Bayesian approach to a generalized inherent optical property model

ZACHARY K. ERICKSON,[1,*] **LACHLAN MCKINNA,**[2,3]
P. JEREMY WERDELL,[2] **AND IVONA CETINIĆ**[2,4]

[1]*NOAA Pacific Marine Environmental Laboratory, Seattle, WA, USA*
[2]*NASA Goddard Space Flight Center, Greenbelt, MD, USA*
[3]*Go2Q Pty Ltd., Sunshine Coast, QLD, Australia*
[4]*Goddard Earth Sciences Technology and Research (GESTAR) II, Morgan State University, Baltimore, MD, USA*
[*]*zachary.k.erickson@noaa.gov*

Abstract: Relationships between the absorption and backscattering coefficients of marine optical constituents and ocean color, or remote sensing reflectances $R_{rs}(\lambda)$, can be used to predict the concentrations of these constituents in the upper water column. Standard inverse modeling techniques that minimize error between the modeled and observed $R_{rs}(\lambda)$ break down when the number of products retrieved becomes similar to, or greater than, the number of different ocean color wavelengths measured. Furthermore, most conventional ocean reflectance inversion approaches, such as the default configuration of NASA's Generalized Inherent Optical Properties algorithm framework (GIOP-DC), require *a priori* definitions of absorption and backscattering spectral shapes. A Bayesian approach to GIOP is implemented here to address these limitations, where the retrieval algorithm minimizes both the error in retrieved ocean color and the deviation from prior knowledge, calculated using output from a mixture of empirically-derived and best-fit values. The Bayesian approach offers potential to produce an expanded range of parameters related to the spectral shape of absorption and backscattering spectra.

1. Introduction

The color of the ocean reflects the concentration and quality of optical constituents in the upper ocean [1]. Empirical algorithms have been developed and are widely used to provide estimates of the near-surface concentration of chlorophyll-*a* [2,3], particulate organic carbon [4], and particulate inorganic carbon [5,6], among others, from spectral remote sensing reflectances, $R_{rs}(\lambda)$ (sr^{-1}), or the ratio of upwelling radiance to downward solar irradiance at various wavelengths λ (nm). These formulae exploit observed wavelength-dependencies in the absorption or scattering of light from these compounds; for example, that chlorophyll-*a* absorbs light more strongly in blue wavelengths than in green in such a manner that the ratio of blue-to-green reflectances can be used to predict chlorophyll-*a* concentrations. However, the functional form of empirical algorithms and the coefficients that translate $R_{rs}(\lambda)$ into concentrations of upper ocean constituents are simply statistical relationships determined by fitting to data [7].

Semi-analytical algorithms (SAAs) provide an alternative approach to utilizing $R_{rs}(\lambda)$ information through assumptions of prior knowledge of the complete spectral shapes of absorption ($a(\lambda)$, m^{-1}) and backscattering ($b_b(\lambda)$, m^{-1}) from different optical constituents. Spectral matching type SAA algorithms typically comprise three components: (i) spectral shapes of the inherent optical properties (IOPs; $a(\lambda)$ and $b_b(\lambda)$), (ii) a forward reflectance model, and (iii) an inverse solution method. The sums of $a(\lambda)$ and $b_b(\lambda)$ from different constituents provide the total absorption and backscattering, $a_{tot}(\lambda)$ and $b_{b,tot}(\lambda)$, which are IOPs that relate directly to $R_{rs}(\lambda)$ through simplifications to the radiative transfer equation. Options to configure such an SAA are broadly available through the Generalized IOP framework (GIOP) [8], which is also summarized below.

