Research Article | Vol. 31, No. 24 / 20 Nov 2023 / Optics Express 40557
Optics EXPRESS

Informing ocean color inversion products by seeding with ancillary observations

KM Bisson,[1,2,*] PJ Werdell,[3] AP Chase,[4] SJ Kramer,[5] BB Cael,[6] E Boss,[7] LIW McKinna,[8] and MJ Behrenfeld[2]

[1]Ocean Biology and Biogeochemistry Program, NASA Headquarters, Earth Science Division, Washington, D.C., USA
[2]Department of Botany and Plant Pathology, Oregon State University, Corvallis, OR, USA
[3]Ocean Ecology Laboratory, NASA Goddard Flight Center, Greenbelt, MD, USA
[4]Applied Physics Laboratory, University of Washington, Seattle, WA, USA
[5]Monterey Bay Aquarium Research Institute, Moss Landing, CA, USA
[6]National Oceanography Centre, Southampton, UK
[7]School of Marine Sciences, University of Maine, Orono, ME, USA
[8]Go2Q Pty Ltd., Sunshine Coast, QLD, Australia
*bissonk@oregonstate.edu

Abstract: Ocean reflectance inversion algorithms provide many products used in ecological and biogeochemical models. While a number of different inversion approaches exist, they all use only spectral remote-sensing reflectances ($R_{rs}(\lambda)$) as input to derive inherent optical properties (IOPs) in optically deep oceanic waters. However, information content in $R_{rs}(\lambda)$ is limited, so spectral inversion algorithms may benefit from additional inputs. Here, we test the simplest possible case of ingesting optical data ('seeding') within an inversion scheme (the Generalized Inherent Optical Property algorithm framework default configuration (GIOP-DC)) with both simulated and satellite datasets of an independently known or estimated IOP, the particulate backscattering coefficient at 532 nm ($b_{bp}(532)$). We find that the seeded-inversion absorption products are substantially different and more accurate than those generated by the standard implementation. On global scales, seasonal patterns in seeded-inversion absorption products vary by more than 50% compared to absorption from the GIOP-DC. This study proposes one framework in which to consider the next generation of ocean color inversion schemes by highlighting the possibility of adding information collected with an independent sensor.

1. Introduction

Since the launch of the NASA Coastal Zone Color Scanner in 1978, ocean color inversion algorithms have been used to link the primary product available from ocean color satellite instruments, spectral remote-sensing reflectances ($R_{rs}(\lambda)$; sr^{-1}), with biogeochemically relevant information (e.g., [1]). Derived inherent optical properties (IOPs) are used to quantify phytoplankton biomass and metrics of community structure [2–4], concentrations of particulate organic and inorganic carbon [5–6], and metrics of detrital and dissolved material [7–8]. However, there are limits to the information provided from $R_{rs}(\lambda)$ alone [9–11]. The addition of ancillary information (including prior information, [12]) within an inversion scheme invites new pathways for improving the utility of such approaches and has been historically underutilized in the ocean color community. Indeed, the use of ancillary data to inform inversions has improved aerosol models for decades [13–15] and, more recently, independent environmental and hydrographic information has been used to improve satellite algorithms to retrieve chlorophyll-a concentrations [16], IOPs [17]. and to detect the presence of diatoms [18] and *Synechococcus* [19].

Spectral $R_{rs}(\lambda)$ inversion algorithms solve for spectral absorption and backscattering. It follows, then, that substituting either unknown parameter (or one of its subcomponents) with a known

#503496 https://doi.org/10.1364/OE.503496
Journal © 2023 Received 21 Aug 2023; revised 27 Oct 2023; accepted 7 Nov 2023; published 15 Nov 2023

Optics EXPRESS

quantity of high quality could improve the other. Recently, we performed an evaluation of satellite retrieved particulate backscattering ($b_{bp}(\lambda)$; m^{-1}) from two satellite instruments that both fly in the A-train constellation: the Moderate Resolution Imaging Spectrometer (MODIS onboard the Aqua spacecraft) and the Cloud-Aerosol Lidar with Orthogonal Polarization (CALIOP) on the Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observations (CALIPSO) mission [20,21]. MODIS-Aqua is a passive radiometer designed specifically for ocean applications, whereas CALIOP is a polarized lidar optimized for atmospheric studies. MODIS-Aqua observes large swaths of the ocean and provides global coverage every ~2 days, whereas CALIOP produces lines of data with a footprint of ~ 100 m and a global repeat cycle of 16 days [22]. Despite shortcomings of CALIOP's construction and coverage, CALIOP $b_{bp}(532)$ was found to outperform MODIS-Aqua $b_{bp}(532)$ in a pointwise comparison with in situ Argo float data [20], as well as in a follow-on study assessing different matchup scales [23]. MODIS-Aqua $b_{bp}(532)$ data were also found to exhibit a seasonal bias when compared with Argo floats [21]. This bias is not observed in CALIOP $b_{bp}(532)$ data and was ultimately proposed to be due to a seasonal bias in ocean color $R_{rs}(\lambda)$ [21]. Compared to MODIS-Aqua, CALIOP $b_{bp}(532)$ is retrieved from a constant viewing angle and has substantially less atmospheric concerns in its retrieval [22], which may contribute to its higher performance.

In the future, we envision realization of a satellite-borne ocean optimized lidar (e.g., [24–25]), which will provide even better $b_{bp}(\lambda)$ coverage and accuracy than CALIOP. In preparation for such data, we consider here how active and passive satellites might be used in tandem to advance our understanding of ocean ecosystems. At the time of writing, the ideal dataset for validating inversion-based global products from merged ocean color and lidar data does not exist because, unlike backscattering sensors, arrays of absorption meters are not yet widely deployed in situ. Several validation schemes were considered for this study but found to be premature due to the lack of global co-located observations across the two satellite sensors and in situ data. Nevertheless, we can still strategically test different datasets across theoretical and observational cases to anticipate the magnitude of improvement that might be expected from such approaches in the future.

Here, we offer two case studies that use simulated and satellite datasets of IOPs and $R_{rs}(\lambda)$ to evaluate the inclusion (or 'seeding') of independently measured b_{bp} (single wavelength) within a commonly-used inversion algorithm, namely the Generalized Inherent Optical Property (GIOP) algorithm framework [26], which is suited to optically deep waters. The GIOP was also chosen because it is a flexible framework that can be configured for future enhancements, including Bayesian approaches [12], which is a natural extension to the work herein. The seeding approach, in practice, can either reduce the number of free parameters to be retrieved from an inversion algorithm or can offer an opportunity to replace one free parameter with another variable of interest (Fig. 1). We use simulated data constructed for inversion algorithm development [27], as well as $b_{bp}(532)$ data from the Ice, Cloud, and land Elevation Satellite (ICESat-2) and CALIPSO missions as input to the seeded-inversion scheme. Our study adds to a growing body of knowledge on how simple adjustments to ocean reflectance inversion schemes can significantly affect outcomes (e.g., [26,28–31]), which in turn may enable insightful biogeochemical and ecological interpretations in the future.

In essence, our analysis represents an exploration of an alternative inversion parameterization to retrieve the total spectral absorption coefficient ($a(\lambda)$; m^{-1}). Extracting information about absorbing constituents in seawater from ocean color remains challenging and yet is critical for quantifying phytoplankton abundance, composition, and physiology (e.g., [32–38,9]). If consideration of independent $b_{bp}(\lambda)$ datasets shows promise for seeding ocean color inversions, alternative retrievals of absorption coefficients may be formulated. If these alternative absorption coefficients are of improved quality relative to retrievals from current algorithms, they could inform new understanding of phytoplankton ecology, biodiversity, and environmental stressors.

