minzhe Luo, et al.

Remote Sensing of Environment 318 (2025) 114602

Contents lists available at ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse

Consistency analysis of water diffuse attenuation between ICESat-2 and MODIS in Marginal Sea: A case study in China Sea

Zhenhua Zhang [a], Siqi Zhang [b], Michael J. Behrenfeld [c], Cédric Jamet [d], Paolo Di Girolamo [e], Davide Dionisi [f], Yongxiang Hu [g], Xiaomei Lu [g], Yuliang Pan [a], Minzhe Luo [a], Haiqing Huang [a], Delu Pan [a], Peng Chen [a,b,*]

[a] Southern Marine Science and Engineering Guangdong Laboratory (Guangzhou), No. 1119, Haibin Rd., Nansha District, Guangzhou 511458, China
[b] State Key Laboratory of Satellite Ocean Environment Dynamics, Second Institute of Oceanography, Ministry of Natural Resources, 36 Bochubeilu, Hangzhou 310012, China
[c] Department of Botany and Plant Pathology, 2071 SW Campus Way, Oregon State University, Corvallis, OR, USA
[d] Université Littoral Côte d'Opale, CNRS, Université Lille, IRD, UMR 8187, Laboratoire d'Océanologie et de Géosciences (LOG), 62930 Wimereux, France
[e] Università della Basilicata, Ateneo Lucano Boulevard, 10 - 85100 Potenza, Italy
[f] Institute of Marine Sciences (ISMAR), Italian National Research Council (CNR), Rome - Tor Vergata, Italy
[g] NASA Langley Research Center, Hampton, VA 23681, USA

ARTICLE INFO

Editor name: Dr. Menghua Wang

Keywords:
Spaceborne lidar ICESat-2
Diffuse attenuation coefficient
Lidar attenuation coefficient
Multiple scattering correction
MODIS
Ocean optics
Ocean color

ABSTRACT

Recent studies highlight the application of deriving the attenuation coefficient from spaceborne photon-counting lidar ATLAS/ICESat-2 over open oceans on global scales. However, its performance in the more optically complex and variable environments of marginal seas, which are more susceptible to human activity, has not been validated yet. In this study, we present an in-depth analysis of the consistency between diffuse attenuation coefficient (K_d) detection from MODIS and ICESat-2 in China's Marginal Seas. Findings demonstrate that ICESat-2 possesses strong capabilities for the retrieval of the attenuation coefficient across differing aquatic environments. However, discrepancies exist between the lidar system attenuation coefficient obtained from ICESat-2 and the diffuse attenuation coefficient determined by MODIS, influenced by factors such as multiple scattering. Implementation of a novel multiple scattering correction model demonstrates a notable ability in significantly reducing the inconsistency. Validation with in-situ Biogeochemical Argo float measurements reveals an enhancement in the accuracy of lidar-derived diffuse attenuation coefficients upon correction, with the mean absolute percent difference between lidar-derived K_d and Argo-K_d decreasing from 26 % to 15.7 %. The multiple scattering model developed can bridge the gap between the passive and active remote sensing detection and improve the reliability of lidar-derived attenuation coefficients. Fusing these two missions will greatly improve ocean observation capabilities, providing unprecedented opportunities for precise and comprehensive assessment of marine light environments. This approach has broad implications for ocean science and the application of satellite remote sensing in environmental studies.

1. Introduction

When a light beam enters a water body, its intensity is diminished due to absorption and scattering by various materials such as suspended particles, phytoplankton, and colored dissolved organic matter. The result is a near-exponential decrease in downward irradiance with the increase in depth. The rate at which light intensity decreases with depth is quantified using the diffuse attenuation coefficient, K_d. As a key metric of oceanography, K_d represents the extent of light absorption or scattering within the depths of the water. It can be employed to ascertain the availability of light for marine organisms at different layers of the water column (Yentsch et al., 2002). Critical to oceanographic research, K_d provides data relevant to understanding various oceanic physical and biochemical processes such as upper ocean heat exchange (Stramska and Zuzewicz, 2013; Wu et al., 2007), phytoplankton photosynthesis (Loiselle et al., 2009), marine primary productivity (Bergamino et al.,

* Corresponding author at: Southern Marine Science and Engineering Guangdong Laboratory (Guangzhou), No. 1119, Haibin Rd., Nansha District, Guangzhou 511458, China.
E-mail address: chenp@sio.org.cn (P. Chen).

https://doi.org/10.1016/j.rse.2025.114602
Received 28 August 2024; Received in revised form 22 December 2024; Accepted 7 January 2025
Available online 11 January 2025

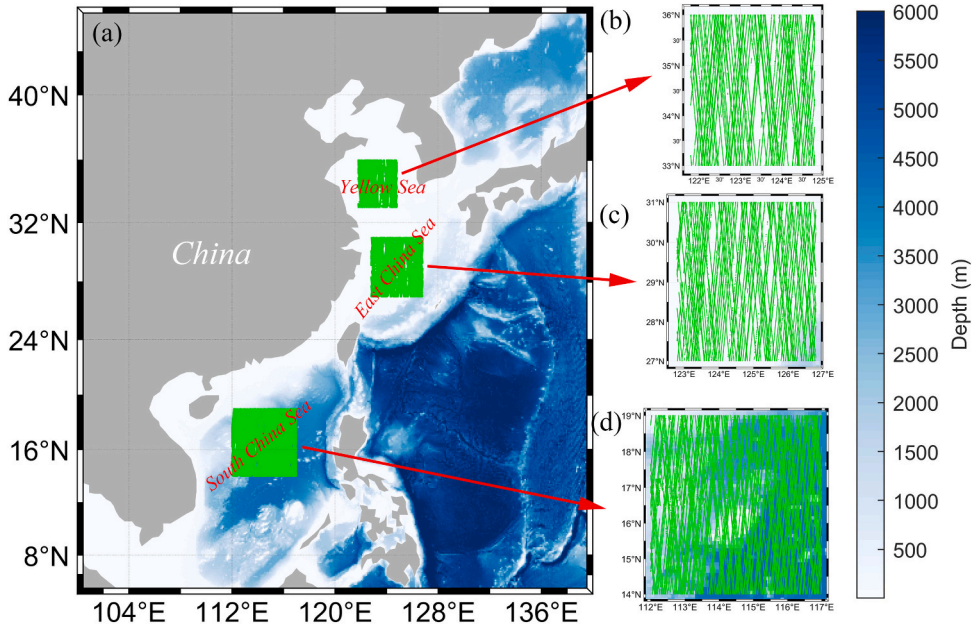

Fig. 1. Location of the study area, where the green lines indicate the ICESat-2 tracks. (a) The location of the study area; (b) (c) (d) are tracks in Yellow Sea, East China Sea, and South China Sea, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

2010; Lee et al., 2011), and water turbidity (Kirk, 1981).

Traditional methods of obtaining K_d involve in-situ measurements like Secchi depth measurements (Lee et al., 2015) and submarine photometry (Koenings and Edmundson, 1991). These methods, although known for their high accuracy and often used to validate other measurement methods (Dickey et al., 2006), are limited by poor efficiency. Advancements in satellite imaging technology, i.e. passive ocean color remote sensing, now empower researchers to conduct high-frequency observations of oceanic waters on a global scale (Lee et al., 2013; Rast et al., 1999; Wang et al., 2013; Wang et al., 2009). Currently, K_d has evolved into a standard product for passive ocean color satellites, having secured wide acceptance and validation (Jamet et al., 2012; Jiang et al., 2020; Lee et al., 2005b; Lee et al., 2005c; Morel et al., 2007; Tiwari and Shanmugam, 2014; Xing et al., 2020). For instance, instruments like the Medium Resolution Imaging Spectrometer (MERIS) of the European Space Agency, the Moderate Resolution Imaging Spectroradiometer (MODIS) of NASA, and the Visible Infrared Imaging Radiometer Suite (VIIRS) of NOAA, all serve up standard products of the diffuse attenuation coefficient K_d at 490 nm for the oceans. However, the inherent reliance of passive ocean color remote sensing on sunlight presents significant limitations. Issues arise as ocean color satellites cannot provide data during nighttime, and for polar high-latitude regions that face prolonged periods of sunlight deprivation (Behrenfeld et al., 2017). Further complications are introduced by persistent cloudy weather in these high-latitude regions, hampering the acquisition of ocean color data (Zhang et al., 2024). As an active detection technology, lidar can address the inherent limitations of passive remote sensing measurement and procure effective ocean observation data even through small gaps in the clouds. This technology proves to be an excellent addition to passive ocean color remote sensing, enabling continuous global ocean observation data, regardless of day or night, as well as accurate observational data for polar regions (Behrenfeld et al., 2017; Luthcke et al., 2021; Sun et al., 2023; Vadakke Chanat and Jamet, 2023; Zhang et al., 2023d).

The Cloud-Aerosol Lidar with Orthogonal Polarization (CALIOP), an integral part of the Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observation (CALIPSO) satellite, has opened up new avenues for ocean observation. Initially designed for cloud and aerosol detection, it has proven invaluable in recent years for oceanographic studies such as the interannual variability in polar phytoplankton biomass (Behrenfeld et al., 2022; Behrenfeld et al., 2017; Zhang et al., 2023d) and the global reserves of particulate carbon (Behrenfeld et al., 2013; Zhang et al., 2023c; Zhang et al., 2024). However, the primary shortcoming of CALIOP is its relatively coarse vertical resolution, making it challenging to use for water column profiling. NASA's latest Ice, Cloud, and land Elevation Satellite-2 (ICESat-2), equipped with the Advanced Topographic Laser Altimeter System (ATLAS), further expands the capability of passive ocean color satellites and CALIOP lidar satellites in ocean detection, making vertical profiling of the ocean possible (Lu et al., 2020).

In recent years, the field of oceanography has witnessed the advent of novel methodologies that involve the retrieval of water K_d and the backscattering coefficient of water particles using satellite lidar echo signals (Corcoran and Parrish, 2021; Dionisi et al., 2024; Eidam et al., 2024; Lu et al., 2021a; Zhang et al., 2022a) and linking them to water chlorophyll concentration through bio-optical models (Zheng et al., 2022). The lidar attenuation coefficient K_{lidar}, the primary parameter retrieved by lidar, reveals the degree to which the laser beam is absorbed or scattered within the water body (Gordon, 1982). ICESat-2 has already demonstrated its prowess in detecting ocean K_d at both the global and basin scales (Lu et al., 2023; Yang et al., 2023), where the open ocean's relative homogeneity allows for broad-scale application of its measurements. However, marginal seas present a unique set of challenges due to their high productivity and susceptibility to human activities (Wang et al., 2024; Yoo et al., 2019; Yu et al., 2019), which can lead to more variable and complex optical properties. These characteristics necessitate careful validation to ensure the accuracy and relevance of ICESat-2 measurements in these distinct environments. Here, we take the China's Marginal Sea as an example to explore the consistency of measurements between ICESat-2 and MODIS in nearshore waters. This study aims to assess the consistency between ICESat-2-derived K_{lidar} and MODIS-derived K_d in China's Marginal Seas and develop and validate a novel multiple scattering correction model to improve this consistency. The remainder of this paper is organized as follows: Section 2 describes the study area, data sources and details the methodology, including data processing, multiple scattering correction model development, and validation approaches. Results are presented in Section 3, followed by a discussion of implications and limitations in Section 4. Finally,

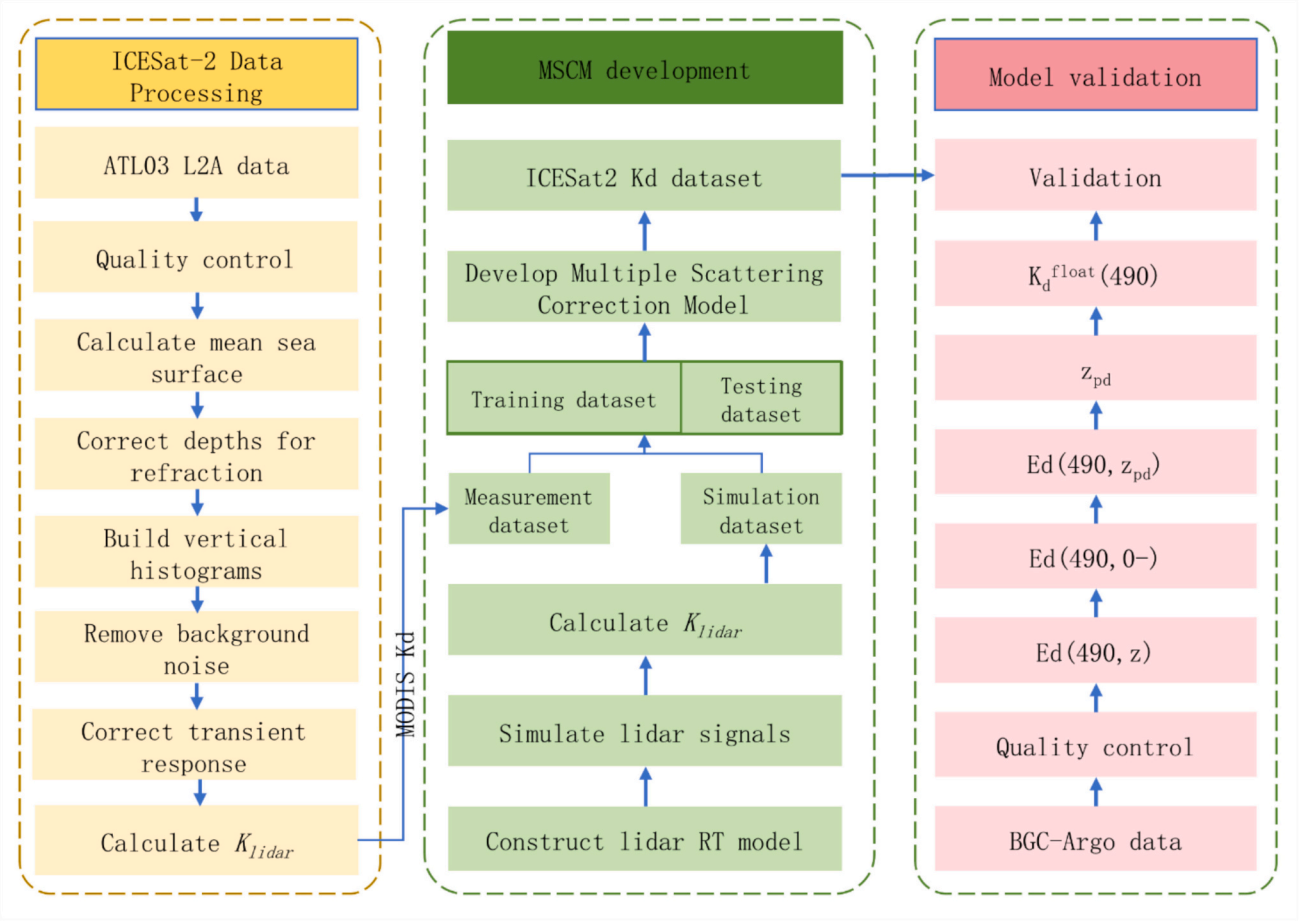

Fig. 2. Schematic flowchart of the methodological approach considered in this study.

conclusions and future research directions are provided in Section 5.