The GIOP algorithm in its default configuration (GIOP-DC) provides a best-fit result, minimizing the error in calculated to observed $R_{rs}(\lambda)$ by assigning spectral shapes and varying magnitude parameters controlling the absorption by phytoplankton ($a_{ph}(\lambda_{ref})$, m^{-1}), absorption by detritus and colored dissolved organic material (CDOM), also known as "gelbstoff" ($a_{dg}(\lambda_{ref})$, m^{-1}), and backscattering of particulates ($b_{bp}(\lambda_{ref})$, m^{-1}) at reference wavelengths λ_{ref}. In this form, the three GIOP retrievals are the amplitudes for each term that combine to best recreate the input $R_{rs}(\lambda)$. The need to pre-define spectral shapes, however, remains a limitation, as such assignments cannot universally and accurately represent all aquatic conditions at all times. For example, photodegradation of CDOM by ultraviolet light alters its spectral slope and results in a loss of absorption preferentially at longer wavelengths [9] as CDOM is modified into lower molecular weight molecules [10]. Similarly, the backscattering spectral slope of particulates is related to the particle size distribution [11,12] and varies widely with phytoplankton community composition and physiological state [13–15].

In practice, most SAAs pre-define a spectrum for $a_{ph}(\lambda)$ and spectral slopes for $a_{dg}(\lambda)$ and $b_{bp}(\lambda)$, which are typically expressed as an exponential and a power-law, respectively [16]. Given the importance of understanding the spectral shapes of CDOM absorption and particulate backscattering, the spectral shape of these IOPs would ideally be varied and fit the same as the magnitude parameters that are in the inversion algorithm. However, SAAs such as GIOP are typically applied to satellite measurements with only 5–7 wavelengths. This introduces challenges when the number of fitted parameters is similar to the number of wavelengths [17], and indeed when all five parameters — three related to magnitude and two related to shape — are fit by GIOP, the results need to be tightly bound to a narrow dynamic range lest they become unrealistic.

The Bayesian approach [18,19] provides a methodology to overcome these challenges. In this framework, prior distributions for each of the fitted parameters are defined, and the algorithm minimizes the sum of the deviations between calculated and observed $R_{rs}(\lambda)$ and between the prior and best-fit parameters. In other words, the model will deviate from the best-guess prior solution only to the extent that such a deviation results in a commensurate decrease in model error. The GIOP model (Section 2) is augmented with this Bayesian approach (Section 3) and applied to both a reference in situ dataset and a satellite scene (Section 4). Bayesian-GIOP performs better on retrieving shape and magnitude parameters than a 5-variable GIOP fit, and equally well to a standard 3-variable fit, while greatly expanding the predicted variability in these parameters at smaller scales (Section 5). With the current and upcoming development of hyperspectral ocean color satellites, use of a Bayesian implementation in inversion algorithms such as GIOP will become more important to set realistic but flexible constraints on IOP ranges (Section 6).

2. GIOP model

SAAs such as GIOP exploit a simplified expression of the radiative transfer equation to invert measured $R_{rs}(\lambda)$ into corresponding estimates of absorption and backscattering coefficients. This simplification is expressed via a forward model that describes the relationships between the desired IOPs and observed $R_{rs}(\lambda)$ [8]. Total absorption $a_{tot}(\lambda)$ is the sum of absorption from seawater ($a_w(\lambda)$; m^{-1}), phytoplankton, and detritus/gelbstoff:

$$a_{tot}(\lambda) = a_w(\lambda) + a_{ph}(443)a_{ph}^*(\lambda) + a_{dg}(443)a_{dg}^*(\lambda) \qquad (1)$$

where $a_w(\lambda)$ is known [20] and the absorptions of phytoplankton, $a_{ph}(\lambda)$ (m^{-1}), and detritus/gelbstoff, $a_{dg}(\lambda)$ (m^{-1}), are represented as a magnitude at a reference wavelength of 443 nm and dimensionless spectral shapes. In its default form, GIOP prescribes a dynamically-assigned spectral shape to $a_{ph}^*(\lambda)$ from Ref. [21] and to $a_{dg}^*(\lambda)$ as [21]