Optics EXPRESS

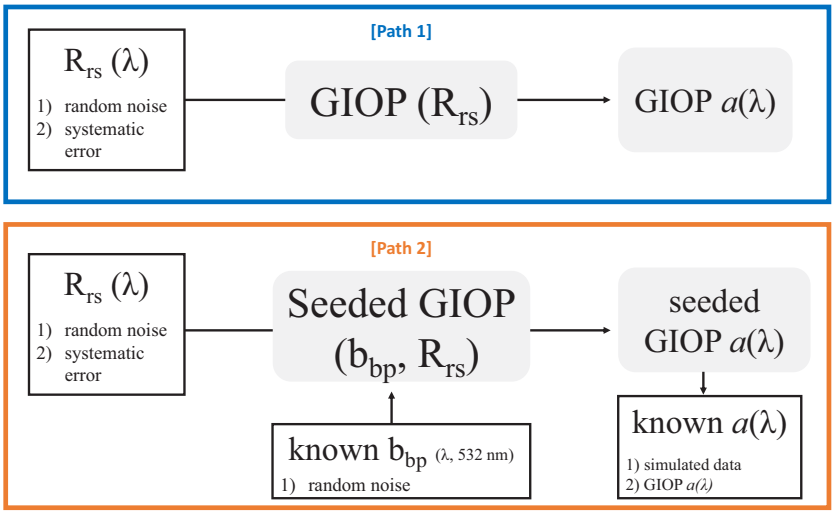

Fig. 1. Illustration of workflow for the two inversion schemes from primary data products (R_{rs}, λ) to output total absorption (a, λ). [Path 1] = default pathway. [Path 2] = seeded-inversion pathway. Seeded products are compared to the simulated dataset 'true' absorption values (for case study 1) and to satellite values (GIOP absorption described in [1], case study 2).

In the following sections, we begin with a description of the ancillary $b_{bp}(532)$ and $R_{rs}(\lambda)$ datasets used for our analysis (§2.1), the methodology for adapting the default configuration of GIOP (GIOP-DC) to ingest b_{bp} at a single wavelength (§2.2), and an uncertainty procedure to perturb $R_{rs}(\lambda)$ realistically with both random and systematic noise (§2.3). We then explore how seeded-inversions can affect downstream products from point-by-point to global scales and, finally, close with some recommendations for improving the next generation of ocean color algorithms and opportunities for expanding upon the work presented herein.

2. Methods

2.1. Datasets

We acquired synthesized spectral $R_{rs}(\lambda)$, $b_{bp}(\lambda)$, $a(\lambda)$, and chlorophyll-a concentrations (Chl; mg m^{-3}) from the IOCCG Ocean Colour Algorithms Working Group synthetic data set [27]. This dataset defines IOPs across four orders of magnitude variation in Chl (20 values within the range 0.03 to 30 mg m^{-3}, N = 500). $R_{rs}(\lambda)$ spectra are modeled from these IOPs using Hydrolight radiative transfer code [39], with outputs defined in 10 nm spectral steps from 400-800 nm. Given this use of a forward radiative transfer model, we consider the resultant $R_{rs}(\lambda)$ to be relatively free of errors, although we note that model assumptions regarding the sky model, the volume scattering function, the state of the air-water interface, and inelastic scattering will affect generated $R_{rs}(\lambda)$. We interpolate simulated $R_{rs}(\lambda)$ at the same wavelengths as MODIS-Aqua (412, 443, 488, 531, 547, and 667 nm) for consistent calculations across all datasets. Accordingly, a single wavelength of simulated b_{bp} is used (532 nm) as our seeded GIOP input.

To assess the extent of changes in seeded GIOP $a(\lambda)$ on regional and global scales, we used level-3 MODIS-Aqua $R_{rs}(\lambda)$ observations (R2022 reprocessing) in tandem with satellite lidar data. MODIS-Aqua data were acquired from the NASA Ocean Biology Processing Group (https://oceancolor.gsfc.nasa.gov). Details of MODIS-Aqua IOP processing can be found at https://oceancolor.gsfc.nasa.gov/atbd/giop/ and are described below. For the lidar data, we first

Optics EXPRESS

explore a satellite lidar dataset from ICESat-2 collected in December 2018 (processing details in [40]). ICESat-2 derived $b_{bp}(532)$ come from the ATL03 geotagged photon height product (https://nsidc.org/data/atl03/versions/5). While these data have not been formally validated, they were considered by [40] to exhibit general agreement with MODIS-Aqua $b_{bp}(532)$ retrievals. The product continues to be developed and refined [41]. We acquired Cloud-Aerosol Lidar with Orthogonal Polarization (CALIOP) $b_{bp}(532)$ daytime products from the Oregon State University Ocean Productivity group (http://orca.science.oregonstate.edu/lidar.data.php). These data are collected at nearly the same time as MODIS-Aqua because the two platforms are in the same satellite constellation. Details of CALIOP processing are in [22]. For climatological comparisons, MODIS-Aqua and CALIOP data are spatially binned and averaged to a 2-degree by 2-degree global grid and temporally averaged to monthly resolutions over the shared time period of 2002-2017. The broad space/time averaging for the climatologies ensures sufficient observations per grid box, provides representative coverage for broad global comparisons, and accommodates differences in coverage and revisit times between CALIOP and MODIS-Aqua. As noted above, CALIOP $b_{bp}(532)$ was previously found to be more accurate than MODIS-Aqua $b_{bp}(531)$ based on both point-by-point comparisons with in situ data and using monthly averaged data over large spatial b_{bp} scales [20–21]. We therefore consider it suitable for the current exploratory study. In Section 5 below, we also consider alternative 'seeding' schemes for lidar data.

All datasets described above were intended to be globally representative. The simulated dataset is constructed across a large dynamic range of observed chlorophyll-*a* concentrations, whereas ICESat-2, CALIOP, and MODIS-Aqua all have global spatial coverage. The datasets generally have good agreement between measured and inverted (λ) [20,26–27,40], so their use in this study likely represents a conservative estimate for average absorption changes due to seeding, while also representing a fuller range of outcomes on a point-by-point basis given their global coverage. It should go without saying that the use of lower quality estimates of $b_{bp}(\lambda)$ inversion products may cause more dramatic changes in absorption retrievals. That is, seeding an ocean reflectance inversion algorithm with 'known' $b_{bp}(\lambda)$ of any quality does not necessarily mean the retrieved products will be of higher quality, nor that they will be significantly different from the non-seeded retrievals, so it is important to explore the plausible range of outcomes across scales in order to direct future work.

2.2. Ocean reflectance inversion model

The GIOP framework adopted for this study uses the following general form in its default configuration (see [26] for additional details). After converting from above-surface R_{rs} to subsurface r_{rs} $\left(r_{rs}(\lambda) = \frac{R_{rs}(\lambda)}{0.52+1.7R_{rs}(\lambda)}, [43]\right)$, the $r_{rs}(\lambda)$-IOP relationship is defined as:

$$r_{rs}(\lambda) = g_1 u(\lambda) + g_2 u(\lambda)^2, \tag{1a}$$

$$u(\lambda) = \frac{b_b(\lambda)}{b_b(\lambda) + a(\lambda)} \tag{1b}$$

where g_1 and g_2 are spectrally fixed at 0.0949 and 0.0794, respectively, as in [42]. Both $b_b(\lambda)$ and $a(\lambda)$ can be further expanded as the sums of multiple components:

$$a(\lambda) = a_w(\lambda) + a_{ph}(\lambda) + a_{dg}(\lambda), a_{ph}(\lambda) \tag{2a}$$

$$a(\lambda) = a_w(\lambda) + M_{ph}a^*_{ph}(\lambda) + M_{dg}a^*_{dg}(\lambda), \tag{2b}$$

$$b_b(\lambda) = b_{bw}(\lambda) + b_{bp}(\lambda) \tag{3a}$$

$$b_b(\lambda) = b_{bw}(\lambda) + M_{bp}b^*_{bp}(\lambda), \tag{3b}$$

where $a_w(\lambda)$, $a_{dg}(\lambda)$, and are respectively absorption by seawater (ranging from 0.0046 at 412 nm to 0.43 at 667 nm), colored dissolved organic matter (CDOM) plus detritus, and phytoplankton.

Optics EXPRESS

Similarly, $b_{bw}(\lambda)$ and $b_{bp}(\lambda)$ are respectively backscattering by seawater and particles. Note that $a_{dg}(\lambda)$, $a_{ph}(\lambda)$, and $b_{bp}(\lambda)$ are also presented as the product of a magnitude (M) and spectral shape (denoted with an asterisk, *). In the default configuration of GIOP (GIOP-DC), the three free parameters are given by M, with the assumed spectral shapes informed from various studies applied over the $b^*_{bp}(\lambda)$ wavelength range considered herein. Currently, GIOP-DC adopts a power-law expression for with a dynamically-assigned spectral exponent as an empirical function of R_{rs} ratios calculated following [43], an exponential expression for $a^*_{dg}(\lambda)$ with a fixed decay exponent of 0.018 nm^{-1}, and a dynamically-assigned, concentration-specific $a^*_{ph}(\lambda)$ calculated following [34] that requires an estimate of Chl as input. Both $a_w(\lambda)$ and $b_{bw}(\lambda)$ are considered known (e.g., [28,30] and references therein) and a Raman correction is applied [29]. Using R_{rs} from 400-700 nm and the spectral shape assumptions as input, GIOP-DC estimates values for the magnitudes (i.e., M terms) using nonlinear least squares inversion of Eqs. 1–3. Additional algorithm metrics and quality control practices are described in [26].