2. Materials and methods

2.1. Study area

The study focuses on the offshore regions of China, specifically the Yellow Sea, the East China Sea, and the South China Sea. As shown in Fig. 1, these areas sit on the peripheral edge of the western Pacific Ocean, with the Yellow Sea lying between China and the Korean Peninsula, the East China Sea situated between China and Japan, and the South China Sea bordered by China, Vietnam, and the Philippines. The diverse ecological make-up of our chosen study area ranges from turbid coastal waters to clear, open oceanic waters (Shi and Wang, 2012; Zhang et al., 2023a). According to the MODIS entire mission composite product AQUA_MODIS.20020704_20230930.L3m.CU.KD.Kd_490.4 km.nc, the average K_d at 490 nm ($K_d(490)$) within the study areas of the Yellow Sea and the East China Sea are 0.13 m^{-1} and 0.09 m^{-1}, respectively, indicating the prevalence of low turbid waters that are influenced by terrestrial sources. In contrast, the South China Sea, with an average $K_d(490)$ of 0.034 m^{-1}, falls under the category of open ocean type I waters (Chen et al., 2022), whose optical properties are primarily determined by phytoplankton and their associated materials.

2.2. Data

2.2.1. MODIS data

The study employs the MODIS-Aqua and MODIS-Terra Level 3 daily products from October 2018 to August 2023 as a reference to compare with the ICESat-2 inversion outcomes. To maintain consistency with previous research (Lu et al., 2023), two different K_d values derived from MODIS were used for comparison with K_{lidar}. Specifically, through a semi-analytical model (Eq. 1), the K_d at 531 nm ($K_d^M(531)$) is estimated using the absorption coefficient at 531 nm ($a(531)$) and the backscattering coefficient at 531 nm ($b_b(531)$) (Lee et al., 2002; Lee et al., 2013; Yang et al., 2023).

$$K_d^M(531) = (1 + 0.005\theta_s) \times a(531) + 4.18\left(1 - 0.52e^{-10.8a(531)}\right)b_b(531) \quad (1)$$

where, θ_s is the solar zenith angle of the detector.

In contrast, the MODIS $K_d^M(490)$ product, a standard distributed product calculated by the band ratio method (Eq. 2) (Austin and Petzold, 1981), is converted to the K_d at 532 nm $K_d^M(532)$ using Eq. 3 (Lu et al., 2016) based on the specific wavelength of the ICESat-2 lidar system, which operates at 532 nm. Moreover, with the aid of a bio-optical model (Eq. 4), the scattering coefficient at 531 nm ($b(531)$) is computed based on the MODIS chlorophyll product using an empirical relationship (Lu et al., 2023; Morel, 1991). MODIS data is obtained from the Ocean Color website (https://oceancolor.gsfc.nasa.gov/). Details of the coefficients in Eq. 2 are shown on the Ocean Color website (https://oceancolor.gsfc.nasa.gov/resources/atbd/kd/).

$$K_d^M(490) = 10^{a_0 + \sum_{i=1}^{4}\left(\log_{10}\left(\frac{R_{rs}(488)}{R_{rs}(547)}\right)\right)^i} + 0.0166 \quad (2)$$

$$K_d^M(532) = 0.68\left(K_d^M(490) - 0.022\right) + 0.054 \quad (3)$$

$$b(531) = 0.00226 + 0.31 \times Chl^{0.62} \quad (4)$$

2.2.2. ICESat-2 data

The Ice, Cloud, and land Elevation Satellite-2 (ICESat-2), launched by NASA in 2018, aims to measure the elevation of Earth's ice sheets, clouds, and land surfaces. Its main instrument is the Advanced Topographic Laser Altimeter System (ATLAS), a pioneering lidar system that employs single-photon counting technology. The core of this system is a laser that emits pulses at a wavelength of 532 nm with a pulse repetition frequency that reaches up to 10 kHz. This frequency allows for nearly continual along-track measurements with an impressive along-track horizontal resolution of close to 0.7 m. The laser beam, when it reaches the Earth's surface, has a spot diameter of approximately 11 m. The emitted laser pulses are split into six beams and, after reflecting off the surface, are detected by a highly sensitive photon-counting receiver, allowing for extremely precise measurement of the Earth's surface elevation data (Magruder and Brunt, 2018; Magruder et al., 2020). The zenith angle of the laser emission is close to 0. ICESat-2 boasts near-global surveillance with a repeat period of approximately 91 days, extending its coverage from 88°N to 88°S. In this study, we use the ICESat-2 geolocation L2A (Neumann et al., 2023) data from October 2018 to August 2023. The data is sourced from the National Snow and Ice Data Center (NSIDC, https://nsidc.org/).

2.2.3. BGC-Argo data

The significant advancements in buoy technology allow us to access the extensive in-situ dataset for global ocean optical and biogeochemical observations provided by Biogeochemical Argo (BGC-Argo) (Claustre et al., 2020; Johnson and Claustre, 2016; Wong et al., 2020; Xing et al., 2020). The onboard multispectral downwelling irradiance radiometer is able to measure downwelling irradiance (*Ed*) across multiple wavelengths, including 380, 412, and 490 nm (Begouen Demeaux and Boss, 2022). For the purpose of evaluating and validating satellite products, this study utilizes the *Ed* data specifically at 490 nm. The BGC-Argo data employed in this research includes the Argo float profiles incorporating downwelling irradiance collected over the period extending from 2018 to 2023. This data is accessible via the Argo Global Data Assembly Center (GDACs) at Ifremer (ftp://ftp.ifremer.fr/ifremer/argo/).

2.3. K_d estimation from ICESat-2 and BGC-Argo

2.3.1. ICESat-2 data processing

The processing of ICESat-2 is illustrated in Fig. 2. First, a data filtering process is applied, which utilizes only nighttime data with a range of 1–12 sea surface return photons for each laser pulse and an atmospheric aerosol optical thickness that is less than 0.2 (Lu et al., 2023). Following the identification of the sea surface position, a refraction range correction is applied to the underwater photons by subtracting the mean sea level from the heights and then multiplying by 0.75 (Zheng et al., 2022). These photons are subsequently accumulated along the track considering a sliding distance of 4 km and a cumulative distance of 20 km, to result in the pseudo-waveforms of the water column profile. The bin size of 20 km along-track distance is to ensure a sufficient number of photons are collected, thereby enhancing the signal-to-noise ratio (Yang et al., 2023). This step length of 4 km enables us to match the lidar-derived profiles with MODIS data more effectively, as it corresponds to the resolution of MODIS pixels. After removing background noise, the next step involves the deconvolution of these profile waveforms in order to eliminate the effect of the detector's afterpulses (Lu et al., 2021b). The background photon noise rate is calculated from the signal phonons in the atmospheric region above the sea surface (Eidam et al., 2024). Finally, the lidar system attenuation coefficient K_{lidar} is computed. This calculation is based on the exponential decay of the lidar return signal observed through the water column (Lu et al., 2023):

$$K_{lidar} = -0.5 \frac{d\, ln(S(z))}{dz} \quad (5)$$

where, $S(z)$ is the lidar pseudo-waveform signal at depth z after transient response correction. The calculation is obtained by a linear fit on the logarithmic return signal within 14 m (Lu et al., 2023).

2.3.2. BGC-Argo data processing

Before any calculation is performed, the buoy profile data undergoes a quality control process in line with Organelli's standards established in 2017 (Organelli et al., 2017). This preparatory step is crucial to remove the influence of dark current and cloud/spikes. The subsequent computation of the diffuse attenuation coefficient at a specific depth z proceeds as follows (Begouen Demeaux and Boss, 2022):

$$K_d^{float}(490, z) = \frac{1}{z} ln\left(\frac{E_d(490, 0^-)}{E_d(490, z)}\right) \quad (6)$$

Here, $E_d(490, 0^-)$ is defined as the downwelling irradiance just beneath the sea surface. Given that BGC-Argo does not provide irradiance at $z = 0^-$, a binomial fit of the data, applied within a 10-m range, is employed to extrapolate and thereby calculate $E_d(490, 0^-)$ (Xing et al., 2020).

The subsequent step determines the penetration depth at 490 nm (z_{pd}) as follows:

$$E_d(490, z_{pd}) = \frac{E_d(490, 0^-)}{e} \quad (7)$$

where z_{pd} is obtained using linear interpolation on the $ln(E_d(490))$ profile.

Finally, the average diffuse attenuation coefficient within the first optical depth, denoted as $K_d^{float}(490, z_{pd})$, is calculated from the Argo data as follows (Begouen Demeaux and Boss, 2022; Xing et al., 2020):

$$K_d^{float}(490) = K_d^{float}(490, z_{pd}) = \frac{1}{Z_{pd}} \quad (8)$$

2.4. Statistical parameters to evaluate ICESat-2 K_d estimates

To evaluate the accuracy and consistency of K_d estimates derived from ICESat-2, several statistical metrics are employed. These metrics allow for a quantitative comparison between ICESat-2 K_d estimates and reference data from MODIS and BGC-Argo, helping to assess the performance of ICESat-2 across different aquatic environments. This assessment makes use of several key statistical metrics: the Pearson Correlation Coefficient (R), the bias, the Root Mean Square Difference (RMSD), and the Mean Absolute Percentage Difference (MAPD):

Pearson Correlation Coefficient (R): The Pearson Correlation Coefficient measures the strength of the linear relationship between ICESat-2 K_d estimates and the reference data, which is calculated as:

$$R = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2 \sum_{i=1}^{n}(y_i - \overline{y})^2}} \quad (9)$$

Bias: Bias measures the average difference between the ICESat-2 K_d estimates and the reference data, indicating whether ICESat-2 tends to overestimate or underestimate K_d, which is calculated as:

$$bias = \frac{\sum_{i=1}^{n}(x_i - y_i)}{n} \quad (10)$$

Root Mean Square Difference (RMSD): RMSD quantifies the overall error in the K_d estimates by calculating the square root of the mean of squared differences between the ICESat-2 estimates and the reference data. Lower RMSD values indicate better performance:

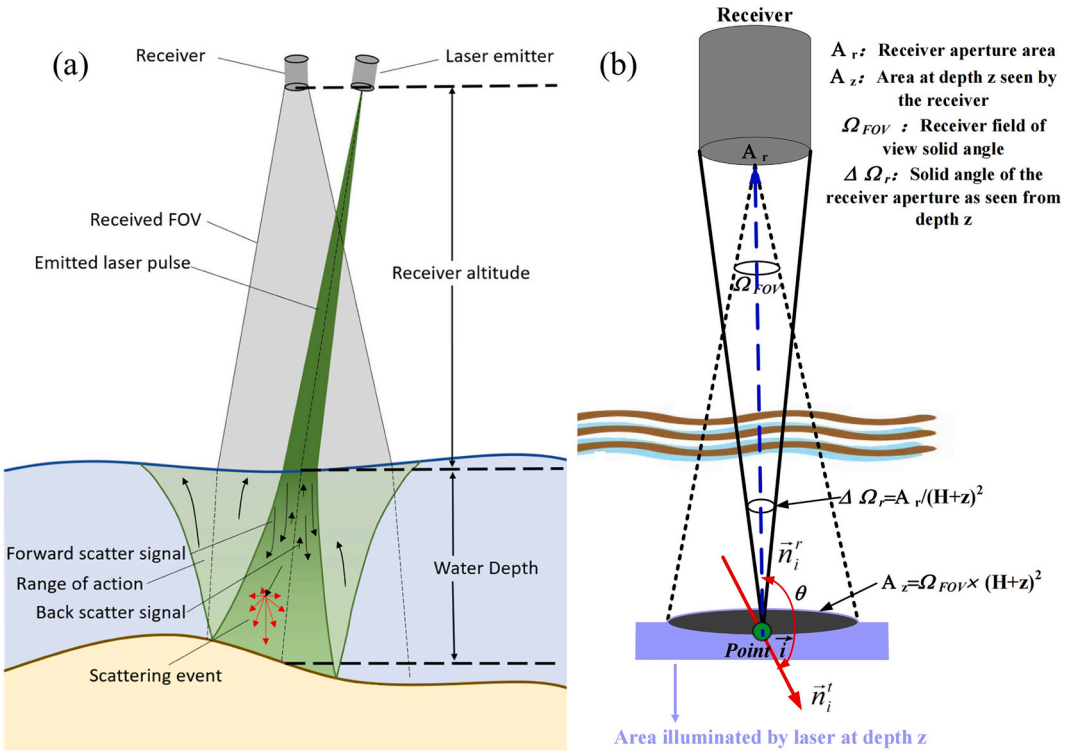

Fig. 3. The principle of the semi-analytical lidar radiative transfer model. (a) The complete procedure of this model encompasses laser emission, scattering, absorption, and reception. (b) The system optics for analytically estimating the probability of photon at point i collected by a remote receiver. $\vec{n}_i^t$ is the current photon travel direction at the layer index i; $\vec{n}_i^r$ is the photon receiving direction; and θ is the angle between the photon travel and receiving directions.