$$a_{dg}^* = e^{S_{dg}(\lambda-443)}, \qquad (2)$$

where S_{dg} is an exponential slope (nm^{-1}) [22]. Other methods to assign $a_{ph}^*(\lambda)$ (Refs. [23] and [24]) were also tested, but did not change any of the conclusions of this manuscript. Total backscatter is the sum of backscatter from seawater ($b_{bw}(\lambda)$; m^{-1}) and particulates:

$$b_{b,tot}(\lambda) = b_{bw}(\lambda) + b_{bp}(555)b_{bp}^*(\lambda), \tag{3}$$

where $b_{bw}(\lambda)$ is known [25] and the phytoplankton/detrital backscattering is similarly separated into a magnitude at a reference wavelength, commonly 555 nm, and a dimensionless power-law controlled by the slope factor η (e.g., Ref. [26]):

$$b_{bp}^*(\lambda) = \left(\frac{\lambda}{555}\right)^\eta. \tag{4}$$

Spectrally-resolved a_{tot} and $b_{b,tot}$ relate to sub-surface reflectances $r_{rs}(\lambda)$ (sr^{-1}) as:

$$u(\lambda) = \frac{b_{b,tot}(\lambda)}{a_{tot}(\lambda) + b_{b,tot}(\lambda)} \tag{5a}$$

$$r_{rs}(\lambda) = g_1 u(\lambda) + g_2 u(\lambda)^2, \tag{5b}$$

with commonly assigned values $g_1 = 0.0949$ and $g_2 = 0.0794$ [27]. The resulting water-leaving reflectance $R_{rs}(\lambda)$ is then approximated as [28]:

$$R_{rs}(\lambda) = \frac{0.52 r_{rs}(\lambda)}{1 - 1.7 r_{rs}(\lambda)}. \tag{6}$$

In this form, the GIOP forward model includes six unknown parameters. Three are magnitude quantities encompassing absorption from phytoplankton ($a_{ph}(443)$), absorption from detritus and gelbstoff ($a_{dg}(443)$), and backscattering from particulates ($b_{bp}(555)$). Two are single shape parameters controlling the variation with wavelength of gelbstoff and detrital absorption (S_{dg}) and particulate backscatter (η). The last is chlorophyll-a, which is used to dynamically assign the spectral shape $a_{ph}^*(\lambda)$ using Ref. [21]. Heritage SAAs, such as GIOP-DC, do not attempt to fit a solution for the shape parameters, but instead define them either dynamically or via constant values. GIOP-DC and Ref. [23] assign constant values of 0.018 and 0.02061 nm^{-1} to S_{dg}, respectively. An alternative dynamic approach [28], used here, is to define:

$$S_{dg} = 0.015 + 0.002 \left(0.6 + r_{rs}(443)/r_{rs}(555)\right)^{-1}. \tag{7}$$

GIOP-DC dynamically assigns η as [28]:

$$\eta = 2 \times \left(1 - 1.2 e^{-0.9 r_{rs}(443)/r_{rs}(555)}\right), \tag{8}$$

with $r_{rs}(\lambda)$ calculated from the observed $R_{rs}(\lambda)$ as in Eq. (6), although constant values for η are also commonly chosen (e.g. Reference [23]).

Once spectral shapes are assigned in the forward model framework described above, only the magnitudes remain unknown. An inverse solution is then sought where, for each observed $R_{rs}(\lambda)$, the three magnitude parameters are then varied to produce a best-fit $R_{rs}(\lambda)$ using standard minimization techniques, such as Levenberg-Marquardt, that minimizes relative error, expressed as a fraction of expected variance due to measurement uncertainty ($\sigma_{Rrs}(\lambda)^2$) [8]:

$$\chi_{rel}^2 = \sum \frac{\left(R_{rs,calc.}(\lambda) - R_{rs,obs.}(\lambda)\right)^2}{\sigma_{Rrs}(\lambda)^2}. \tag{9}$$

The above approach presents an over-constrained solution space, where five or more $R_{rs}(\lambda)$ are typically used to retrieve estimates of three IOPs, namely $a_{ph}(443)$, $a_{dg}(443)$, and $b_{bp}(555)$.