For the current study, GIOP-DC was first run as described above and then rerun with M_{bp} provided as input ('seeded GIOP'). For this latter case, we considered M_{bp} provided as part of the synthetic dataset and M_{bp} from ICESat-2 or CALIOP. In all cases, we report and evaluate $a(\lambda)$ at 412, 443, 488, 531, 547, and 667 nm rather than the absorption coefficient of any given constituent to minimize additional uncertainty introduced by partitioning $a(\lambda)$ into its phytoplankton, detrital, and/or dissolved contributions.

2.3. Uncertainty propagation and comparative statistics

While the simulated $R_{rs}(\lambda)$ dataset is designed to have no radiometric errors, real datasets have uncertainty and error associated with them. As such, we perturb the $R_{rs}(\lambda)$ data with both random noise and systematic error to explore how $a(\lambda)$ retrievals are affected by realistic uncertainties. We simulate random noise by perturbing $R_{rs}(\lambda)$ values with 5% random (i.e., varying in magnitude and direction per band) Gaussian noise. We add systematic error to simulated $R_{rs}(\lambda)$ to emulate the behavior of a recently documented systematic bias in ocean color data [21]. Relative systematic error values (10% at 412, 443, and 488 nm, 20% at 531 and 547 nm, and 50% at 667 nm) were based on observations at the Marine Optical BuoY (MOBY) site near Hawaii, which may not fully reflect biases at all locations and times but are sufficient as a simple case study to explore how systematic errors potentially affect outcomes. Systematic error is not added to the satellite data because MODIS-Aqua data already contain this bias [21]. To test how error in $b_{bp}(532)$ affects the performance of seeded GIOP, we add 20% random Gaussian noise to both simulated and satellite input (seed) $b_{bp}(532)$ data.

For all datasets, we calculate median relative bias (given by the ratio between two products minus 1, where a value of 0 indicates equivalence), Spearman's rank correlation (a non-parametric assessment of correlation, ρ), median absolute error (MAE, in the case of simulated data), and median absolute difference (MAD, for the satellite data comparisons). Statistics are reported relative to GIOP-DC $a(\lambda)$ retrievals except for the simulated data case, in which statistics are reported relative to simulated $a(\lambda)$. Spearman's rank correlation is a generally more conservative measure of the strength and direction existing between two datasets compared to Pearson's correlation, as it is not affected by the range and variability in absolute data values. We use the two-sample Kolmogorov-Smirnov (K-S) test for statistically significant differences among the different GIOP configurations.

3. Case study results

3.1. Case study 1: IOCCG simulated algorithm development data

Seeded $a(\lambda)$ values from the simulated dataset were retrieved by ingesting a single value of simulated $b_{bp}(532)$ as M_{bp} within the inversion scheme. Results show visual improvement

Optics EXPRESS

(less scatter around the 1:1 line) in the seeded GIOP $a(\lambda)$ compared to the GIOP-DC results (Fig. 2) that are confirmed by both lower error (Fig. 3) and higher correlations (Spearman's rank correlation, Fig. S1). Note that the seeded GIOP $a(\lambda)$ values are slightly biased above the 1:1 line (Fig. 2,3, S1), which could result from the synthetic $b_{bp}(\lambda)$ dataset that also exhibits slight documented biases (Fig. 5.1, Table 5.2 in IOCCG, 2006). When random or systematic error are not added to $R_{rs}(\lambda)$ (Fig. 3(a), b), the seeded GIOP results outperform the GIOP-DC for the green and red wavelengths but generally not the blue. Seeded GIOP results outperform GIOP-DC at all wavelengths except 412 nm when 5% random error is added to $R_{rs}(\lambda)$ (Fig. 3(b)). This performance improvement is especially pronounced after systematic error is added to $R_{rs}(\lambda)$ (Fig. 3(c)). Modifying $b_{bp}(532)$ values used to seed the GIOP with 5% or 20% random error (Fig. 3(d)) results in negligible performance differences, with seeded GIOP results still generally outperforming GIOP-DC at the green and red wavelengths (especially in terms of ρ, Fig. S1).

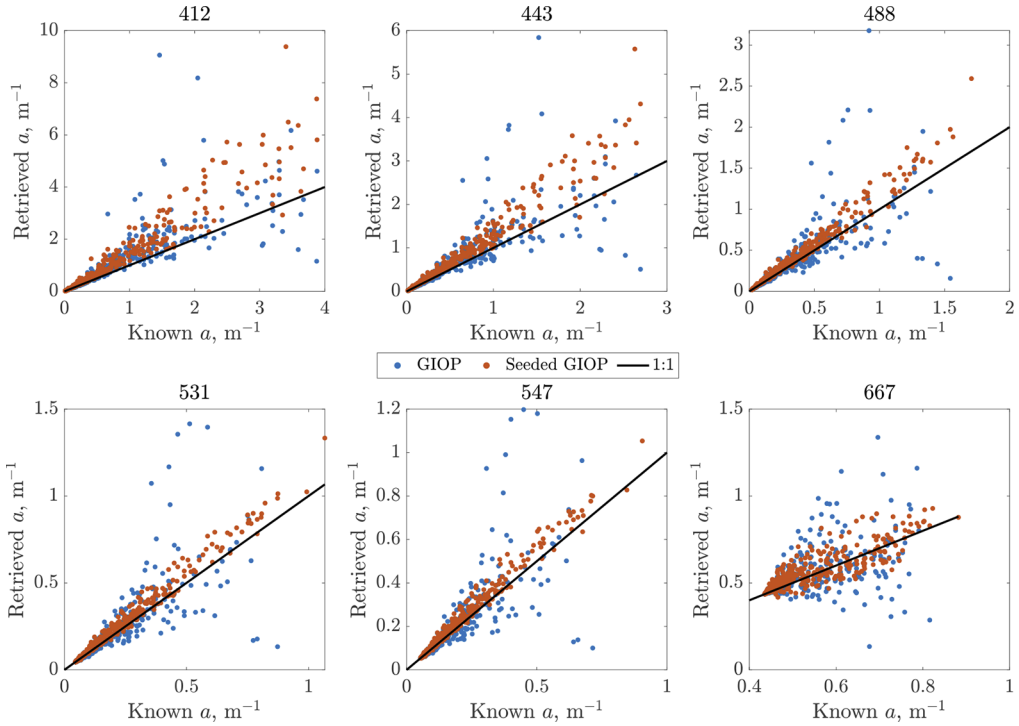

Fig. 2. [Case study 1] Comparison of GIOP-DC (blue circles) and seeded GIOP (orange circles) predictions of $a(\mathrm{m}^{-1})$ versus known values across different wavelengths (nm, 6 different subplots). Black line = 1:1 relationship.

One reason for the general close correspondence between GIOP-DC and seeded GIOP $a(\lambda)$ is that both configurations are derived from the same $R_{rs}(\lambda)$ and using the same assumed spectral shapes of absorption components. The only difference between GIOP-DC and seeded GIOP $a(\lambda)$ is the method used to define or retrieve M_{bp} – that is, the former has three free parameters, whereas the latter has only two. Despite similarities in the construction and data products used herein, Kolmogrov-Smirnov (K-S) testing (N = 500) reveals that the seeded GIOP results are significantly different from GIOP-DC results at the 5% significance level across all perturbations. Altogether, the simulated data suggest that improvements to GIOP-DC products can come by incorporating known values of $b_{bp}(532)$, even in the most conservative case when $R_{rs}(\lambda)$ is assumed to have no error.