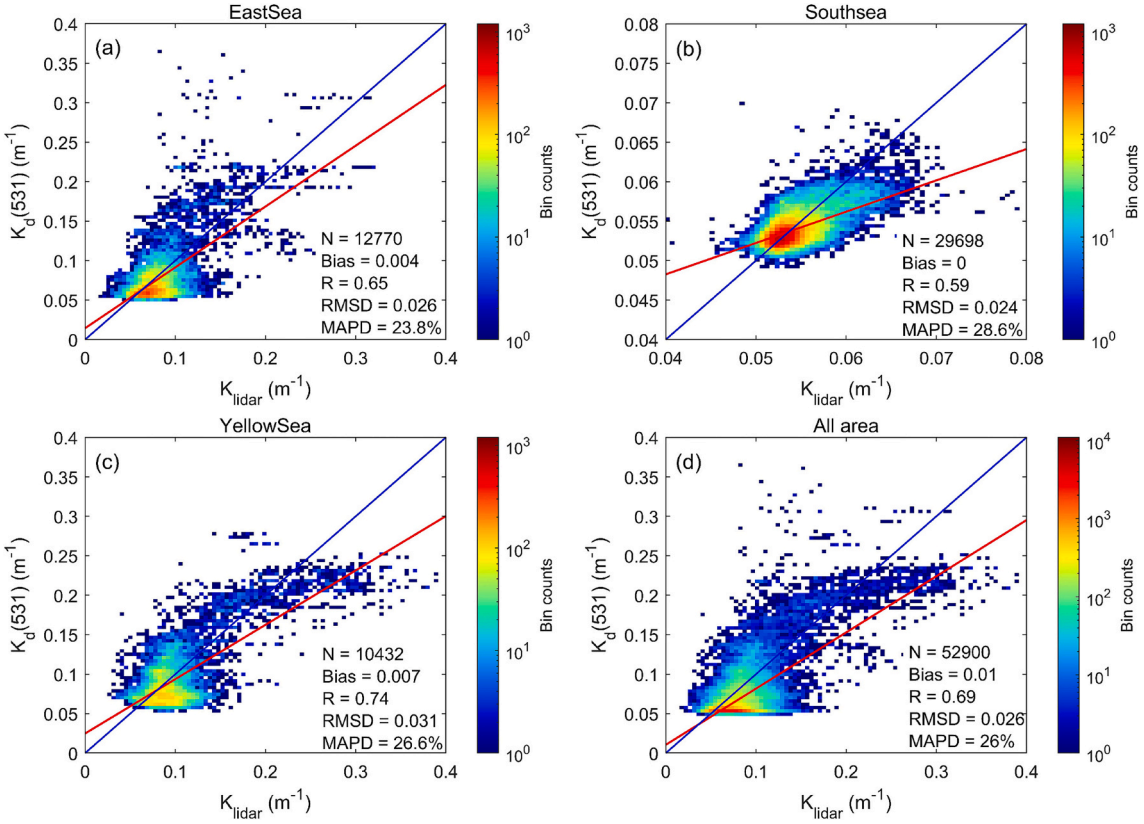

Fig. 4. Scatterplots of ICESat-2-derived K_{lidar} versus MODIS-derived $K_d(531)$ under different sea areas. Red solid lines denote the best-fit function, and blue solid lines denote the 1:1 line. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

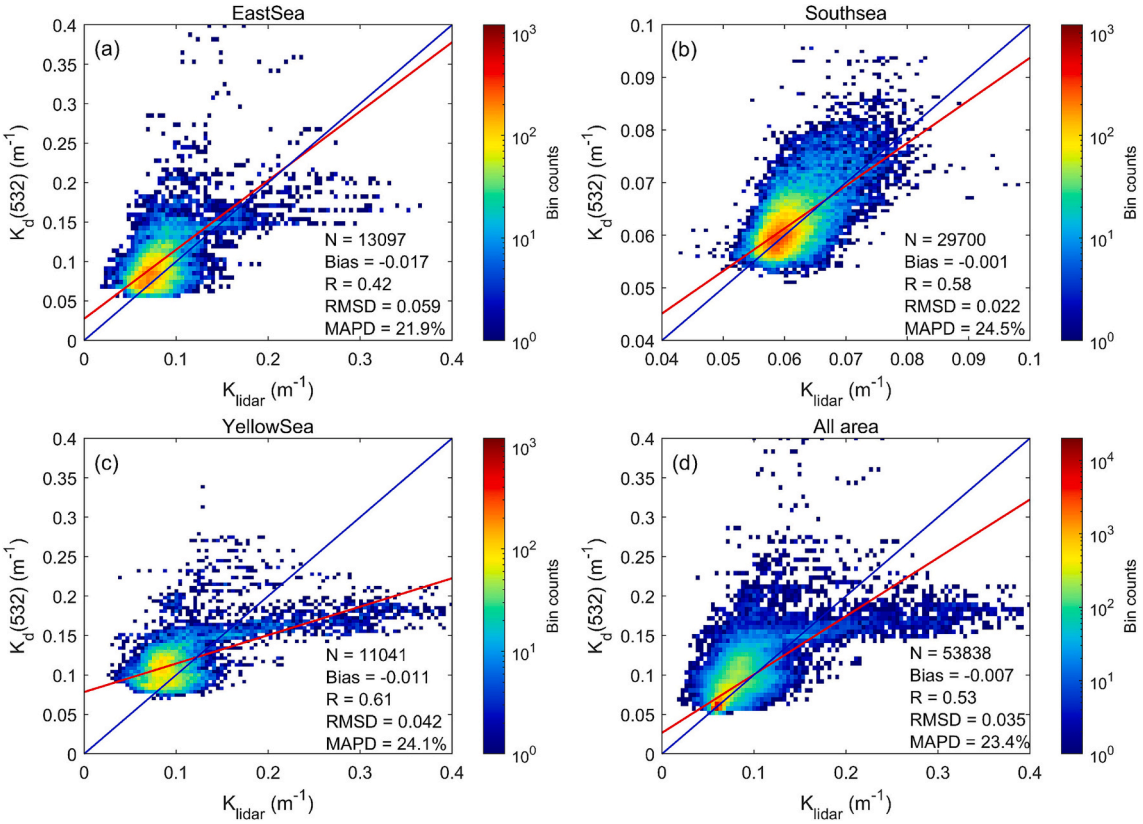

Fig. 5. Scatterplots of ICESat-2-derived K_{lidar} versus MODIS-derived $K_d(532)$ under different sea areas. Red solid lines denote the best-fit function, and blue solid lines denote the 1:1 line. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

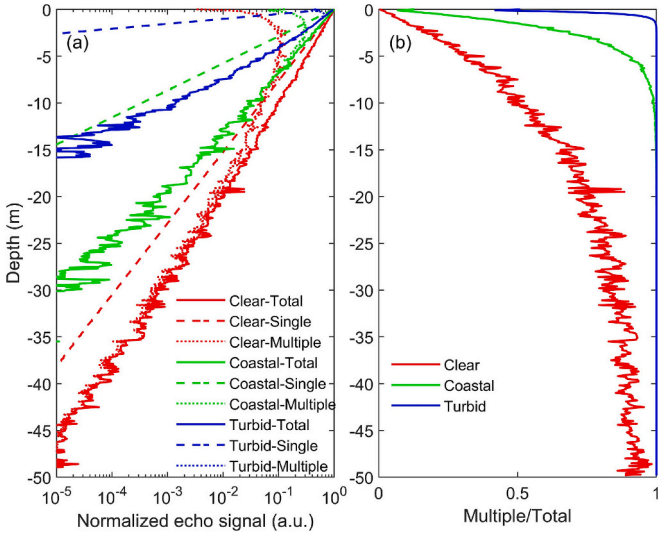

Fig. 6. Effect of multiple scattering on lidar echo signal. (a) Lidar echo signal simulation in different types of water. (b) The ratio of the multiple scattering signal to the total signal.

$$RMSD = \sqrt{\frac{\sum_{i=1}^{n}(x_i - y_i)^2}{n-1}} \tag{11}$$

Mean Absolute Percentage Difference (MAPD): MAPD measures the relative difference between the ICESat-2 estimates and the reference data as a percentage, providing insight into the proportional difference in the estimates:

$$MAPD = \frac{\sum_{i=1}^{n}\left|\frac{x_i - y_i}{y_i} \times 100\%\right|}{n} \tag{12}$$

where, x represents the ICESat-2 inversion results, and y represents the reference data.

2.5. Development of multiple scattering correction models

For lidar, multiple scattering refers to the phenomenon where the laser beam is scattered multiple times before being detected by the sensor. This can distort the measurements of K_d, particularly in turbid waters. To correct for this effect and improve the accuracy of ICESat-2 K_d estimates, two Multiple Scattering Correction Models (MSCM-531 and MSCM-532) are developed to accurately quantify the relationship between seawater K_d, K_{lidar}, and the other relevant lidar system parameters. The steps described below are repeated for both $K_d(531)$ and $K_d(532)$, as the two models aim to address the differences in their derivation methods and applications. These models are designed to correct for the effects of multiple scattering in the retrieval of K_{lidar} from ICESat-2 data, thereby enhancing the consistency with the diffuse attenuation coefficient. As illustrated in Fig. 2, the process of constructing the MSCMs ensues as follows:

1. Start with the construction of the lidar radiative transfer model (Wu et al., 2024). As shown in Fig. 3, the complete procedure of this model encompasses laser emission, scattering, and absorption. The lidar radiative transfer model used is a semi-analytical approach that efficiently simulates the multiple scattering of LiDAR signals in oceanic environments. This method integrates improved stochastic processes with analytical techniques to address the mathematical intricacies of light scattering, which conventional Monte Carlo models find challenging. The model uses the radiation transfer equation to simulate the

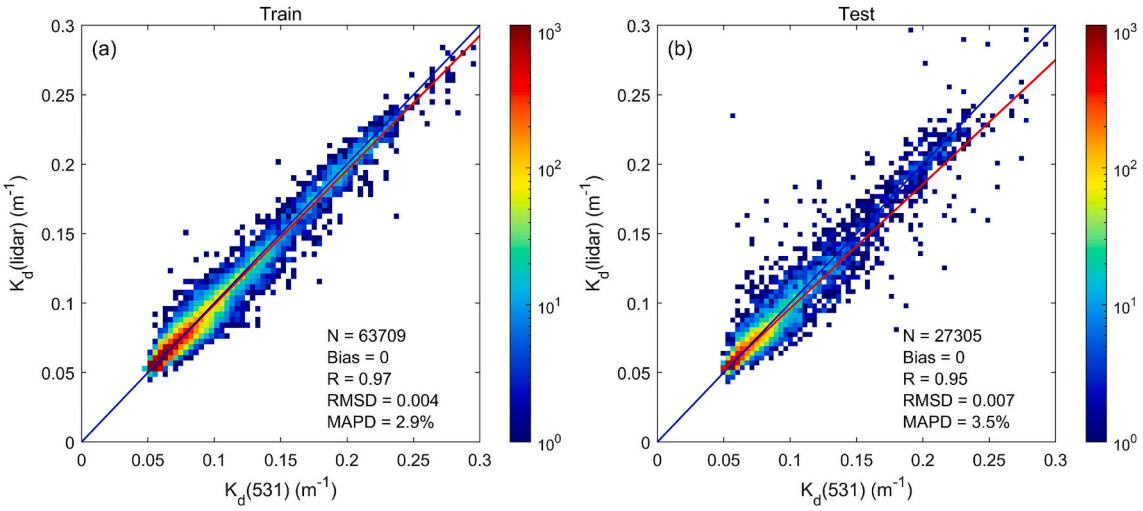

Fig. 7. Multiple scattering correction model MSCM-531 developed with MODIS-derived $K_d^M(531)$. Red solid lines denote the best-fit function, and blue solid lines denote the 1:1 line. The Y-axis represents the data after the correction. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

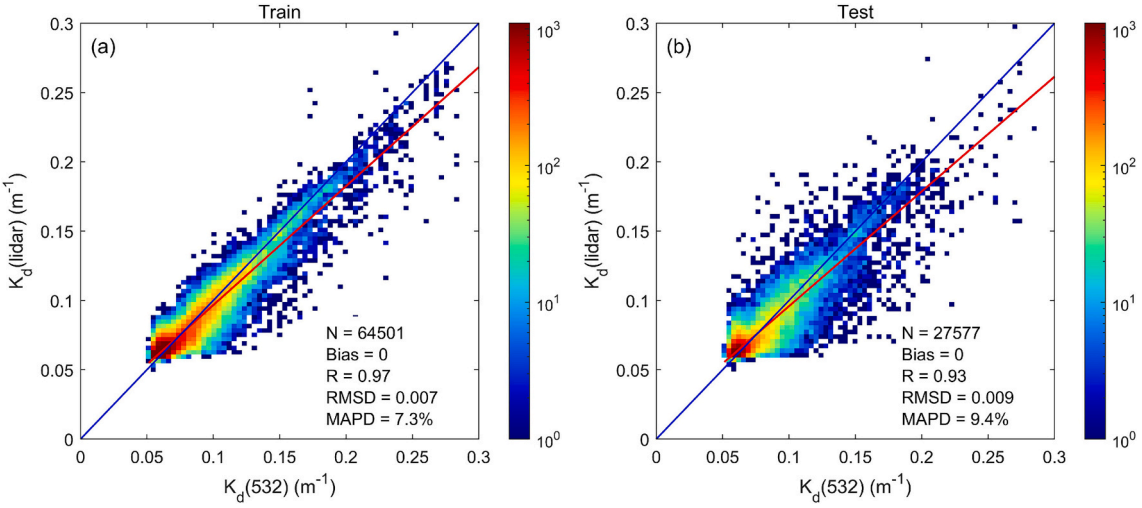

Fig. 8. Multiple scattering correction model MSCM-532 developed with MODIS-derived $K_d^M(532)$. Red solid lines denote the best-fit function, and blue solid lines denote the 1:1 line. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

interaction of laser light with the water, considering the light's propagation velocity, radiance, direction, position, and optical properties of the water. It incorporates a semi-analytical module that divides photon clusters into parts that either continue scattering or return to the receiver, enhancing computational efficiency by reducing the number of simulated photons. This approach not only offers significant time savings but also provides a more accurate representation of the laser's behavior in various underwater conditions. The semi-analytical Monte Carlo simulation is only briefly recalled here for completeness.