3. Bayesian-GIOP

The Bayesian approach utilizes priors, and probability distributions about those priors, as an additional metric to control the fitting of parameters [18,19]. The narrowness or broadness of each probability distribution sets the allowable range of each parameter to be fit. Any best-fit algorithm can be framed in a Bayesian sense; for example, most standard best-fit algorithms, such as the approach presented in Section 2, are Bayesian where each fitted parameter has a maximally uninformative prior ranging from $[-\infty, +\infty]$. A standard bounded algorithm can also be understood as Bayesian, where the prior likelihood for each parameter is constant within the bounded region and zero elsewhere. In most applications, however, Bayesian priors are smoothly distributed functions, the most common and general of which is a Gaussian distribution.

The only addition necessary to make an algorithm "Bayesian" is to incorporate the prior distributions into the function to be minimized. When these distributions are Gaussian, this is conveniently done by adding a term to the error equation quantifying the discrepancy between the solution and the prior:

$$\chi^2_{Bayes} = (R_{rs,calc.}(\lambda) - R_{rs,obs.}(\lambda))^T S_R^{-1}(R_{rs,calc.}(\lambda) - R_{rs,obs.}(\lambda)) + (x - x_p)^T S_p^{-1}(x - x_p), \quad (10)$$

where a superscript ^{T} denotes a matrix transpose. Diagonal elements of the noise covariance matrix S_R give the expected variance, $\sigma_{Rrs}(\lambda)^2$, due to uncertainty in each $R_{rs}(\lambda)$ observation, and off-diagonal elements incorporate co-variances between uncertainties at different wavelengths, which could, for example, be caused by upstream errors in an atmospheric correction term [29]. When off-diagonal components of S_R are zero, the first term in Eq. (10) is equivalent to χ^2_{rel} (Eq. (9)), just expressed in matrix form. The second term in Eq. (10) is a penalty for deviations of the fitted parameters x from the prior maximum likelihood x_p, scaled to the prior covariance matrix S_p.

A crucial step in any Bayesian implementation is to carefully define the prior distributions, since these can play a leading role in determining the final output, and are what distinguish a Bayesian inversion. The Bayesian-GIOP model is therefore separated into two steps. In the first step, GIOP is run as in Section 2 to calculate the spectral shape parameters S_{dg} and η (i.e., from Eq. (7) and (8)) and estimate the most likely magnitude parameters $a_{ph}(443)$, $a_{dg}(443)$, and $b_{bp}(555)$ by minimizing the relative error (Eq. (9)). The prior most likely values are then defined as:

$$x_p = [a_{ph}(443), a_{dg}(443), b_{bp}(555), S_{dg}, \eta]. \quad (11)$$

The uncertainty covariance matrix for these three parameters is calculated as $J^{-1}S_R J^{-1,T}$, where J is the Jacobian matrix. It is less apparent how to derive proper prior uncertainties for shape parameters S_{dg} and η. Here we suggest using static standard deviations of 0.001 nm^{-1} and 0.1, respectively. The full prior uncertainty matrix is then:

$$S_p = \begin{bmatrix} & & & 0 & 0 \\ [& J^{-1}S_R J^{-1,T} &] & 0 & 0 \\ & & & 0 & 0 \\ 0 & 0 & 0 & 0.001^2 & 0 \\ 0 & 0 & 0 & 0 & 0.1^2 \end{bmatrix} \quad (12)$$

A subsequent Bayesian implementation of GIOP is run for all five parameters, minimizing the Bayesian error (Eq. (10)) using x_p and S_p defined from the three-parameter standard GIOP run.