Optics EXPRESS

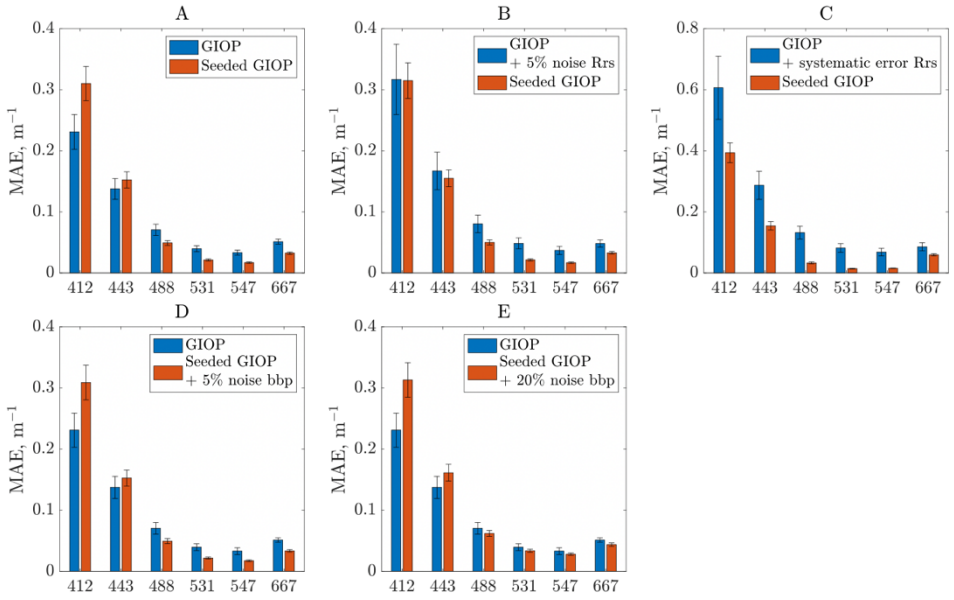

Fig. 3. [Case study 1] Median absolute error (MAE) in $a(\lambda, \mathrm{m}^{-1})$ for the GIOP-DC (blue) and seeded GIOP (orange) under conditions of A) no error added, B) 5% random noise added to $R_{rs}(\lambda)$, C) systematic error added to $R_{rs}(\lambda)$, D) 5% random noise added to b_{bp}, E) 20% random noise added to b_{bp}. Bar graphs are grouped by wavelength (nm) along the x-axis. Error bars are standard error.

3.2. Case study 2: satellite data from lb_{bp}CESat-2 and CALIOP

Seeding GIOP with ICESat-2 data results in $a(\lambda)$ values that are broadly comparable ($\rho > 0.9$ for most wavelengths) to those derived from GIOP-DC (Fig. 4). Obvious deviations appear for larger MODIS-Aqua $a(\lambda)$ values, where the corresponding derived $a(\lambda)$ values from seeding with ICESat-2 data are much higher and biased positive. As with the simulated dataset, the MAD between ICESat-2 seeded GIOP and GIOP-DC run with MODIS-Aqua data are relatively small (on the order of 0.001-0.01m^{-1}), but a clear clustering emerges among the seeded GIOP configurations with ICESat-2 (right bars compared to left bar, Fig. 5, S2). All of the seeded GIOP configurations are significantly different from GIOP-DC or GIOP-DC + 5% $R_{rs}(\lambda)$ noise cases at the 5% significance level. In essence, adding random noise to $R_{rs}(\lambda)$ does not significantly change the $a(\lambda)$ distribution, whereas seeding does. Compared to the simulated dataset, we see large differences in $a(\lambda)$ values (which can be nearly a factor of 2), particularly for larger $a(\lambda)$ values. These large differences provide evidence that seeding the GIOP with $b_{bp}(532)$ can create substantial differences in absorption for certain places and times, ultimately indicating the need to use products with low error and high fidelity across space and time so as to not confound interpretation of these differences. Future performance assessments of ICESat-2 $b_{bp}(532)$ are warranted, especially with consideration to higher attenuating waters and before ICESat-2 data are distributed on global scales.

GIOP-DC $a(\lambda)$ and CALIOP $b_{bp}(532)$ seeded GIOP fall along the 1:1 line (Fig. 6) and have generally low median differences (Fig. 7). Similar to results with ICESat-2 (Fig. 5), the seeded GIOP configurations group together (right three bars, Fig. 7) compared to the GIOP-DC with random noise (left bar). K-S tests for differences (N = 51,048) again confirm that the seeded GIOP results are significantly different from GIOP-DC at the 5% significance level. Taken together, seeding the GIOP with $b_{bp}(532)$ produces significantly different $a(\lambda)$ results for all wavelengths.

Optics EXPRESS

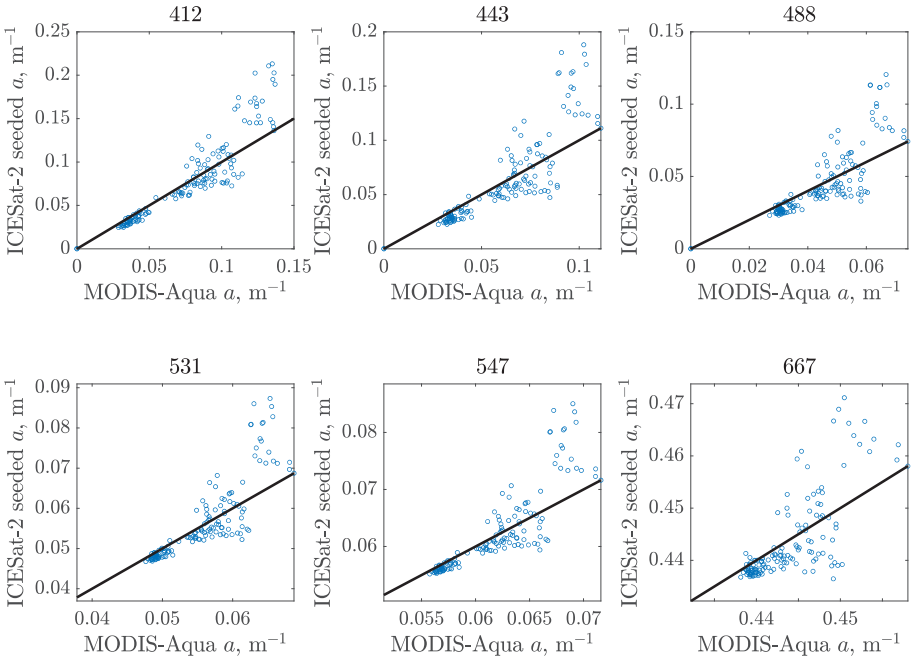

Fig. 4. [Case study 2] $a(\lambda)$ from the seeded GIOP via ICESat-2 b_{bp} (y-axis) and MODIS-Aqua a (x-axis) from the GIOP-DC, grouped by wavelength (nm). Black line = 1:1 relationship between data across the two sensors.

We note that the large sample size in this case may favor statistically significant differences. The low absolute magnitude differences between different configurations above led to the question: are there important differences in global patterns between seeded and GIOP-DC $a(\lambda)$?

We find spatial differences between GIOP-DC and seeded GIOP that vary in magnitude depending on location and time of year (Fig. 8). In particular, GIOP-DC total absorption (at 443 nm, see Fig. S4-S8 for other wavelengths) is enhanced relative to the seeded GIOP in most of the world's ocean (blue areas are where GIOP-DC > seeded GIOP). Exceptions are along North America's Pacific Coast, the Indian Ocean, the North Atlantic, the west coast of Africa, and waters near Japan, where the seeded GIOP exceeds GIOP-DC total absorption by more than 50% during certain months. Looking ahead, differences between GIOP-DC and seeded GIOP total absorption, both spatially and seasonally, could be used to re-cast interpretations of absorption in particular regions and times if they are found to improve upon GIOP-DC absorption values.