The simulation starts by setting parameters for the emitter, receiver, environment, and Monte Carlo model (e.g., photon count, weight threshold). A photon is emitted, with its initial angle and Stokes vector. In the Move phase, the photon's step length is calculated, and its new position is updated. If it exits the water, it's processed by the receiver. If still in water, scattering angles are randomly assigned based on the phase function. For polarized light, Stokes vector adjustments are made post-scattering. In semi-analytical mode, the photon is checked against the receiver's range for scattering properties and weight updates. In the Absorb phase, photon energy decays as it moves. Threshold screening and stochastic photon termination process decide if it "lives" for the next move. This cycle repeats for all photons until the simulation is completed. More details can be found in (Chen et al., 2021; Wu et al., 2024).

2. Perform simulations of echo signals under various conditions such as different seawater types and parameter configurations.

3. Retrieve K_{lidar} from the simulated signals.

4. Proceed with the creation of a matched dataset. This dataset is based on the retrieved K_{lidar}-K_d data and the measured ICESat-2 K_{lidar} and MODIS K_d data. For ICESat-2 K_{lidar} and MODIS K_d, the data are matched as they occur within the same day and fall within 4 km. MODIS K_d is utilized as a reference to correct K_{lidar} values, capitalizing on the extensive use in oceanographic research. By doing so, we aim to enhance the accuracy of K_{lidar} data and maintain consistency between the active lidar measurements and the well-established passive ocean color remote sensing data. Statistically, the matched dataset consisted of approximately 50,000 match points for ICESat-2 K_{lidar} and MODIS K_d data. To maintain consistency and ensure that our simulations reflect the actual data distribution, we used a quantity of simulated data that closely matches the actual number of matched points. This approach guarantees that the simulation data and the actual measured data have equal weight

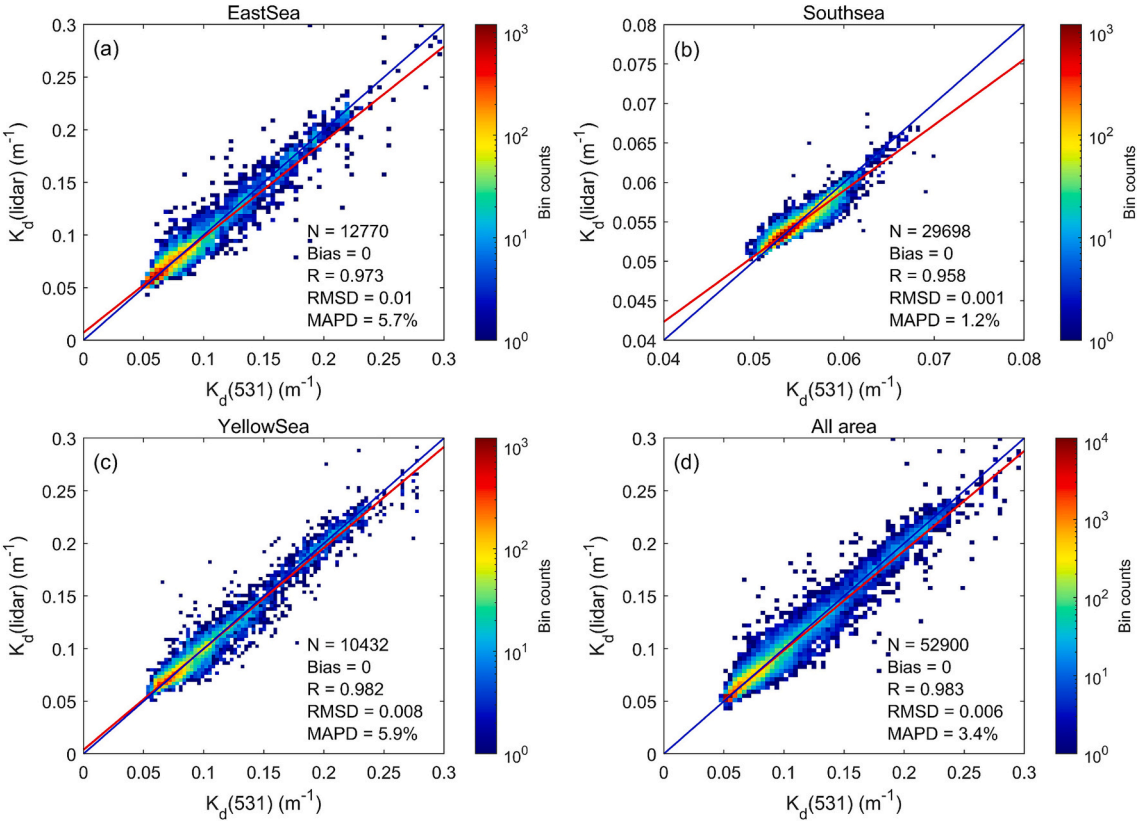

Fig. 9. Scatterplots of ICESat-2-derived K_d^L corrected by MSCM-531 developed with $K_d^M(531)$ versus MODIS-derived $K_d^M(531)$ under different sea areas. Red solid lines denote the best-fit function, and blue solid lines denote the 1:1 line. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

in our analysis, providing a balanced and representative dataset for the development and validation of our multiple scattering correction model.

5. Utilizing the matched dataset, formulate a data-driven empirical multiple scattering model where K_{lidar} and lidar system parameters are used as inputs and MODIS-derived K_d is considered as the reference value. The empirical multiple scattering model leverages the power of machine learning algorithms to identify patterns and relationships within the data. This process is accomplished using eXtreme Gradient Boosting (XGBoost), an efficient machine learning algorithm based on the Gradient Boosting framework (Chen and Guestrin, 2016). XGBoost optimizes conventional gradient boosting algorithms, leading to significant improvements in decision tree training speed and model performance. Importantly, it adds a regularization term to the loss function (Song et al., 2023), which is a key mechanism for preventing overfitting. Notably, XGBoost has found extensive application in fields like ocean color remote sensing, assisting with tasks like atmospheric correction, POC concentration retrieval (Liu et al., 2021), and phytoplankton population retrieval (Zhang et al., 2023b).

To ascertain the accuracy of the model, the data are initially divided into training and testing subsets through a random partitioning process, with 70 % allocated to the training set and the remaining 30 % designated for the testing set. This ratio ensures comprehensive training while still providing a substantial dataset for validation purposes. We start by building the model using the training data, and then apply it to the testing data for evaluation. The equation for this process is as follows:

$$K_d = f(K_{lidar}, FOV, DIV, H, \theta, a, b, b_b) \quad (13)$$

where FOV is the field of view, DIV is the angle of divergence, H is the height of lidar and θ is the nadir angle of the laser beam. The selection of these features is based on the contribution to multiple scattering accounting for both system parameters and water body parameters. (Wu et al., 2024). In the common matched dataset, the FOV ranges from 0.01 to 200 mrad, the DIV is between 0.001 and 0.1 mrad, the H is between 200 m and 600 km, and the θ is between 0 and 30 degrees, suitable for various types of platform-based LiDAR systems. This exploration is intentional, as it allows us to assess the model's robustness and generalizability under various potential sensor configurations. By testing the model's performance across this parameter space, we can ensure its applicability and accuracy in diverse real-world conditions and future sensor developments. Specifically, the ICESat-2 system parameters serve as a set of inputs to simulate the echo signals in various aquatic environments. The absorption coefficient a ranges from 0.05 and 0.4 m^{-1}, the scattering coefficient b is between 0.03 and 2 m^{-1}, and the backscattering coefficient b_b is between 0.001 and 0.05 m^{-1}. These ranges are informed by MODIS-detected data and are slightly broader than those typically reported in the previous studies (Chen et al., 2021; Gabriel et al., 2013) to ensure the inclusion of a typical variety of marine environments. To address the differences in performance and applicability of $K_d(531)$ and $K_d(532)$, we have developed two separate Multiple Scattering Correction Models (MSCM-531 and MSCM-532).

Lastly, we validate the accuracy of the MSCM model. This is achieved by processing ICESat-2 data using the MSCM and subsequently comparing the results with BGC-Argo data. Validation of the model was performed using BGC-Argo data collected from 2018 to 2023. The data are matched as they fall within 25 km^2 and 24 h. The matchup criteria of 24 h/25 km^2 is a deliberate attempt to synthesize the matching criteria used in studies involving MODIS and Argo (Werdell and Bailey, 2005), as well as the spaceborne lidar CALIOP and Argo data (Vadakke Chanat and Jamet, 2023). This criterion provides reasonable temporal proximity and accounts for the spatial extent of the lidar footprint, offering a balance that ensures a sufficient number of matchups for our analysis.

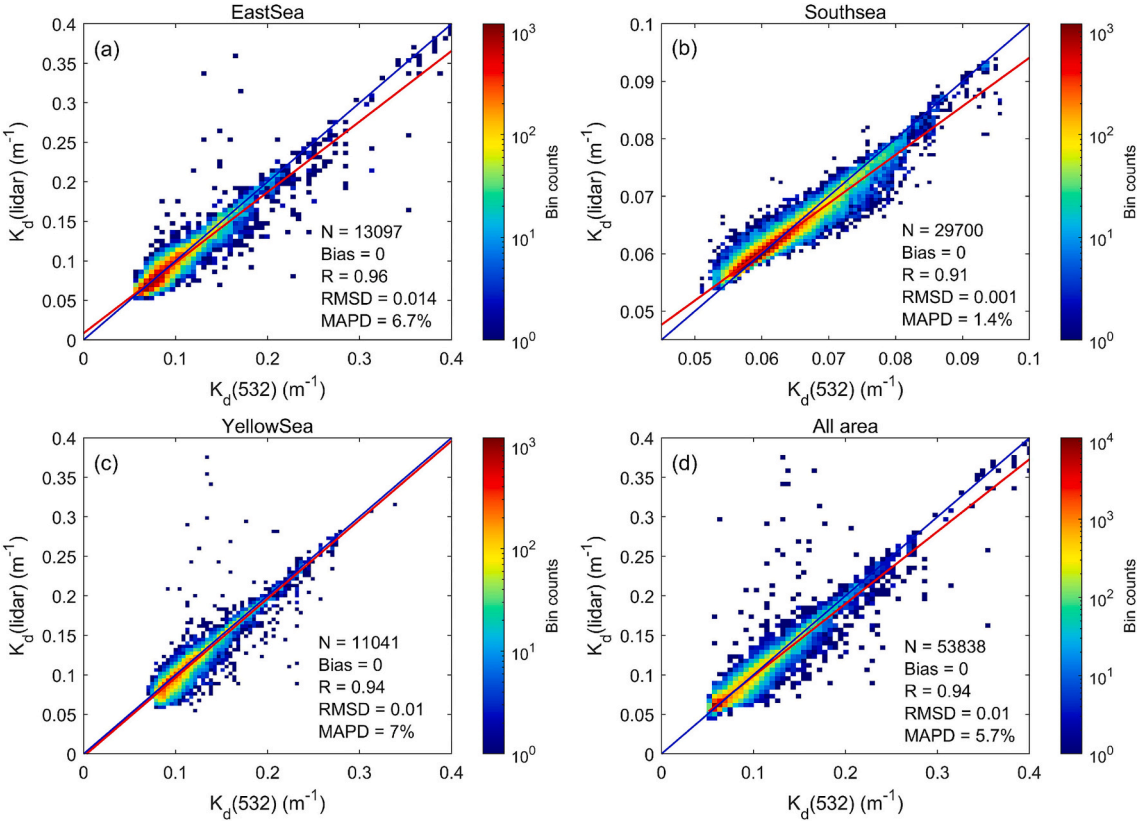

Fig. 10. Scatterplots of ICESat-2-derived K_d^L corrected by MSCM-532 developed with $K_d^M(532)$ versus MODIS-derived $K_d(532)$ under different sea areas. Red solid lines denote the best-fit function, and blue solid lines denote the 1:1 line. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

3. Results

3.1. K_d difference between MODIS and ICESat-2

Figs. 4 and 5 illustrate the comparative results of K_{lidar}, derived from ICESat-2 along China's coastal regions, with MODIS-derived $K_d^M(531)$ and $K_d^M(532)$ respectively. Fig. 4 presents the regression analysis results for K_{lidar} versus $K_d^M(531)$, focusing on the East China Sea, South China Sea, and Yellow Sea areas. Minor variations are observed across different water bodies when considering the ICESat-2 inversion results. Regarding correlation coefficients, the Yellow Sea, which is notably turbid, displays the highest correlation at a coefficient of 0.74. The coefficient for the clearer water in the South China Sea slightly drops to 0.59. For the MAPD, the East China Sea presents the lowest at 23.8 %, followed by the Yellow Sea at 26.6 %, while the South China Sea ranks the highest at 28.6 %. The RMSD reaches its minimum in the South China Sea at 0.024 m^{-1}, while it stands at 0.026 m^{-1} and 0.031 m^{-1} for the East China Sea and Yellow Sea, respectively. The overall results are depicted in Fig. 4(d), where K_{lidar} generally exceeds $K_d^M(531)$ (bias = 0.01 m^{-1}), and the correlation coefficient indicates a certain degree of correlation at 0.69. The RMSD registers at 0.026 m^{-1} and the MAPD at 26 %, signifying a discernible correlation amidst substantial differences. Fig. 5 presents the regression analysis for K_{lidar} and $K_d^M(532)$, displaying a higher disparity between K_{lidar} and $K_d^M(532)$ when compared to $K_d^M(531)$. The regression analysis for all data suggests a bias between K_{lidar} and $K_d^M(532)$ of −0.007 m^{-1}, an RMSD of 0.035 m^{-1}, and a correlation coefficient of 0.53 which, though indicative of some degree of correlation, is lower than that of $K_d^M(531)$. Contrarily, the MAPD at 23.4 % is slightly lower than that of $K_d^M(531)$.