4. Data

Data to ground-truth this approach are taken from the NASA bio-Optical Marine Algorithm Dataset (NOMAD) [30,31], a collection of high-quality in situ optical data for use in algorithm development. NOMAD data include measurements of $R_{rs}(\lambda)$, $a_{ph}(\lambda)$, and $b_{bp}(\lambda)$ at multiple wavelengths that closely correspond to those measured by satellite data. For this study a subset of 86 locations from NOMAD (Fig. 1) are used, where $R_{rs}(\lambda)$, $a_{ph}(\lambda)$, and $b_{bp}(\lambda)$ are all measured at six wavelengths corresponding to those observed by the SeaWiFS satellite, and where a corresponding measurement of Chl is taken either from chlorophyll fluorometry or high-performance liquid chromatography (HPLC).
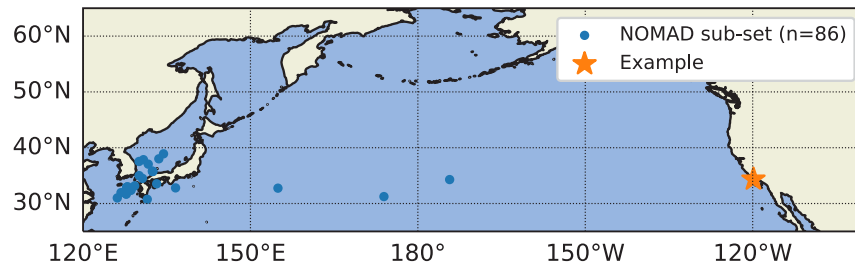

Fig. 1. Map of NOMAD sites used here (blue dots), with the example site considered in more detail as a gold star.

For a spatial analysis, the GIOP methods outlined above are also applied to Level 2 satellite imagery obtained from the NASA Ocean Biology Processing Group (OBPG; https://oceancolor.gsfc.nasa.gov). These images include $R_{rs}(\lambda)$ and have been geo-located and atmospherically corrected following standard NASA protocols. An overpass scene of the Gulf of Maine, off the northeastern coast of the USA, from MODIS-Aqua on September 29, 2018 is used here.

5. Results

The fidelity of algorithms describing the spectral shape of CDOM and detrital absorption and particulate backscatter can be tested using a subset of the NOMAD dataset with $a_{dg}(\lambda)$ and $b_{bp}(\lambda)$ at all six SeaWiFS wavelengths. Taking all wavelengths together, the two-parameter solution for $a_{dg}(\lambda)$ (Eq. (2)) resulted in misfits generally (87%) under 5%, and misfits for the two-parameter solution for $b_{bp}(\lambda)$ (Eq. 2) were virtually always (97%) under 2%. Although these misfits both had spectral shapes, suggesting biases in these functional forms, the comparatively low misfit results point to the overall usefulness of this approach. The range of S_{dg} calculated from this exercise (10–90th percentiles from 0.011–0.016 nm^{-1}) supports the choice to use a dynamic, rather than constant, S_{dg} value even in the 3-variable GIOP approach. However, in many cases $R_{rs}(\lambda)$ predicted from GIOP was significantly different from the observations (>25%), pointing to limitations in this method (see Section 6).