Optics EXPRESS

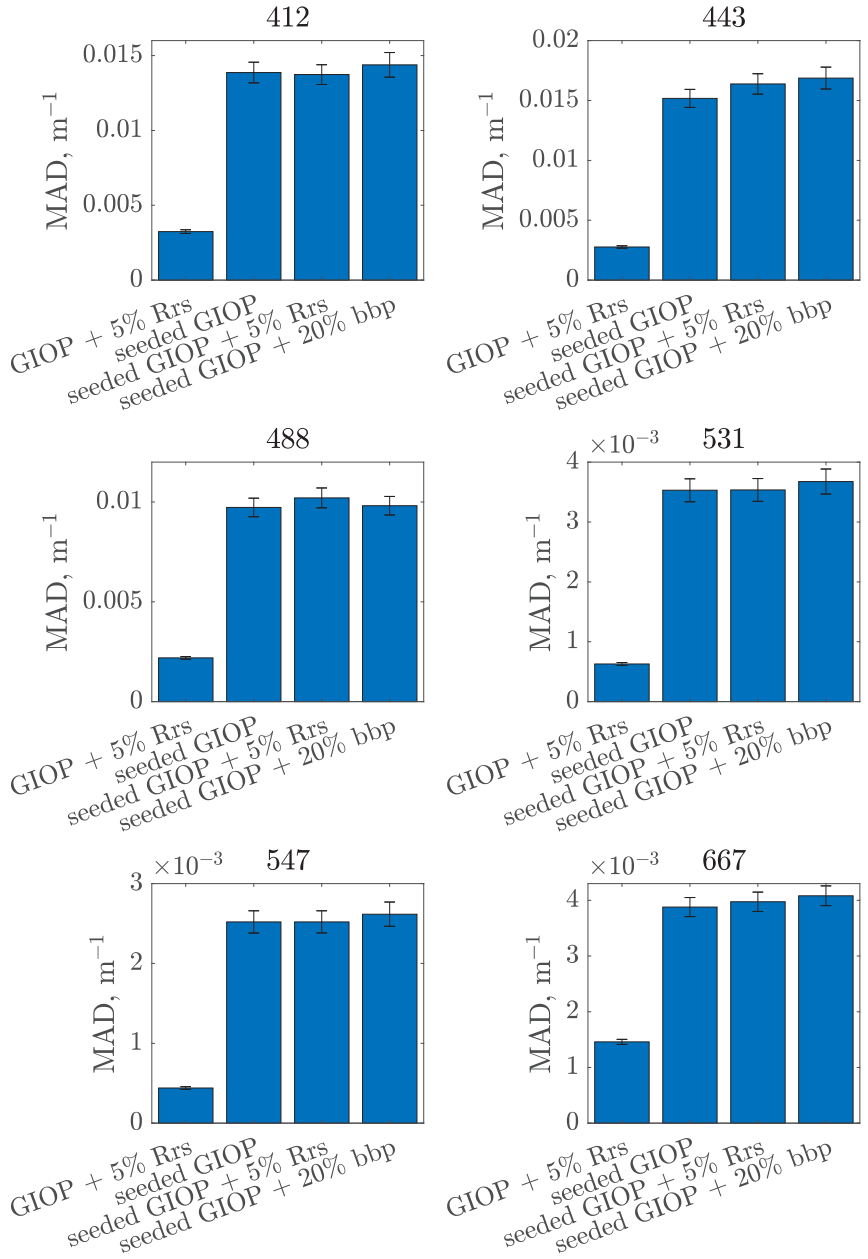

Fig. 5. [Case study 2] Median absolute difference in $a(\lambda)$ from different GIOP configurations (different bars) and wavelengths (nm, subplots), all relative to $a(\lambda)$ from GIOP-DC. Seeded GIOP products are based on ICESat-2 b_{bp}. Error bars are standard error.

Optics EXPRESS

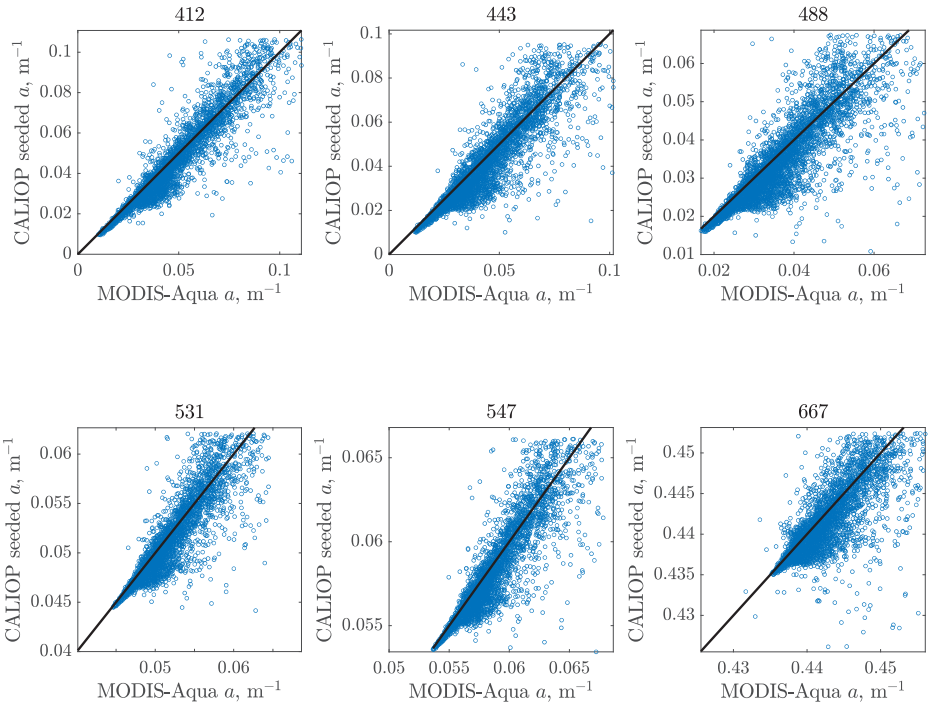

Fig. 6. [Case study 2] Scatter graphs of $a(\lambda)$ from the GIOP seeded with CALIOP b_{bp} data (y-axis) and MODIS-Aqua $a(\lambda)$ (x-axis) from GIOP-DC. Panels correspond to different wavelength (nm; indicated at top). Black line = 1:1 relationship between data across the two sensors.