3.2. Lidar multiple scattering analysis under different water types

The lidar echo signal simulation is conducted to analyze the effects of multiple scattering in various types of water including clear ocean, coastal water, and turbid harbor water. The simulation shown in Fig. 6 specifically utilizes the lidar system parameters of ICESat-2 (Martino et al., 2019; Zhang et al., 2022b). For the clear ocean water, the parameters are $a = 0.1140\ m^{-1}$, $b = 0.0370\ m^{-1}$, and $K_d = 0.1152\ m^{-1}$, with the echo signal depicted by the red line in the Fig. 6, and the K_{lidar} derived from the echo signal is 0.1410 m^{-1}. For the coastal water, the parameters are $a = 0.1790\ m^{-1}$, $b = 0.2190\ m^{-1}$, $K_d = 0.1925\ m^{-1}$, and $K_{lidar} = 0.2355\ m^{-1}$; for the highly turbid water, the parameters are $a = 0.3660\ m^{-1}$, $b = 1.8240\ m^{-1}$, $K_d = 0.4954\ m^{-1}$, and $K_{lidar} = 0.6655\ m^{-1}$. As observed from Fig. 6, at the initial entry of the laser into the water body, the echo signal is primarily dominated by single scattering, whereas as the laser propagates through the water, multiple scattering gradually increases. The impact of multiple scattering increases with the turbidity of the water body. In highly turbid water, the echo signal is mainly characterized by multiple scattering. Due to the effect of multiple scattering, the lidar system's attenuation coefficient lies between K_d and c (Montes et al., 2011). Thus, when inverting water body parameters, it is necessary to eliminate the impact of multiple scattering to obtain more accurate water body parameters.

3.3. K_d multiple scattering correction model construction

Utilizing XGBoost, K_{lidar} and system parameters are employed as inputs, with MODIS data serving as the reference values. The multiple scattering correction model is constructed using the gbtree algorithm, which is a gradient-boosting decision tree algorithm used within the XGBoost library (Chen and Guestrin, 2016). The early stopping

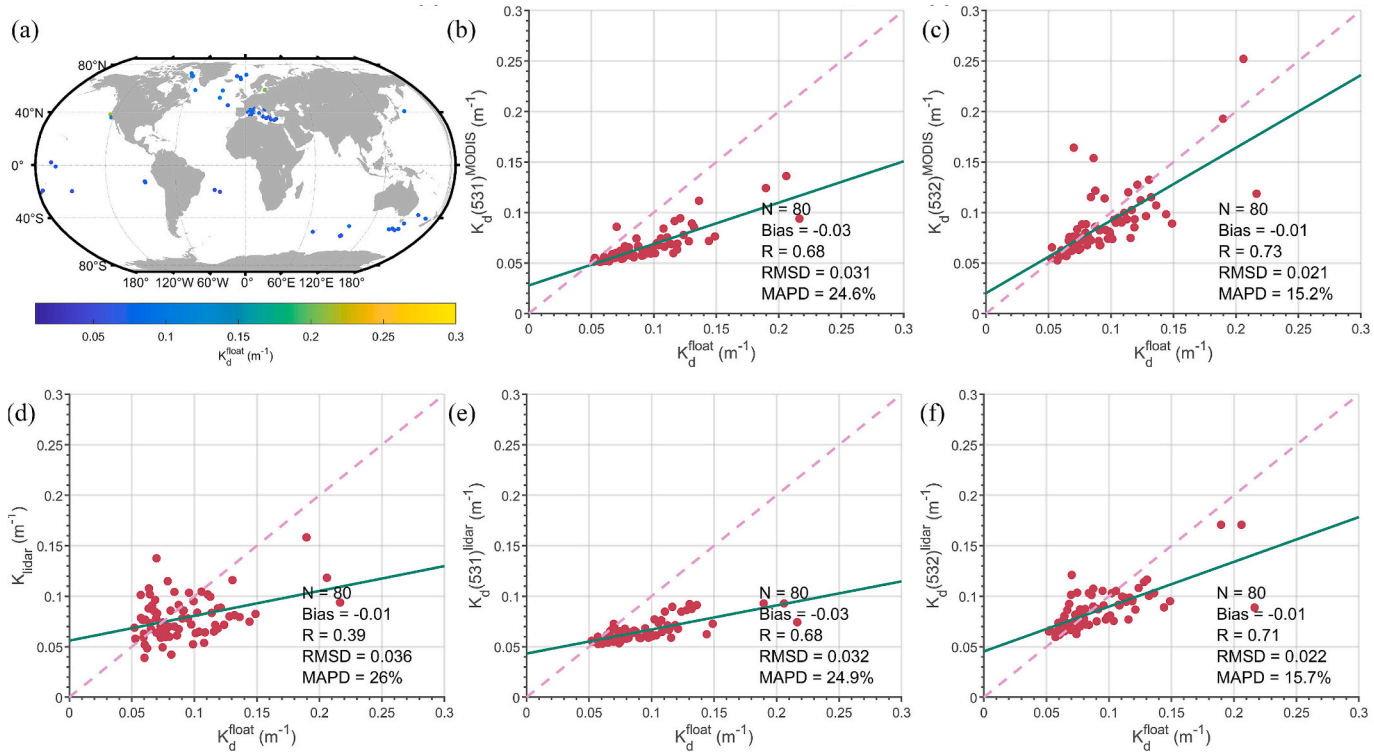

Fig. 11. Matched data between measured in-situ K_d^F and those derived from ICESat-2 and MODIS. (a) Distribution of matched data; (b) Scatterplots of Argo-derived K_d^F versus MODIS-derived $K_d^M(531)$; (c) Scatterplots of Argo-derived K_d^F versus MODIS-derived $K_d^M(532)$; (d) Scatterplots of Argo-derived K_d^F versus ICESat-2-derived K_{lidar}; (e) Scatterplots of Argo-derived K_d^F versus ICESat-2-derived $K_d^L(531)$ corrected by MSCM-531 developed with MODIS-derived $K_d^M(531)$; (f) Scatterplots of Argo-derived K_d^F versus ICESat-2-derived $K_d^L(532)$ corrected by MSCM-532 developed with MODIS-derived $K_d^M(532)$. Green solid lines denote the best-fit function, and pink dashed lines denote the 1:1 line. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

parameter is implemented to prevent overfitting. Through cross-validation, the optimal max_depth is determined. Subsequently, all data are randomly divided into a training set and a test set in a 7:3 ratio. Fig. 7 illustrates the training with $K_d^M(531)$ as the reference value, where the left and right plots represent the model's performance on the training and test sets, respectively. The model performs better on the training set than on the test set, indicating a slight tendency towards overfitting. However, the bias of the model on both the training and test sets is zero, suggesting that the model does not exhibit systematic bias. The correlation coefficients (R) are 0.97 and 0.95, respectively, indicating a strong correlation between MODIS-derived $K_d^M(531)$ and the ICESat-2-derived K_d^L processed by the MSCM-531 model. The RMSDs of training and testing data are 0.004 m^{-1} and 0.007 m^{-1}, respectively, indicating that the model's predictive error is relatively small, and the predicted results are very close to the reference values. The MAPDs are 2.9 % and 3.5 %, respectively, indicating that although the model's predictive accuracy on the test set is slightly lower than on the training set, it remains within a low error range. Overall, the multiple scattering correction model demonstrates good predictive performance and consistency on both training and test data. Furthermore, through the correction of the multiple scattering model, the discrepancy between K_{lidar} and $K_d^M(531)$ is significantly reduced.

The multiple scattering model MSCM-532 established based on MODIS-derived $K_d^M(532)$ is depicted in Fig. 8. The predictive model presented in Fig. 8 is applied across all regions, allowing us to assess the consistency and accuracy of predictions derived from our model in different marine settings. The accuracy and consistency of the model's performance on both the training and test sets are highlighted. In particular, the model boasts a correlation coefficient of 0.97 on the training set and 0.93 on the test set. The RMSDs are 0.007 m^{-1} and 0.009 m^{-1}, respectively, echoing the model's impressive predictive accuracy. The model's MAPD reveals a low deviation of 7.3 % in the training set and 9.4 % in the test set. These results suggest that, despite a certain amount of overfitting, the test set error manages to stay within a relatively low range. Overall, the application of the multiple scattering model substantially minimizes the discrepancy between K_{lidar} and $K_d^M(532)$.

3.4. Model performance in different areas

Figs. 9 and 10 showcase the regression analysis of the lidar-derived K_d^L, following a multiple scattering correction, versus the MODIS-derived K_d^M. Notably, when contrasted with Fig. 4, it's apparent that the discrepancy between the corrected K_d^{lidar} and $K_d(531)$ is markedly reduced. Specifically, the correlation coefficient in the East China Sea has seen an increase from 0.65 to 0.97. Similarly, the RMSD has dropped from 0.026 m^{-1} to 0.01 m^{-1}, while the MAPD has minimized from 23.8 % to 5.7 %. In regard to the South China Sea, the correlation coefficient has observed a rise from 0.59 to 0.96, while the RMSD has declined from 0.024 m^{-1} to 0.001 m^{-1}. As for the MAPD in the same region, it has decreased from 28.6 % to 1.2 %. Looking at the Yellow Sea, the correlation coefficient is seen to improve from 0.74 to 0.98. Concurrently, the RMSD is reduced from 0.031 m^{-1} to 0.008 m^{-1}, and the MAPD cut down from 26.6 % to 5.9 %. Taking into account the regions overall, the correlation coefficient is boosted from 0.69 to 0.98. Additionally, the RMSD is minimized from 0.026 m^{-1} to 0.006 m^{-1}, and the MAPD from 26 % shrinks to 3.4 %. The bias across different regions evens out at zero, indicating that the data corrected by the multiple scattering correction model does not possess bias.

Fig. 10 concisely illustrates the effectiveness of the multiple scattering correction model formulated using $K_d^M(532)$. When contrasted with Fig. 5, considerable enhancements are observed. In the East China

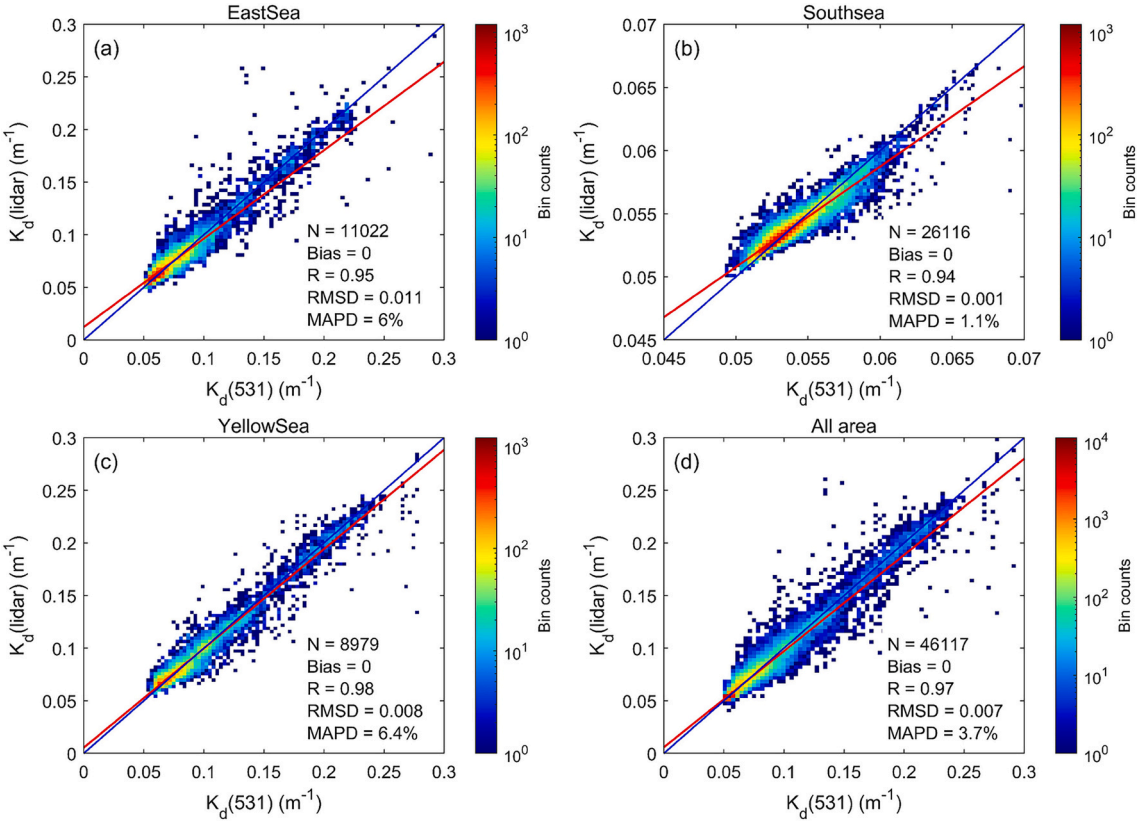

Fig. 12. Scatterplots of ICESat-2-derived K_d^L corrected by MSCM-531 versus MODIS-derived $K_d^M(531)$ with a cumulative distance of 9 km under different sea areas. Red solid lines denote the best-fit function, and blue solid lines denote the 1:1 line. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Sea, the correlation coefficient is seen to have risen from 0.42 to 0.96, indicating a strong linear relationship. Concurrently, the RMSD has decreased from 0.059 m^{-1} to 0.014 m^{-1}, while the MAPD has also witnessed a reduction from 21.9 % to 6.7 %. A similar trend is observed in the South China Sea, where the correlation coefficient has advanced from 0.58 to 0.91. The RMSD has dropped from 0.022 m^{-1} down to 0.001 m^{-1}, and the MAPD has been lowered from 24.5 % down to 1.4 %. Likewise, in the Yellow Sea, notable improvements are seen with the correlation coefficient climbing from 0.61 to 0.94, the RMSD reduced from 0.042 m^{-1} to 0.01 m^{-1}, and the MAPD diminished from 24.1 % to 7 %. The overall enhancements across all regions are significant, with the correlation coefficient improving from 0.53 to 0.94, RMSD seeing a reduction from 0.035 m^{-1} to 0.01 m^{-1}, and the MAPD witnessing a decrease from 23.4 % to 5.7 %. Crucially, across all regions, the model bias remains zero, indicating the model is unbiased.