GIOP was applied to the NOMAD sub-dataset described in Section 4 in (1) a 3-variable configuration where the shape parameters S_{dg} and η are calculated and the three magnitude parameters $a_{ph}(443)$, $a_{dg}(443)$, and $b_{bp}(555)$ are fit, which represents its current mode of operation (Section 2), (2) a 5-variable configuration where the shape parameters S_{dg} and η are also fit using the same model, and (3) Bayesian-GIOP where all five parameters are fit using priors derived from the 3-variable configuration (Section 3). The retrieval of each of the five parameters—$a_{ph}(443)$, $a_{dg}(443)$, $b_{bp}(555)$, S_{dg}, and η—can then be compared with measurements from the in situ dataset. The fidelity of an SAA's forward model and its parameterization for inversion can be partially approximated by how closely the calculated $R_{rs}(\lambda)$ align with the measured values. Typical uncertainties for R_{rs} measurements are about 5% [32,33]. An SAA that appropriately retrieves
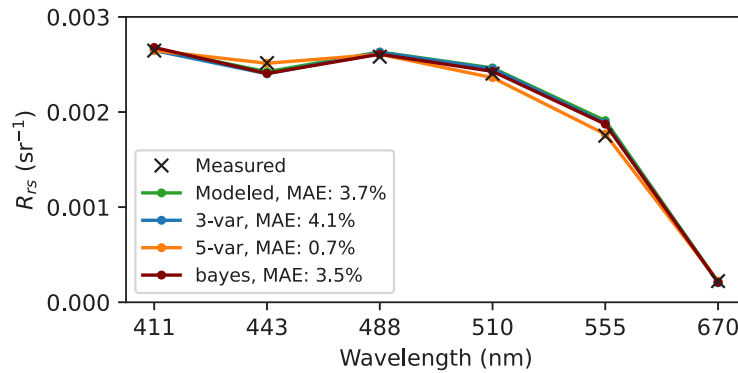
Fig. 2. R_{rs} (black x's) measured at a Southern California site (see gold star in Fig. 1). Superimposed are the modeled results using measured values (green), 3-variable GIOP (blue), 5-variable GIOP (orange), and Bayesian-GIOP (maroon).

IOPs should therefore ideally result in a final mean average error (MAE) of approximately the same percentage. Here MAE is calculated as [34]:

$$MAE = \exp\left(\frac{\sum_i^N |\log(M_i) - \log(O_i)|}{n}\right) - 1, \tag{13}$$

where M_i and O_i are a set of modeled results and observations, respectively. Algorithm results with a MAE much higher than 5% indicate an incorrect representation of bio-optical and geophysical representation of the model. Crucially, however, a MAE substantially lower than an expected uncertainty of 5% is also not desirable, as this suggests that the model is fitting to noise in the data.

An example result, from a station off the coast of Southern California (Fig. 1, gold star), demonstrates expected and lower-than-expected error (Fig. 2). The modeled $R_{rs}(\lambda)$ using measured parameters (green line) has a 3.7% MAE with respect to the observations (black x's), which sets the expected error due to sensor noise, observational uncertainties, and modeling inaccuracies (Table 1). The 3-variable GIOP implementation (blue line) has a very similar error of 4.1%, but the 5-variable version (orange) has a MAE, 0.7%, that is much lower than expected. This low error is not surprising, considering that an algorithm with five free parameters fit to a dataset of six wavelengths only has one free variable, but does suggest that the model is over-fitting to the data. The Bayesian implementation (maroon line) has a MAE similar to that of the 3-variable model, of 3.5%.

Table 1. Example results for Southern California site (Fig. 1, gold star). Note that NOMAD results are measured values, while other rows present the results of different model runs as described in the text.