Optics EXPRESS

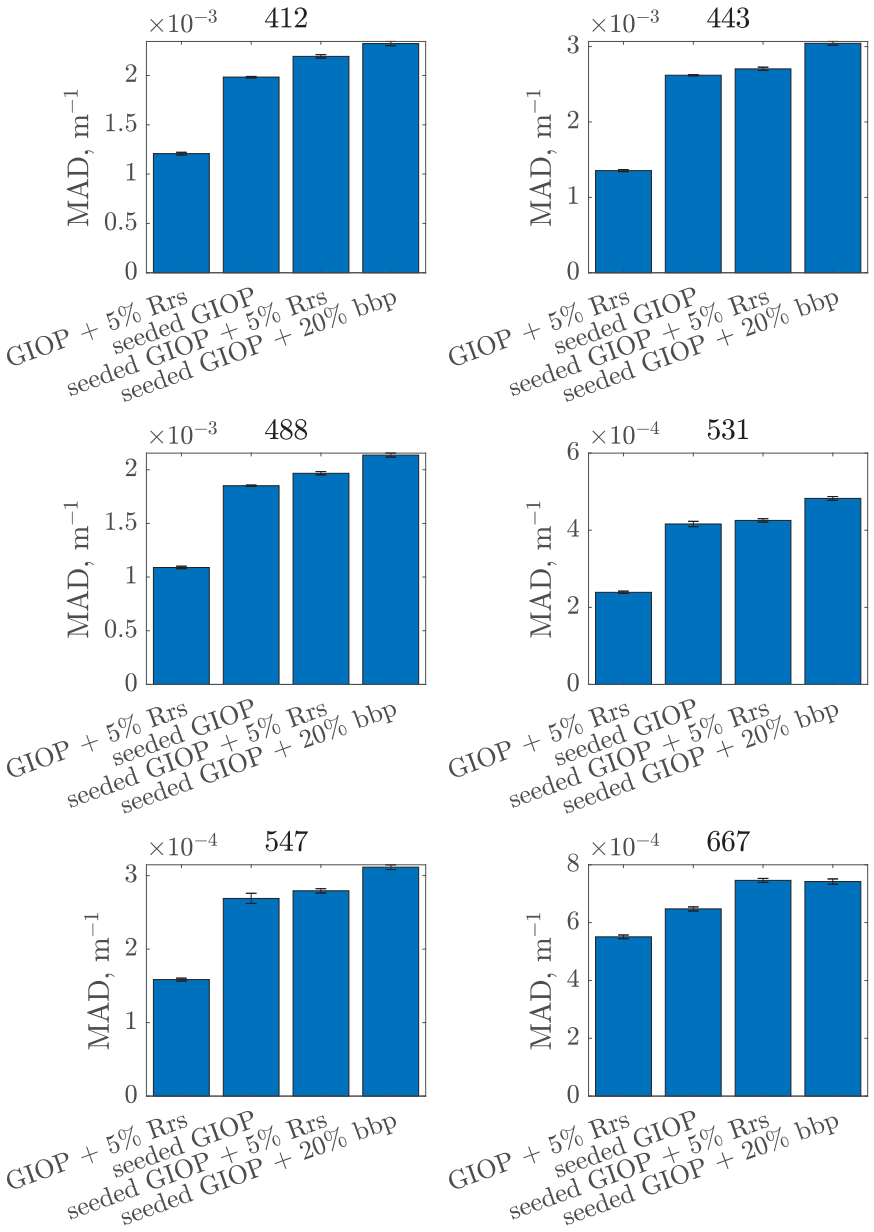

Fig. 7. [Case study 2] Median absolute difference in $a(\lambda)$ from different GIOP configurations (different bars) and wavelengths (nm, subplots), all relative to $a(\lambda)$ from GIOP-DC. Seeded GIOP products are based on CALIOP b_{bp}. Error bars are standard error.

Optics EXPRESS

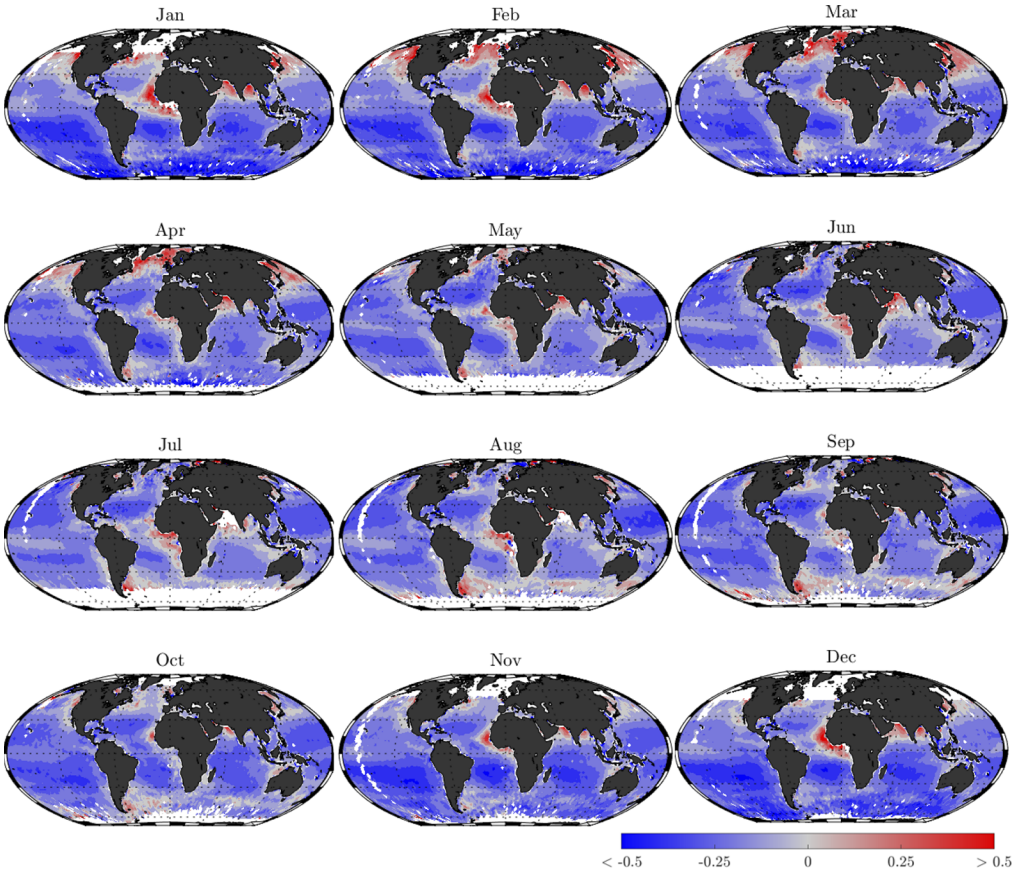

Fig. 8. [Case study 2] Monthly (indicated at top of each panel) global maps of relative $a(443\,\text{nm})$ differences defined as the seeded GIOP value minus GIOP-DC value. Note that while color bar is restricted to a range of -0.5 to 0.5 (i.e., -50% to 50%), some cases exceeded these values. Seeded GIOP products are based on CALIOP b_{bp} (532 nm). Note the north-south stripe of missing data in the Pacific is due to a gap in ancillary data used in CALIOP processing [22].

Optics EXPRESS

4. Summary

In this study, we explore how GIOP products differ after seeding GIOP with known observations of particulate backscattering, which also reduced the number of free parameters in the inversion of $R_{rs}(\lambda)$ from three to two. While we did not directly validate retrieved absorption quantities between the seeded or non-seeded case, we performed the exercise of demonstrating improved $b_{bp}(532)$ with satellite lidar [20,21] and by default, if we can seed with improved $b_{bp}(532)$ then we might expect improvement in absorption. This analysis used a simulated dataset that was created to test $R_{rs}(\lambda)$ inversion algorithms (arguably providing a 'best case' testbed for the idea of adding ancillary data to the GIOP) and found varied performance results across different metrics and through the addition of either random noise or systematic error. Seeding the GIOP with independent $b_{bp}(532)$ improved performance of derived $a(\lambda)$ values even when the input $R_{rs}(\lambda)$ observations were assumed to be error free. Therefore, direct substitution of $b_{bp}(\lambda)$ within an inversion can generate improved absorption products. Altogether, improvements to GIOP retrievals can come by seeding the model directly with independent data.

We used satellite data to explore differences in datasets with varied errors and uncertainties. We found broadly consistent qualitative results using both ICESat-2 and CALIOP, as was seen with the simulated dataset. Compared to the simulated data findings, the satellite case study shows that differences in $b_{bp}(532)$ values used to seed the GIOP can result in substantial point-wise differences in $a(\lambda)$. While differences in the absolute magnitude of absorption may seem small on average (on the order of 1×10^{-3} m^{-1}), global patterns in these differences can be relatively large, exceeding 50% in some regions and times (Fig. 8). Finally, although satellite lidars currently only provide $b_{bp}(\lambda)$ at a single wavelength and thus rely on $R_{rs}(\lambda)$ to inform spectral b_{bp}, we still find statistically significant differences in absorption products after seeding the GIOP with b_{bp} at a single wavelength. Improved approaches may come through incorporating additional b_{bp} wavelengths either through an ocean-optimized lidar or in situ datasets.

5. Recommendations for future work

Satellite sensors available for ocean research continue to diversify, presenting new opportunities to advance scientific understanding. For example, the complement of the hyperspectral ultraviolet-to-near infrared ocean color instrument (OCI) and multi-wavelength, muti-spectral polarimeters on the upcoming Plankton, Aerosol, Cloud, ocean Ecosystem mission (PACE, [44]) is expected to improve both atmospheric corrections of OCI data and characterizations of upper ocean plankton and particle assemblages. Achieving these goals will require the remote sensing science community to reconcile viewing angle and spatial coverage differences between the PACE instruments. From the current study, we propose that coincident global lidar data can similarly complement ocean color measurements, improving inversion products and increasing information content.

While the CALIOP and ICESat-2 sensors have provided proof-of-concept demonstrations of ocean retrievals with satellite lidar, our study was particularly motivated by the potential of future satellite lidar to employ advanced technologies and yield even more accurate b_{bp} retrievals than heritage lidar [45]. In anticipation of such missions, it is prudent now to investigate advantages proffered through merged active and passive remote sensing data and to develop strategies addressing differences between lidar (narrow field of view, single viewing angle) and ocean color (continuous swaths) data coverage (akin to the above noted issues regarding merged multi-sensor PACE data). With respect to lidar data, perhaps the simplest approach is to utilize it much like current in situ data are used for tuning ocean color algorithms and validating products, by considering appropriate space/time matchup scales [46]. The tremendous advantage here of the lidar data is that it provides spatially unbiased, global test data on a monthly basis that dwarfs the coverage over in situ ship-based measurements. Alternatively, it may be possible to directly seed

Optics EXPRESS

satellite $R_{rs}(\lambda)$ inversion algorithms with spatially-extrapolated lidar data. Specifically, evaluation of results presented in Fig. 8 indicates that differences between seeded GIOP and GIOP-DC exhibit broadly coherent but temporally-varying global patterns, potentially suggesting regional and seasonal biases in the standalone global algorithm. In such a case, a global lidar product without sampling gaps could first be created through machine learning or interpolation methods and then used at different scales directly within an inversion scheme. Irrespective of details, the overarching message is that the time is ripe to think creatively about how to synergistically use multiple space assets (radiometers, polarimeters, lidar, etc.) that each observe the ocean from a different perspective and present different strengths and weaknesses.