3.5. Validation with in-situ measurement

The corresponding Argo float data that was successfully matched is depicted in Fig. 11(a). Across diverse global sea areas, we found 80 valid match-ups. Fig. 11(d) presents the K_{lidar} derived directly from the lidar inversion and the float K_d^F, with an RMSD of 0.036 m^{-1}, a MAPD of 26 %, and a correlation coefficient indicating a low correlation, only 0.39. The results following the application of the multiple scattering correction model are displayed in Figs. 11(e) and (f). Fig. 11(e) showcases the results of the multiple scattering model developed with $K_d^M(531)$, where the correlation coefficient R has increased to 0.68, indicating a higher correlation. The RMSD and MAPD have decreased to 0.032 m^{-1} and 24.9 %, respectively, and are consistent with the regression analysis results of the same position float data and MODIS-derived $K_d^M(531)$ (as shown in Fig. 11(b)). A comparison between MODIS-derived $K_d^M(532)$ and Argo data is presented in Fig. 11(c). with a correlation coefficient of 0.73, an RMSD of 0.021, and a MAPD of 15.2 %, revealing a high degree of consistency. These results indicate that the discrepancy between MODIS-derived $K_d^M(532)$ and Argo data is significantly smaller than that between MODIS-derived $K_d^M(531)$ and the same Argo data since MODIS-derived $K_d^M(532)$ and Argo data are both converted from $K_d(490)$, while $K_d^M(531)$ is calculated based on another method. Accordingly, the correction effect of the multiple scattering model established with MODIS-derived $K_d^M(532)$ as the reference data (as shown in Fig. 11(f)) is better than the model established with $K_d^M(531)$. The validation carried out with actual measurements reveals that the accuracy of the water body attenuation coefficient derived from the lidar, after model correction, improved significantly. Furthermore, MODIS-derived $K_d^M(532)$ exhibited a much smaller discrepancy than MODIS-derived $K_d^M(531)$.

4. Discussion

4.1. Influence of horizontal binning sizes

Horizontal binning sizes refer to the spatial distance along the satellite track over which lidar photons are accumulated to form a pseudo-waveform signal. These accumulation distances are crucial for improving the signal-to-noise ratio (SNR) (Zheng et al., 2022) in photon-counting lidar data and directly impact the accuracy of K_d estimates. To explore the impact of varying cumulative distances on the retrieval of the K_d by spaceborne photon-counting lidar, multiple scattering correction models were developed for 9 km and 4 km along-track accumulation distances. Using $K_d^M(531)$ as the reference, Figs. 12 and

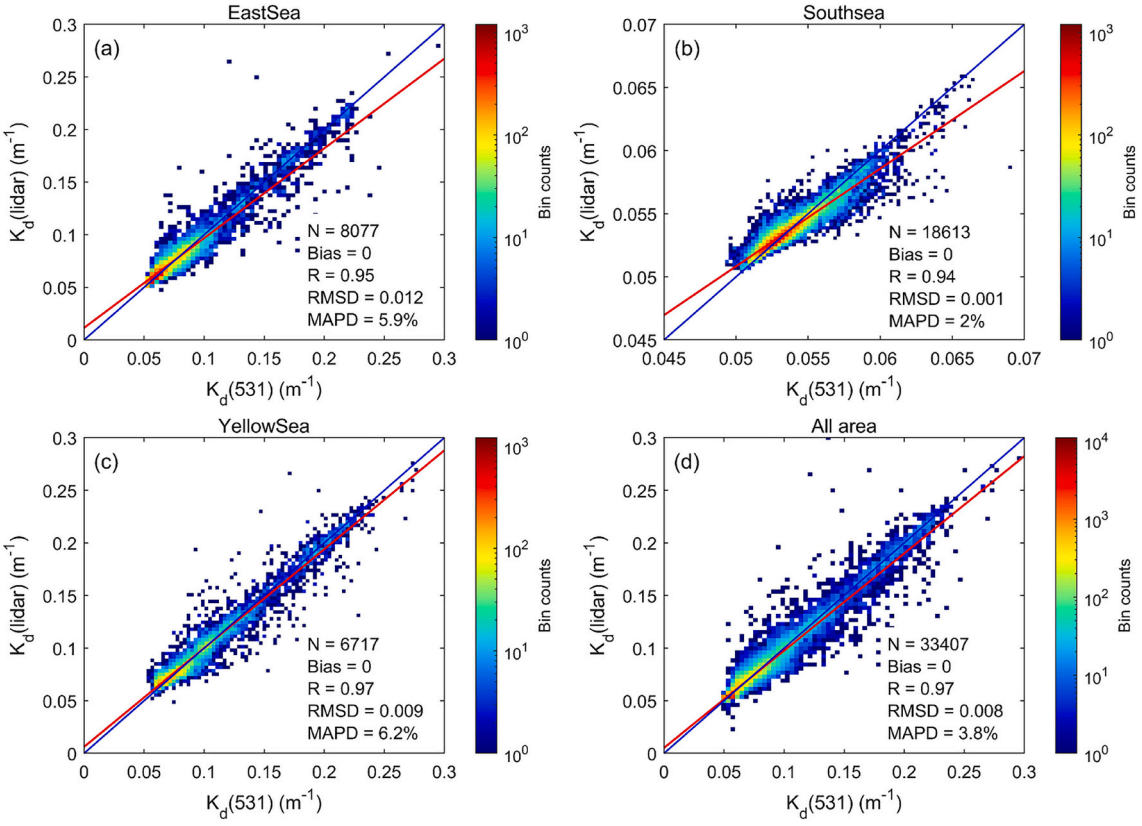

Fig. 13. Scatterplots of ICESat-2-derived K_d^L corrected by MSCM-531 versus MODIS-derived $K_d^M(531)$ with a cumulative distance of 4 km under different sea areas. Red solid lines denote the best-fit function, and blue solid lines denote the 1:1 line. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

13 respectively illustrate the correction results for the 9 km and 4 km models. When compared to the 20 km model (as depicted in Fig. 9), those statistical parameters exhibit minimal changes. This suggests that when utilizing $K_d^M(531)$ as the reference value, the multiple scattering correction model demonstrates strong robustness and consistency across varying cumulative distances. Using $K_d^M(532)$ as the reference, similar small changes of the correction results for the 9 km and 4 km models can be observed in Figs. 14 and 15. The consistency of the model remains largely unaffected, indicating that changes in cumulative distance have minimal bearing on the model's output and general performance.

4.2. Comparison with in-situ measurements and model validation

The multiple scattering correction model established with MODIS $K_d^M(532)$ as the reference value has the best effect. One critical factor in the discrepancies between the satellite-derived K_d (i.e. ICESat-2-derived K_d^L and MODIS-derived K_d^M) and in-situ K_d^F is the difference in the temporal sampling interval between the satellite-derived products and in-situ measurements. Another aspect to consider is the spatial scales of the measurements, with satellites capturing integrated conditions over time and area, while in-situ data reflects localized, instantaneous conditions. Additionally, the conversion from $K_d(490)$ to $K_d(532)$ introduces uncertainties due to wavelength-specific bio-optical variations estimated in the range of 10 % to 20 % (Lee et al., 2005a; Werdell and Bailey, 2005). In terms of different water bodies, as shown in Fig. 11e and f, when the K_d value is low, the matching points are concentrated near the 1:1 line. As the K_d value increases, indicating more turbid waters, the matching points gradually deviate from the 1:1 line. This suggests that the model performs very well in oligotrophic waters but experiences a certain degree of accuracy decline in turbid water bodies. In the future, with the expansion of the measured dataset, establishing models based on measured values as references will further enhance the model's accuracy. The ultimate goal is to develop a global model with extensive in-situ measurements. This global model will aim to provide a more comprehensive tool for understanding and predicting the optical properties of marine environments worldwide.

4.3. K_d inconsistencies between ICESat-2 and MODIS

The results clearly indicate that while ICESat-2 exhibits a strong ability to retrieve water diffuse attenuation coefficient, there is a significant inconsistency between the K_d estimated by ICESat-2 and that calculated value using MODIS data. In addition to the impact of multiple scattering, the inconsistency could also be due to several other factors: (1) The difference in the nature of measurements taken by MODIS and ICESat-2. Specifically, MODIS measures the average within a 4 km × 4 km pixel grid, which provides a broad overview of the water properties within that area. In contrast, ICESat-2 measures the water K_d within a more elongated area, specifically a 20 km × 11 m rectangular zone along its track. (2) The intricate nature of ICESat-2 data, particularly its nonlinear and recursive system response, can occasionally lead to divergent deconvolution effects for natural signals as highlighted by (Eidam et al., 2024). (3) Temporal resolution difference influences the comparison result as well, as ICESat-2 measurements may not fully align with the daily averages provided by MODIS. (4) Different zenith angles also contribute to the observed discrepancy. The solar zenith angles of the MODIS mainly range from 10° to 70°, whereas lidar operates at a nadir angle close to 0°. While a correction factor of 1.2 has been proposed in previous studies with an average zenith angle of 40° (Yang et al., 2023), it does not account for the effects of multiple scattering. Consequently, while differences in K_d values estimated by MODIS and ICESat-2 exist, the established multiple scattering model is crucial for

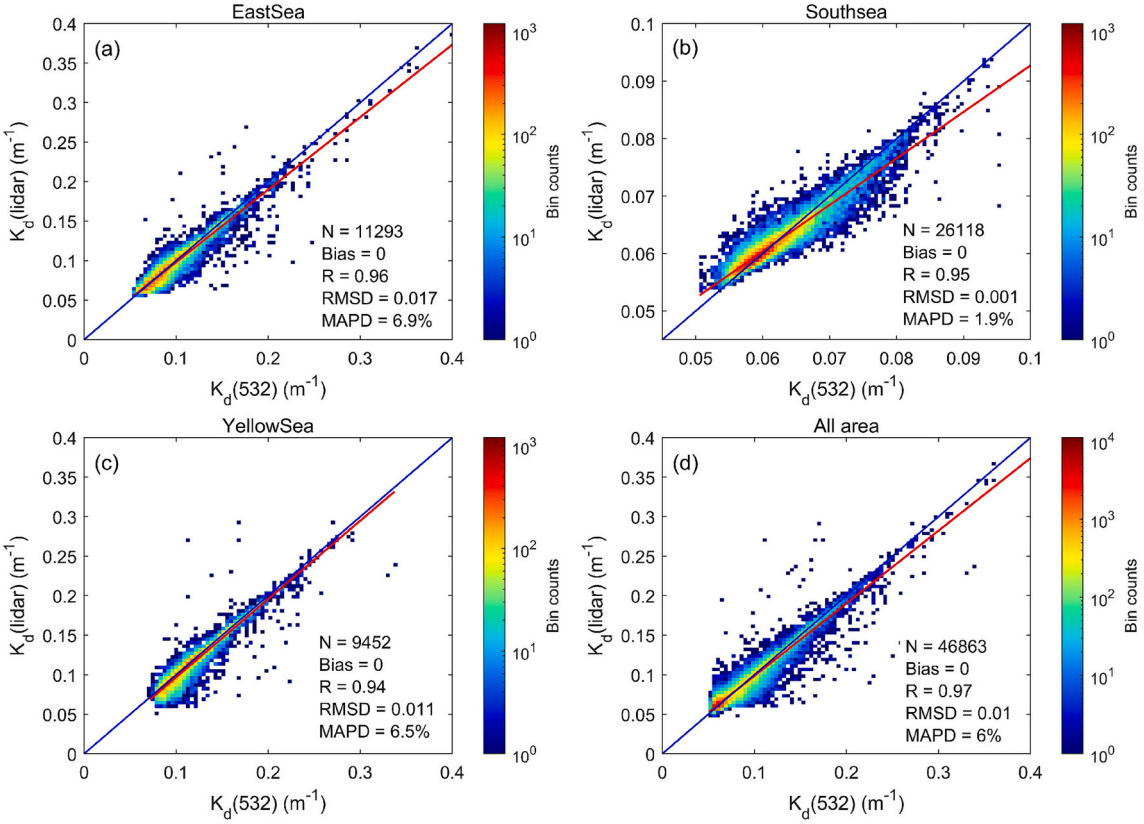

Fig. 14. Scatterplots of ICESat-2-derived K_d^L corrected by MSCM-532 versus MODIS-derived $K_d^M(532)$ with a cumulative distance of 9 km under different sea areas. Red solid lines denote the best-fit function, and blue solid lines denote the 1:1 line. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

eliminating or reducing this discrepancy. Besides, the disparity between K_{lidar} versus $K_d^M(531)$ and K_{lidar} versus $K_d^M(532)$ is mainly due to the distinct methodologies employed in deriving these values, (i.e., Eq. (1) versus Eq. (3)). Validation with in-situ measurement indicates MSCM-532 derived from $K_d^M(532)$ have a higher accuracy.