With respect to follow-on analyses, the current study employed the GIOP algorithm because it was specifically designed as a test bed with adaptations for other schemes (e.g., changing both IOP and AOP assumptions, [26]). Different inversion algorithm parameterizations, and different inversion techniques for solutions, should be considered when testing performance changes across seeded and unseeded products, particularly when an assumed backscattering spectral slope is used (see [47]). We found that choosing either a static (as in GSM, [48]) or dynamic b_{bp} slope (as a function of R_{rs}, [43]) did not change our primary finding that seeded-inversion absorption products are possibly more accurate (evinced by the simulated data) than those generated by the standard implementation. Also, in GIOP-DC, the mismatch between measured and modeled $R_{rs}(\lambda)$ is determined by minimizing a cost function, namely $\sum (R_{rs,\,meas} - R_{rs,\,GIOP})^2$. The greatest cost occurs for absolute $R_{rs}(\lambda)$ differences in the blue wavelengths, which typically have larger magnitude differences than those in the green and red in open ocean environments. Future cost functions could weigh the $R_{rs}(\lambda)$ wavelengths by their uncertainties, signal-to-noise ratios, or relative differences, which would adjust both GIOP-DC and seeded GIOP performance of absorption across all wavelengths rather than preferentially adjusting the blue. Another approach to eliminate the influence of outliers is using absolute values or the square root in a cost function, rather than magnifying large differences as is done presently.

We performed our analysis using a simulated dataset and available satellite data. The range in optical water bodies considered in this study is vast and the focus is on general statistical differences between seeded and GIOP-DC outcomes. In addition to the lidar-based products used in the present study, additional ancillary data might include temperature, salinity (beyond their contribution to sea-water absorption and scattering, which is already included), mixed-layer depth, or other variates (e.g., [17–18]), which in turn may inform derived products by not treating $R_{rs}(\lambda)$ observations in isolation [10].

Finally, we seeded the GIOP in the simplest possible case by considering the known $b_{bp}(532)$ product to be 'truth' and conducted sensitivity tests with varied uncertainty of $b_{bp}(\lambda)$ values from theoretical or satellite datasets. Future work should consider implementing a Bayesian framework where $b_{bp}(\lambda)$ is treated as 'prior' information and has an assigned likelihood distribution that considers both its uncertainty and possible bias [12,31]. Machine learning methods that incorporate $b_{bp}(\lambda)$ and other ancillary data, besides the direct replacement scheme we explore here, should also be investigated. In this study, we replaced an unknown parameter with a direct observation, which reduces the unknown parameters in GIOP by one. In addition to asking how seeding the GIOP may improve performance and change biogeochemical interpretation of water bodies, one may also ask, 'What can be done with the extra free parameter we have now?'

Funding. National Aeronautics and Space Administration (80NSSC18K0957, 80NSSC20K0970).

Acknowledgements. This work was supported by NASA Ocean Biology and Biogeochemistry grant 80NSSC20K0970for KB and MJB and grant 80NSSC18K0957 for KB, MJP, EB, and PJW. SJK was supported by a Simons Foundation Postdoctoral Fellowship in Marine Microbial Ecology. APC was supported by a Washington Research Foundation Postdoctoral Fellowship. We thank 3 anonymous reviewers for their ideas to improve the manuscript.

Disclosures. The authors declare no conflicts of interest.

Data Availability. Data from this paper are publicly available in Ref. [27].

Optics EXPRESS

Supplemental document. See Supplement 1 for supporting content.

References

1. A. Y. Morel and H. R. Gordon, "Report of the working group on water color," Boundary-Layer Meteorology **18**(3), 343–355 (1980).
2. R. Doerffer and J. Fischer, "Concentrations of chlorophyll, suspended matter, and gelbstoff in case II waters derived from satellite coastal zone color scanner data with inverse modeling methods," J. Geophys. Res.: Oceans **99**(C4), 7457–7466 (1994).
3. J. R. Graff, T. K. Westberry, A. J. Milligan, M. B. Brown, G. Dall'Olmo, and V. Van Dongen-Vogels, M. J. Behrenfeld, "Analytical phytoplankton carbon measurements spanning diverse ecosystems," Deep Sea Res., Part I **102**, 16–25 (2015).
4. A. M. Ciotti, M. R. Lewis, and J. J. Cullen, "Assessment of the relationships between dominant cell size in natural phytoplankton communities and the spectral shape of the absorption coefficient," Limnol. Oceanogr. **47**(2), 404–417 (2002).
5. D. Stramski, R. A. Reynolds, M. Kahru, and B. G. Mitchell, "Estimation of particulate organic carbon in the ocean from satellite remote sensing," Science **285**(5425), 239–242 (1999).
6. C. Mitchell, C. Hu, B. Bowler, D. Drapeau, and W. M. Balch, "Estimating particulate inorganic carbon concentrations of the global ocean from ocean color measurements using a reflectance difference approach," J. Geophys. Res.: Oceans **122**(11), 8707–8720 (2017).
7. D. Aurin, A. Mannino, and D. J. Lary, "Remote sensing of CDOM CDOM spectral slope, and dissolved organic carbon in the global ocean," Appl. Sci. **8**(12), 2687 (2018).
8. N. B. Nelson and D. A. Siegel, "The global distribution and dynamics of chromophoric dissolved organic matter," Annu. Rev. Mar. Sci. **5**(1), 447–476 (2013).
9. B. B. Cael, A. Chase, and E. Boss, "Information content of absorption spectra and implications for ocean color inversion," Appl. Opt. **59**(13), 3971–3984 (2020).
10. B. B. Cael, K. Bisson, E. Boss, and Z. K. Erickson, "How many independent quantities can be extracted from ocean color?" Limnol. Oceanogr. Lett. **8**(4), 603–610 (2023).
11. Neranga K. Hannadige, Peng-Wang Zhai, P. Jeremy Werdell, Meng Gao, Bryan A. Franz, Kirk Knobelspiesse, and Amir. Ibrahim, "Optimizing retrieval spaces of bio-optical models for remote sensing of ocean color," Appl. Opt. **62**(13), 3299–3309 (2023).
12. Z. K. Erickson, L. McKinna, P. J. Werdell, and I. Cetinić, "Bayesian approach to a generalized inherent optical property model," Opt. Express **31**(14), 22790–22801 (2023).
13. O. Dubovik and M. D. King, "A flexible inversion algorithm for retrieval of aerosol optical properties from Sun and sky radiance measurements," J. Geophys. Res. **105**(D16), 20673–20696 (2000).
14. A. H. Omar, J. G. Won, D. M. Winker, S. C. Yoon, O. Dubovik, and M. P. McCormick, "Development of global aerosol models using cluster analysis of Aerosol Robotic Network (AERONET) measurements," J. Geophys. Res. **110**(D10), D10 (2005).
15. O. Dubovik, M. Herman, A. Holdak, T. Lapyonok, D. Tanré, J. L. Deuzé, and A. Lopatin, "Statistically optimized inversion algorithm for enhanced retrieval of aerosol properties from spectral multi-angle polarimetric satellite observations," Atmos. Meas. Tech. **4**(5), 975–1018 (2011).
16. N. Pahlevan, B. Smith, J. Schalles, C. Binding, Z. Cao, R. Ma, and R. Stumpf, "Seamless retrievals of chlorophyll-a from Sentinel-2 (MSI) and Sentinel-3 (OLCI) in inland and coastal waters: A machine-learning approach," Remote Sensing of Environment **240**, 111604 (2020).
17. L. I. McKinna, P. R. Fearns, S. J. Weeks, P. J. Werdell, M. Reichstetter, B. A. Franz, and G. C. Feldman, "A semianalytical ocean color inversion algorithm with explicit water column depth and substrate reflectance parameterization," J. Geophys. Res.: Oceans **120**(3), 1741–1770 (2015).
18. A. P. Chase, E. S. Boss, N. Haëntjens, E. Culhane, C. Roesler, and L. Karp-Boss, "Plankton Imagery Data Inform Satellite-Based Estimates of Diatom Carbon," Geophys. Res. Lett. **49**(13), e2022GL098076 (2022).
19. P. K. Lange, P. J. Werdell, Z. K. Erickson, G. Dall'Olmo, R. J. Brewin, M. V. Zubkov, and I. Cetinić, "Radiometric approach for the detection of picophytoplankton assemblages across oceanic fronts," Opt. Express **28**(18), 25682–25705 (2020).
20. K. M. Bisson, E. Boss, P. J. Werdell, A. Ibrahim, and M. J. Behrenfeld, "Particulate backscattering in the global ocean: a comparison of independent assessments," Geophys. Res. Lett. **48**(2), e2020GL090909 (2021).
21. K. M. Bisson, E. Boss, P. J. Werdell, A. Ibrahim, R. Frouin, and M. J. Behrenfeld, "Seasonal bias in global ocean color observations," Appl. Opt. **60**(23), 6978–6988 (2021).
22. M. J. Behrenfeld, Y. Hu, K. M. Bisson, X. Lu, and T. K. Westberry, "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 Sensing of Environment **281**, 113235 (2022).
23. M. Sun, P. Chen, Z. Zhang, C. Zhong, C. Xie, and D. Pan, "Evaluation of the CALIPSO Lidar-observed particulate backscattering coefficient on different spatiotemporal matchup scales," Frontiers in Marine Science **10**, 1181268 (2023).