4.4. Uncertainty of MSCM

The uncertainty in the lidar K_d derived from the multiple scattering correction model predominantly originates from three principal sources: measurement (M), representation (R), and prediction (P) errors. The total uncertainty can be expressed as the square root of the sum of squares of these errors (Gregor and Gruber, 2021):

$$E = \sqrt{M^2 + R^2 + P^2} \tag{14}$$

The uncertainty in Eq. (14) is derived from three principal error sources: measurement, representation, and prediction. While these error sources have both random and systematic components, for the purpose of this study, we have made the simplifying assumption that the systematic errors are negligible due to the rigorous calibration processes of both the ICESat-2 (Luthcke et al., 2021) and MODIS (Xiong and Butler, 2020) instruments. This assumption allows us to treat the measurement errors as normally distributed with a mean bias of zero. Specifically, the measurement error (M) reflects potential biases (systematic errors) from sampling and measurement, as well as the random errors associated with these processes. Given the high level of calibration of the ICESat-2 and MODIS measurements, we consider it reasonable to assume that any systematic errors are effectively zeroed out, leading to an error distribution that is purely random in nature. The measurement errors primarily result from a combination of potential data deviations arising from the MODIS and ICESat-2 datasets. Since the construction of the MSCM is reliant on MODIS data, similar types of errors in this dataset can propagate through the model and influence the final output. Moreover, errors inherent in K_{lidar} also contribute to model uncertainty. Therefore, M is calculated as follows:

$$\frac{\Delta^2 K_d(lidar)}{K_d^2(lidar)} = \frac{\Delta^2 K_{lidar}}{K_{lidar}^2} + \frac{\Delta^2 K_d(MODIS)}{K_d^2(MODIS)} \tag{15}$$

In this equation, the uncertainty of K_{lidar} is less than 1 % (Lu et al., 2020), and the uncertainty of the MODIS-derived K_d product is 21.61 % (Zhao et al., 2018), hence, M is approximately bounded by 21.63 %.

Representation error R arises from gridded matching and is typically assumed to be normally distributed, with no bias at the global scale (Gregor and Gruber, 2021). The prediction error P, determined using validation data, is approximately around 10 %. Therefore, the total uncertainty of the MSCM is approximately estimated at 23.8 % by combining these components.

It is important to recognize that the MODIS $K_d^M(490)$ product and the scattering coefficient retrieval, which serves as the basis for our comparative analysis, are known to perform more effectively in oligotrophic type I waters as those models are derived for the clear open ocean. MODIS product has large uncertainty in the turbid water (Wang et al., 2009). This limitation could potentially impact the accuracy of our model, especially in more turbid waters. In the future, as more measurements become available, more coastal in-situ data will be used to improve the model's accuracy.

5. Summary and conclusion

This study aims to explore the potential of ICESat-2 in detecting the diffuse attenuation coefficient of waters in marginal seas. The conclusions are as follows: (1) ICESat-2 demonstrates a powerful capability for

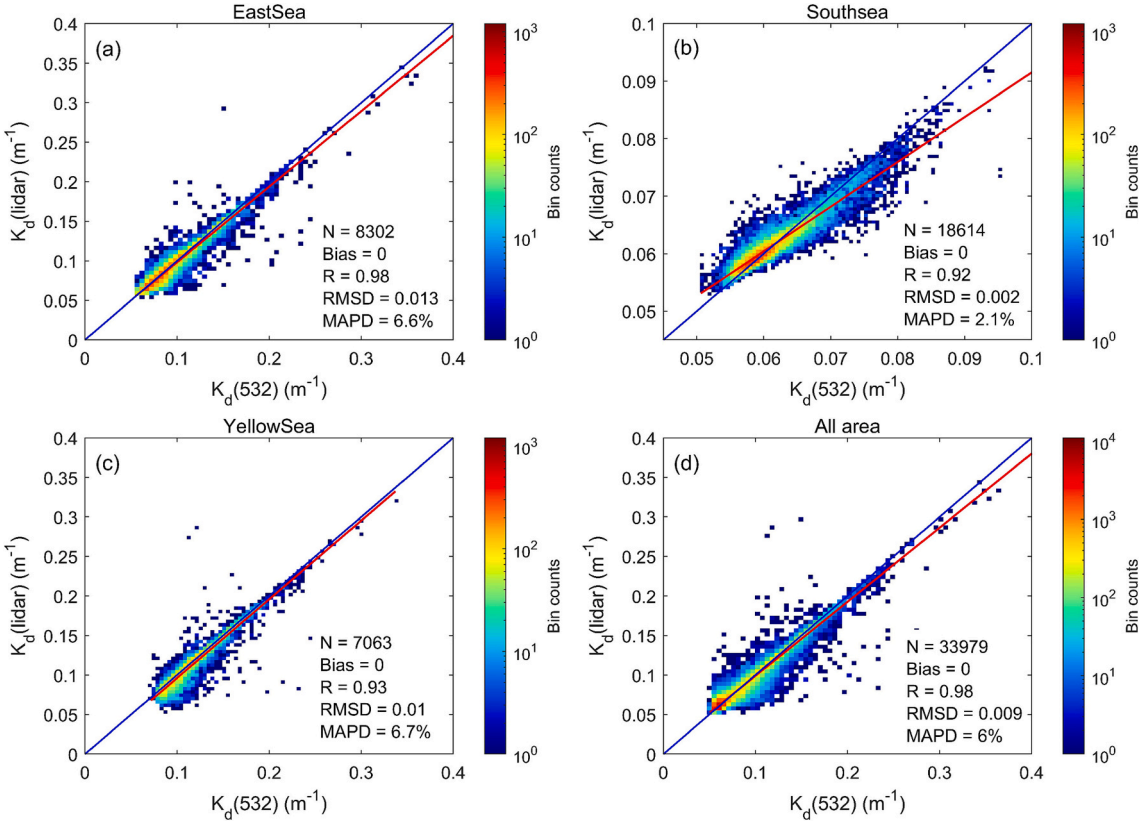

Fig. 15. Scatterplots of ICESat-2-derived K_d^L corrected by MSCM-532 versus MODIS-derived $K_d^M(532)$ with a cumulative distance of 4 km under different sea areas. Red solid lines denote the best-fit function, and blue solid lines denote the 1:1 line. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

retrieving the diffuse attenuation coefficient across a diverse range of aquatic environments.; (2) Due to the influence of multiple scattering and other factors, there is a certain discrepancy between the lidar system attenuation coefficient estimated by ICESat-2 and the diffuse attenuation coefficient of MODIS; (3) This discrepancy can be effectively eliminated or reduced through the construction of a multiple scattering correction model; (4) The multiple scattering model performs well across varied cumulative distances; (5) Validation with in-situ BGC-Argo measurements indicates that the accuracy of the lidar-derived diffuse attenuation coefficient is greatly improved after the application of the multiple scattering correction model. In summary, our research is the first to validate the capability of spaceborne photon-counting lidar in detecting marginal seas. The findings of this study carry significant implications for the characterization of oceanic optical properties in marginal seas, which are often subject to substantial human impact. The successful application of space-borne lidar data, as demonstrated through our analysis, offers a powerful tool for monitoring and assessing the complex dynamics of these vulnerable marine environments. The ability to accurately determine the diffuse attenuation coefficient is particularly crucial for understanding light availability, which is a key driver of photosynthesis and primary productivity in marine ecosystems. Moreover, the enhanced accuracy of the lidar-derived attenuation coefficient, achieved through our developed multiple scattering correction model, provides a more reliable basis for environmental studies and management decisions in coastal and nearshore areas. However, due to the lack of measured data, this study relies on MODIS data as the reference value to correct K_{lidar}, and the Argo measurements are only used for model validation. Consequently, any inherent uncertainty present in MODIS data would inevitably be propagated to the final correction results. Future research should collect sufficient in-situ measurements and use them for the correction of the multiple scattering model, which will further improve the reliability and accuracy of the results.

Funding

National Natural Science Foundation (42322606; 42276180; 61991453), China Postdoctoral Science Foundation (2023M740809), National Key Research and Development Program of China (2022YFB3901703; 2022YFB3902603), Key Special Project for Introduced Talents Team of Southern Marine Science and Engineering Guangdong Laboratory (GML2021GD0809), Donghai Laboratory Pre-research project (DH2022ZY0003), Key Research and Development Program of Zhejiang Province (2020C03100).

CRediT authorship contribution statement

Zhenhua Zhang: Writing – original draft. **Siqi Zhang:** Validation, Data curation. **Michael J. Behrenfeld:** Writing – review & editing, Investigation. **Cédric Jamet:** Writing – review & editing, Methodology. **Paolo Di Girolamo:** Writing – review & editing, Validation, Methodology. **Davide Dionisi:** Writing – review & editing. **Yongxiang Hu:** Writing – review & editing, Software. **Xiaomei Lu:** Writing – review & editing, Software. **Yuliang Pan:** Formal analysis. **Minzhe Luo:** Data curation. **Haiqing Huang:** Software, Formal analysis. **Delu Pan:** Supervision, Project administration, Funding acquisition. **Peng Chen:** Writing – review & editing, Writing – original draft, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence

the work reported in this paper.

Data availability

The MODIS data are available at https://oceancolor.gsfc.nasa.gov/. The ICESat-2 data are available at https://nsidc.org/data/atl03/. The BGC-Argo data are available at ftp://ftp.ifremer.fr/ifremer/argo/. The other data and code are available from the corresponding author on request.

Acknowledgments

The authors express sincere gratitude to the NASA National Snow and Ice Data Center (NSIDC) for providing the ICESat-2 data (https://doi.org/10.5067/ATLAS/ATL03.006). We gratefully acknowledge the NASA Ocean Biology Processing Group (OBPG) for maintaining and distributing MODIS products and biogeochemical-Argo data from the Argo Data Assembly Center. We thank GDAC for providing Argo data. These data were collected and made freely available by the International Argo Program and the national programs that contribute to it. (https://argo.ucsd.edu, https://www.ocean-ops.org). The Argo Program is part of the Global Ocean Observing System. We are appreciative of the anonymous reviewers for their insightful comments, which have notably enhanced the presentation of this paper.

References

Austin, R., Petzold, T.J., 1981. The determination of the diffuse attenuation coefficient of sea water using the Coastal Zone Color Scanner. In: Oceanography from Space, pp. 239–256.
Begouen Demeaux, C., Boss, E., 2022. Validation of remote-sensing algorithms for diffuse attenuation of downward irradiance using BGC-argo floats. Remote Sens. (Basel) 14, 4500.
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. Geophys. Res. Lett. 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. Nat. Geosci. 10, 118–122.
Behrenfeld, M.J., Hu, Y., Bisson, K.M., Lu, X., Westberry, T.K., 2022. Retrieval of ocean optical and plankton properties with the satellite Cloud-Aerosol Lidar with Orthogonal Polarization (CALIOP) sensor: background, data processing, and validation status. Remote Sens. Environ. 281, 113235.
Bergamino, N., Horion, S., Stenuite, S., Cornet, Y., Loiselle, S., Plisnier, P.D., Descy, J.P., 2010. Spatio-temporal dynamics of phytoplankton and primary production in Lake Tanganyika using a MODIS based bio-optical time series. Remote Sens. Environ. 114, 772–780.
Chen, T., Guestrin, C., 2016. XGBoost: A scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Association for Computing Machinery, San Francisco, California, USA, pp. 785–794.
Chen, P., Jamet, C., Mao, Z., Pan, D., 2021. OLE: a novel oceanic lidar emulator. IEEE Trans. Geosci. Remote Sens. 59, 9730–9744.
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 Trans. Geosci. Remote Sens. 60, 1–21.
Claustre, H., Johnson, K.S., Takeshita, Y., 2020. Observing the global ocean with biogeochemical-argo. Ann. Rev. Mar. Sci. 12, 23–48.
Corcoran, F., Parrish, C.E., 2021. Diffuse attenuation coefficient (Kd) from ICESat-2 ATLAS Spaceborne lidar using random-forest regression. Photogramm. Eng. Remote Sens. 87, 831–840.
Dickey, T., Lewis, M., Chang, G., 2006. Optical oceanography: recent advances and future directions using global remote sensing and in situ observations. Rev. Geophys. 44.
Dionisi, D., Bucci, S., Cesarini, C., Colella, S., D'Alimonte, D., Di Ciolo, L., Di Girolamo, P., Di Paolantonio, M., Franco, N., Gostinicchi, G., Giuliano, G., Kajiyama, T., Organelli, E., Santoleri, R., Liberti, G.L., 2024. Exploring the potential of Aeolus lidar mission for ocean color applications. Remote Sens. Environ. 313, 114341.
Eidam, E.F., Bisson, K., Wang, C., Walker, C., Gibbons, A., 2024. ICESat-2 and ocean particulates: a roadmap for calculating Kd from space-based lidar photon profiles. Remote Sens. Environ. 311, 114222.
Gabriel, C., Khalighi, M.A., Bourennane, S., Leon, P., Rigaud, V., 2013. Monte-Carlo-based channel characterization for underwater optical communication systems. IEEE/OSA J. Optical Commun. Network. 5, 1–12.
Gordon, H.R., 1982. Interpretation of airborne oceanic lidar: effects of multiple scattering. Appl. Optics 21, 2996–3001.
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 Syst. Sci. Data 13, 777–808.
Jamet, C., Loisel, H., Dessailly, D., 2012. Retrieval of the spectral diffuse attenuation coefficient K(λ) in open and coastal ocean waters using a neural network inversion. J. Geophys. Res. Oceans 117.
Jiang, L., Guo, X., Wang, L., Sathyendranath, S., Evers-King, H., Chen, Y., Li, B., 2020. Validation of MODIS ocean-colour products in the coastal waters of the Yellow Sea and East China Sea. Acta Oceanol. Sin. 39, 91–101.
Johnson, K., Claustre, H., 2016. Bringing biogeochemistry into the Argo age. In: Eos, Transactions American Geophysical Union.
Kirk, J., 1981. Monte Carlo study of the nature of the underwater light field in, and the relationships between optical properties of, turbid yellow waters. Mar. Freshw. Res. 32, 517–532.
Koenings, J.P., Edmundson, J.A., 1991. Secchi disk and photometer estimates of light regimes in Alaskan lakes: effects of yellow color and turbidity. Limnol. Oceanogr. 36, 91–105.
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. Appl. Optics 41, 5755–5772.
Lee, Z.-P., Darecki, M., Carder, K.L., Davis, C.O., Stramski, D., Rhea, W.J., 2005a. Diffuse attenuation coefficient of downwelling irradiance: an evaluation of remote sensing methods. J. Geophys. Res. Oceans 110.
Lee, Z., Du, K., Arnone, R., Liew, S.C., Penta, B., 2005b. Attenuation of Visible Solar Radiation in the Upper Water Column: A Model Based on IOPs.
Lee, Z., Du, K., Arnone, R., Liew, S.C., Penta, B., 2005c. Penetration of solar radiation in the upper ocean: a numerical model for oceanic and coastal waters. J. Geophys. Res. 110, 13.
Lee, Z., Lance, V.P., Shang, S., Vaillancourt, R., Freeman, S., Lubac, B., Hargreaves, B.R., Del Castillo, C., Miller, R., Twardowski, M., Wei, G., 2011. An assessment of optical properties and primary production derived from remote sensing in the Southern Ocean (SO GasEx). J. Geophys. Res. Oceans 116.
Lee, Z., Hu, C., Shang, S., Du, K., Lewis, M., Arnone, R., Brewin, R., 2013. Penetration of UV-visible solar radiation in the global oceans: insights from ocean color remote sensing. J. Geophys. Res. Oceans 118, 4241–4255.
Lee, Z., Shang, S., Hu, C., Du, K., Weidemann, A., Hou, W., Lin, J., Lin, G., 2015. Secchi disk depth: a new theory and mechanistic model for underwater visibility. Remote Sens. Environ. 169, 139–149.
Liu, H., Li, Q., Bai, Y., Yang, C., Wang, J., Zhou, Q., Hu, S., Shi, T., Liao, X., Wu, G., 2021. Improving satellite retrieval of oceanic particulate organic carbon concentrations using machine learning methods. Remote Sens. Environ. 256, 112316.
Loiselle, S.A., Bracchini, L., Cózar, A., Dattilo, A.M., Tognazzi, A., Rossi, C., 2009. Variability in photobleaching yields and their related impacts on optical conditions in subtropical lakes. J. Photochem. Photobiol. B Biol. 95, 129–137.
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., Yang, Y., Bontempi, P., Omar, A., Baize, R., 2020. Antarctic spring ice-edge blooms observed from space by ICESat-2. Remote Sens. Environ. 245, 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., 2021a. New ocean subsurface optical properties from space Lidars: CALIOP/CALIPSO and ATLAS/ICESat-2. Earth Space Sci. 8 e2021EA001839.
Lu, X., Hu, Y., Yang, Y., Vaughan, M., Palm, S., Trepte, C., Omar, A., Lucker, P., Baize, R., 2021b. Enabling value added scientific applications of ICESat-2 data with effective removal of afterpulses. Earth Space Sci. 8 e2021EA001729.
Lu, X., Hu, Y., Omar, A., Yang, Y., Vaughan, M., Lee, Z., Neumann, T., Trepte, C., Getzewich, B., 2023. Lidar attenuation coefficient in the global oceans: insights from ICESat-2 mission. Opt. Express 31, 29107–29118.
Luthcke, S.B., Thomas, T.C., Pennington, T.A., Rebold, T.W., Nicholas, J.B., Rowlands, D. D., Gardner, A.S., Bae, S., 2021. ICESat-2 pointing calibration and geolocation performance. Earth Space Sci. 8 e2020EA001494.
Magruder, L.A., Brunt, K.M., 2018. Performance analysis of airborne photon- counting lidar data in preparation for the ICESat-2 mission. IEEE Trans. Geosci. Remote Sens. 56, 2911–2918.
Magruder, L.A., Brunt, K.M., Alonzo, M., 2020. Early ICESat-2 on-orbit geolocation validation using ground-based corner cube retro-reflectors. Remote Sens. (Basel) 12, 3653.
Martino, A.J., Neumann, T.A., Kurtz, N.T., McLennan, D., 2019. ICESat-2 mission overview and early performance. In: Sensors, Systems, and Next-Generation Satellites XXIII. SPIE, pp. 68–77.
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. Optics 50, 2990–2999.
Morel, A., 1991. Light and marine photosynthesis: a spectral model with geochemical and climatological implications. Prog. Oceanogr. 26, 263–306.
Morel, A., Huot, Y., Gentili, B., Werdell, P.J., Hooker, S.B., Franz, B.A., 2007. Examining the consistency of products derived from various ocean color sensors in open ocean (case 1) waters in the perspective of a multi-sensor approach. Remote Sens. Environ. 111, 69–88.
Neumann, T., Brenner, A., Hancock, D., Robbins, J., Saba, J., Harbeck, K., Gibbons, A., Lee, J., Luthcke, S., Rebold, T., 2023. ATLAS/ICESat-2 L2A Global Geolocated Photon Data, Version 6. NASA National Snow and Ice Data Center Distributed Active Archive Center, Boulder, CO.