Optics EXPRESS

24. P. Di Girolamo, A. Cosentino, F. Longo, N. Franco, D. Dionisi, D. Summa, and S. Zoffoli, (2022, May), "Cloud Aerosol Lidar for Global Scale Observations of the Ocean-Land-Atmosphere System-CALIGOLA," In *EGU General Assembly Conference Abstracts* (pp. EGU22–5644).
25. N. Franco, P. Di Girolamo, D. Summa, M. Di Paolantonio, and D Dionisi, "Preliminary Studies and Performance Simulations in support of the mission "CALIGOLA," EGU General Assembly 2023, Vienna, Austria, 24–28 Apr 2023, EGU23-16695, https://doi.org/10.5194/egusphere-egu23-16695, 2023.
26. P. J. Werdell, B. A. Franz, S. W. Bailey, G. C. Feldman, E. Boss, V. E. Brando, and A. Mangin, "Generalized ocean color inversion model for retrieving marine inherent optical properties," Appl. Opt. **52**(10), 2019–2037 (2013).
27. IOCCG, (2006). *Remote Sensing of Inherent Optical Properties: Fundamentals, Tests of Algorithms, and Applications*. Z.-P. Lee (ed.), Reports of the International Ocean-Colour Coordinating Group, No. 5, IOCCG, Dartmouth, Canada
28. P. J. Werdell, B. A. Franz, J. T. Lefler, W. D. Robinson, and E. Boss, "Retrieving marine inherent optical properties from satellites using temperature and salinity-dependent backscattering by seawater," Opt. Express **21**(26), 32611–32622 (2013).
29. L. I. W. McKinna, P. J. Werdell, and C. W. Proctor, "Implementation of an analytical Raman scattering correction for satellite ocean-color processing," Opt. Express **24**(14), A1123 (2016).
30. X. Yu, Z. Lee, J. Wei, and S. Shang, "Impacts of pure seawater absorption coefficient on remotely sensed inherent optical properties in oligotrophic waters," Opt. Express **27**(24), 34974–34984 (2019).
31. Z. K. Erickson, P. J. Werdell, and I. Cetinić, "Bayesian retrieval of optically relevant properties from hyperspectral water-leaving reflectances," Appl. Opt. **59**(23), 6902–6917 (2020).
32. C. S. Roesler, M. J. Perry, and K. L. Carder, "Modeling in situ phytoplankton absorption from total absorption spectra in productive inland marine waters," Limnol. Oceanogr. **34**(8), 1510–1523 (1989).
33. A. Bricaud, C. Roesler, and J. R. V. Zaneveld, "In situ methods for measuring the inherent optical properties of ocean waters," Limnol. Oceanogr. **40**(2), 393–410 (1995).
34. A. Bricaud, A. Morel, M. Babin, K. Allali, and H. Claustre, "Variations of light absorption by suspended particles with chlorophyll a concentration in oceanic (case 1) waters: Analysis and implications for bio-optical models," J. Geophys. Res.: Oceans **103**(C13), 31033–31044 (1998).
35. Z. Lee and K. L. Carder, "Absorption spectrum of phytoplankton pigments derived from hyperspectral remote-sensing reflectance," Remote sensing of environment **89**(3), 361–368 (2004).
36. J. Uitz, D. Stramski, R. A. Reynolds, and J. Dubranna, "Assessing phytoplankton community composition from hyperspectral measurements of phytoplankton absorption coefficient and remote-sensing reflectance in open-ocean environments," Remote Sensing of Environment **171**, 58–74 (2015).
37. A. Chase, E. Boss, R. Zaneveld, A. Bricaud, H. Claustre, J. Ras, and T. K. Westberry, "Decomposition of in situ particulate absorption spectra," Methods in Oceanography **7**, 110–124 (2013).
38. S. B. Woźniak, D. Stramski, M. Stramska, R. A. Reynolds, V. M. Wright, E. Y. Miksic, and A. M. Cieplak, "Optical variability of seawater in relation to particle concentration, composition, and size distribution in the nearshore marine environment at Imperial Beach," J. Geophys. Res. **115**(C8), C8 (2010).
39. C.D. Mobley, "Hydrolight 4.0 Users' Guide, Sequoia scientific inc, Mercer Island," (1998).
40. X. Lu, Y. Hu, Y. Yang, P. Bontempi, A. Omar, and R. Baize, "Antarctic spring ice-edge blooms observed from space by ICESat-2," Remote Sensing of Environment **245**, 111827 (2020).
41. X. Lu, Y. Hu, Y. Yang, T. Neumann, A. Omar, R. Baize, and D. Winker, "New Ocean Subsurface Optical Properties From Space Lidars: CALIOP/CALIPSO and ATLAS/ICESat-2," Earth and Space Science **8**(10), e2021EA001839 (2021).
42. H. R. Gordon, O. B. Brown, R. H. Evans, J. W. Brown, R. C. Smith, K. S. Baker, and D. K. Clark, "A semianalytic radiance model of ocean color," J. Geophys. Res. **93**(D9), 10909–10924 (1988).
43. Z. Lee, K. L. Carder, and R. A. Arnone, "Deriving inherent optical properties from water color: a multiband quasi-analytical algorithm for optically deep waters," Appl. Opt. **41**(27), 5755–5772 (2002).
44. P. J. Werdell, M. J. Behrenfeld, P. S. Bontempi, E. Boss, B. Cairns, G. T. Davis, and L. A. Remer, "The Plankton, Aerosol, Cloud, ocean Ecosystem mission: status, science, advances," Bull. Am. Meteorol. Soc. **100**(9), 1775–1794 (2019).
45. C. A. Hostetler, M. J. Behrenfeld, Y. Hu, J. W. Hair, and J. A. Schulien, "Spaceborne lidar in the study of marine systems," Annu. Rev. Mar. Sci. **10**(1), 121–147 (2018).
46. S. Vadakke-Chanat and C. Jamet, "Validation Protocol for the Evaluation of Space-Borne Lidar Particulate Back-Scattering Coefficient bbp," Front. Remote Sens. **4**, 1194580 (2023).
47. K. M. Bisson, E. Boss, T. K. Westberry, and M. J. Behrenfeld, "Evaluating satellite estimates of particulate backscatter in the global open ocean using autonomous profiling floats," Opt. Express **27**(21), 30191–30203 (2019).
48. S. Maritorena, D. A. Siegel, and A. R. Peterson, "Optimization of a semianalytical ocean color model for global-scale applications," Appl. Opt. **41**(15), 2705–2714 (2002).