Organelli, E., Claustre, H., Bricaud, A., Barbieux, M., Uitz, J., D'Ortenzio, F., Dall'Olmo, G., 2017. Bio-optical anomalies in the world's oceans: an investigation on the diffuse attenuation coefficients for downward irradiance derived from Biogeochemical Argo float measurements. J. Geophys. Res. Oceans 122, 3543–3564.
Rast, M., Bezy, J.L., Bruzzi, S., 1999. The ESA Medium Resolution Imaging Spectrometer MERIS a review of the instrument and its mission. Int. J. Remote Sens. 20, 1681–1702.
Shi, W., Wang, M., 2012. Satellite views of the Bohai Sea, Yellow Sea, and East China Sea. Prog. Oceanogr. 104, 30–45.
Song, Z., He, X., Bai, Y., Dong, X., Wang, D., Li, T., Zhu, Q., Gong, F., 2023. Atmospheric correction of absorbing aerosols for satellite ocean color remote sensing over coastal waters. Remote Sens. Environ. 290, 113552.
Stramska, M., Zuzewicz, A., 2013. Influence of the parametrization of water optical properties on the modelled sea surface temperature in the Baltic Sea. Oceanologia 55, 53–76.
Sun, M., Chen, P., Zhang, Z., Zhong, C., Xie, C., Pan, D., 2023. Evaluation of the CALIPSO Lidar-observed particulate backscattering coefficient on different spatiotemporal matchup scales. Front. Mar. Sci. 10.
Tiwari, S.P., Shanmugam, P., 2014. A robust algorithm to determine diffuse attenuation coefficient of downwelling irradiance from satellite data in coastal oceanic waters. IEEE J. Select. Topics Appl. Earth Observ. Remote Sensing 7, 1616–1622.
Vadakke Chanat, S., Jamet, C., 2023. Validation protocol for the evaluation of space-borne lidar particulate back-scattering coefficient BBP. front. Remote sensing 4.
Wang, M., Son, S., Harding Jr., L.W., 2009. Retrieval of diffuse attenuation coefficient in the Chesapeake Bay and turbid ocean regions for satellite ocean color applications. J. Geophys. Res. Oceans 114.
Wang, M., Ahn, J.-H., Jiang, L., Shi, W., Son, S., Park, Y.-J., Ryu, J.-H., 2013. Ocean color products from the Korean geostationary ocean color imager (GOCI). Opt. Express 21, 3835–3849.
Wang, Z., Liu, S., Song, J., Wen, L., Yuan, H., Duan, L., He, Z., Li, X., 2024. Acidification state and interannual variability in marginal sea: a case study of the Bohai and the Yellow Seas surface waters in April 2023. Environ. Res. 259, 119536.
Werdell, P.J., Bailey, S.W., 2005. An improved in-situ bio-optical data set for ocean color algorithm development and satellite data product validation. Remote Sens. Environ. 98, 122–140.
Wong, A.P.S., Wijffels, S.E., Riser, S.C., Pouliquen, S., Hosoda, S., Roemmich, D., Gilson, J., Johnson, G.C., Martini, K., Murphy, D.J., Scanderbeg, M., Bhaskar, T.V.S. U., Buck, J.J.H., Merceur, F., Carval, T., Maze, G., Cabanes, C., André, X., Poffa, N., Yashayaev, I., Barker, P.M., Guinehut, S., Belbéoch, M., Ignaszewski, M., Baringer, M.O.N., Schmid, C., Lyman, J.M., McTaggart, K.E., Purkey, S.G., Zilberman, N., Alkire, M.B., Swift, D., Owens, W.B., Jayne, S.R., Hersh, C., Robbins, P., West-Mack, D., Bahr, F., Yoshida, S., Sutton, P.J.H., Cancouët, R., Coatanoan, C., Dobbler, D., Juan, A.G., Gourrion, J., Kolodziejczyk, N., Bernard, V., Bourlès, B., Claustre, H., D'Ortenzio, F., Le Reste, S., Le Traon, P.-Y., Rannou, J.-P., Saout-Grit, C., Speich, S., Thierry, V., Verbrugge, N., Angel-Benavides, I.M., Klein, B., Notarstefano, G., Poulain, P.-M., Vélez-Belchí, P., Suga, T., Ando, K., Iwasaska, N., Kobayashi, T., Masuda, S., Oka, E., Sato, K., Nakamura, T., Sato, K., Takatsuki, Y., Yoshida, T., Cowley, R., Lovell, J.L., Oke, P.R., van Wijk, E.M., Carse, F., Donnelly, M., Gould, W.J., Gowers, K., King, B.A., Loch, S.G., Mowat, M., Turton, J., Rama Rao, E.P., Ravichandran, M., Freeland, H.J., Gaboury, I., Gilbert, D., Greenan, B.J.W., Ouellet, M., Ross, T., Tran, A., Dong, M., Liu, Z., Xu, J., Kang, K., Jo, H., Kim, S.-D., Park, H.-M., 2020. Argo data 1999–2019: two million temperature-salinity profiles and subsurface velocity observations from a global array of profiling floats. Front. Mar. Sci. 7.
Wu, Y., Tang, C.C.L., Sathyendranath, S., Platt, T., 2007. The impact of bio-optical heating on the properties of the upper ocean: a sensitivity study using a 3-D circulation model for the Labrador Sea. Deep-Sea Res. II Top. Stud. Oceanogr. 54, 2630–2642.
Wu, D., Chen, P., Kong, W., Pan, D., 2024. A novel semi-analytical method for modeling polarized oceanic profiling LiDAR multiple scattering signals. IEEE Trans. Geosci. Remote Sens. 62, 1–17.
Xing, X., Boss, E., Zhang, J., Chai, F., 2020. Evaluation of ocean color remote sensing algorithms for diffuse attenuation coefficients and optical depths with data collected on BGC-Argo floats. Remote Sens. (Basel) 12, 2367.
Xiong, X., Butler, J.J., 2020. MODIS and VIIRS calibration history and future outlook. Remote Sens. (Basel) 12, 2523.
Yang, J., Zheng, H., Ma, Y., Zhao, P., Zhou, H., Li, S., Wang, X.H., 2023. Examining the consistency of lidar attenuation coefficient Klidar from ICESat-2 and diffuse attenuation coefficient Kd from MODIS. IEEE Geosci. Remote Sens. Lett. 20, 1–5.
Yentsch, C.S., Yentsch, C.M., Cullen, J.J., Lapointe, B., Phinney, D.A., Yentsch, S.W., 2002. Sunlight and water transparency: cornerstones in coral research. J. Exp. Mar. Biol. Ecol. 268, 171–183.
Yoo, S., Kong, C.E., Son, Y.B., Ishizaka, J., 2019. A critical re-assessment of the primary productivity of the Yellow Sea, East China Sea and Sea of Japan/East Sea Large Marine ecosystems. Deep-Sea Res. II Top. Stud. Oceanogr. 163, 6–15.
Yu, M., Eglinton, T.I., Haghipour, N., Montluçon, D.B., Wacker, L., Hou, P., Zhang, H., Zhao, M., 2019. Impacts of natural and human-induced hydrological variability on particulate organic carbon dynamics in the Yellow River. Environ. Sci. Technol. 53, 1119–1129.
Zhang, X., Ma, Y., Li, Z., Zhang, J., 2022a. Satellite derived bathymetry based on ICESat-2 diffuse attenuation signal without prior information. Int. J. Appl. Earth Obs. Geoinf. 113, 102993.
Zhang, Z., Chen, P., Mao, Z., 2022b. SOLS: an open-source spaceborne oceanic lidar simulator. Remote Sens. (Basel) 14, 1849.
Zhang, X., Li, C., Zhou, W., Zheng, Y., Cao, W., Liu, C., Xu, Z., Yang, Y., Yang, Z., Chen, F., 2023a. Study of the profile distribution of the diffuse attenuation coefficient and Secchi Disk depth in the Northwestern South China Sea. In: Remote Sensing.
Zhang, Y., Shen, F., Sun, X., Tan, K., 2023b. Marine big data-driven ensemble learning for estimating global phytoplankton group composition over two decades (1997–2020). Remote Sens. Environ. 294, 113596.
Zhang, Z., Chen, P., Jamet, C., Dionisi, D., Hu, Y., Lu, X., Pan, D., 2023c. Retrieving bbp and POC from CALIOP: a deep neural network approach. Remote Sens. Environ. 287, 113482.
Zhang, Z., Chen, P., Zhong, C., Xie, C., Sun, M., Zhang, S., Chen, S., Wu, D., 2023d. Chlorophyll and POC in polar regions derived from spaceborne lidar. Front. Mar. Sci. 10.
Zhang, Z., Zhang, S., Behrenfeld, M.J., Chen, P., Jamet, C., Di Girolamo, P., Dionisi, D., Hu, Y., Lu, X., Pan, Y., Luo, M., Huang, H., Pan, D., 2024. Combining deep learning with physical parameters in POC and PIC inversion from spaceborne lidar CALIOP. ISPRS J. Photogramm. Remote Sens. 212, 193–211.
Zhao, W., Cao, W., Hu, s., & Wang, G., 2018. Comparison of diffuse attenuation coefficient of downwelling irradiance products derived from MODIS-Aqua in the South China Sea. Opt. Precis. Eng. 26, 14–24.
Zheng, H., Ma, Y., Huang, J., Yang, J., Su, D., Yang, F., Wang, X.H., 2022. Deriving vertical profiles of chlorophyll-a concentration in the upper layer of seawaters using ICESat-2 photon-counting lidar. Opt. Express 30, 33320–33336.