	MAE	$a_{ph}(443)$; m^{-1}	$a_{dg}(443)$; m^{-1}	$b_{bp}(555)$; m^{-1}	S_{dg}; nm^{-1}	η
NOMAD	3.7%	0.96	0.049	0.0029	0.014	1.40
3-var	4.1%	1.03 ± 0.09	0.041 ± 0.003	0.0027 ± 0.0001	0.016	1.34
5-var	0.7%	1.14 ± 0.16	0.007 ± 0.004	0.0014 ± 0.0004	0.052 ± 0.019	−0.39 ± 0.96
Bayes	3.5%	1.00 ± 0.07	0.043 ± 0.003	0.0026 ± 0.0001	0.015 ± 0.001	1.26 ± 0.08
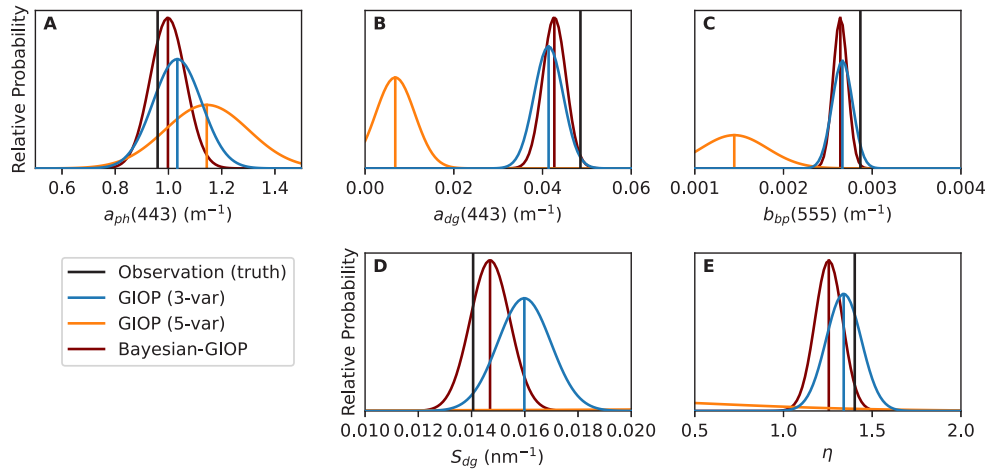
Fig. 3. Best-fit parameters and standard deviations for the 3-variable (blue), 5-variable (orange), and Bayesian (maroon) approach. Observations are indicated by the black line. Standard deviations for S_{dg} (D) and η (E) from the 3-variable approach are assumed to be 0.001 nm^{-1} and 0.1, respectively, as described in the text and in Eq. (12). Note that for several of the 5-variable results the best-fit value lies outside the range shown.

That the 5-variable GIOP is fitting the noise is clearly seen in the best-fit retrievals for each approach (Table 1; Fig. 3). The 5-variable approach, although resulting in a much better MAE, has considerably worse IOP results, including with unrealistic values for the shape parameters S_{dg} and η, and substantially higher uncertainty on each predicted IOP.

The Bayesian implementation improves on the predictions of three of the five variables compared with the 3-variable implementation. The prediction of phytoplankton absorption $a_{ph}(443)$ decreased from 1.03 m^{-1} (3-var) to 1.00 m^{-1} (Bayes), compared with a measured value of 0.96 m^{-1}; the detrital and gelbstoff absorption $a_{dg}(555)$ increased from 0.041 m^{-1} (3-var) to 0.043 m^{-1} (Bayes), compared with a measured value of 0.049 m^{-1}; and the spectral slope of detrital/gelbstoff absorption S_{dg} decreased from a calculated value of 0.016 nm^{-1} (3-var) to a fit value of 0.015 nm^{-1} (Bayes), compared with a measured value of 0.014 nm^{-1}. In this example the Bayesian approach yielded a comparable prediction for the particulate backscatter $b_{bp}(555)$ to the 3-variable model, but yielded a worse prediction for the spectral particulate backscattering slope η, which decreased from a calculated value of 1.34 (3-var) to a fitted value of 1.26 (Bayes), compared with a value derived from observations of 1.40.

The results from this single-station case study are broadly broadly consistent with the results from the full sub-set of NOMAD data used here (Fig. 4). The average (± st. dev.) MAE is 4.8% (±2.9%) for the 3-variable, 1.0% (±0.8%) for the 5-variable, and 4.4% (±2.7%) for the Bayesian retrievals. The 3-variable and Bayesian-GIOP versions are essentially equivalent for the magnitude parameters $a_{ph}(443)$, $a_{dg}(443)$, and $b_{bp}(555)$ (A–C), with a slight low bias for $a_{dg}(443)$ and a high bias for $b_{bp}(555)$, with the 3-variable approach giving slightly better biases for S_{dg} (D) and the Bayesian-GIOP approach providing less biased η estimates (E). The 5-variable approach provides significantly worse results for all variables, with several results biased by more than a factor of 3.

For each variable, the bias, MAE, and wins were calculated for each retrieval method (Table 2), where

$$\text{bias} = \exp\left(\frac{\sum_i^n \log(M_i) - \log(O_i)}{n}\right) - 1 \tag{14}$$
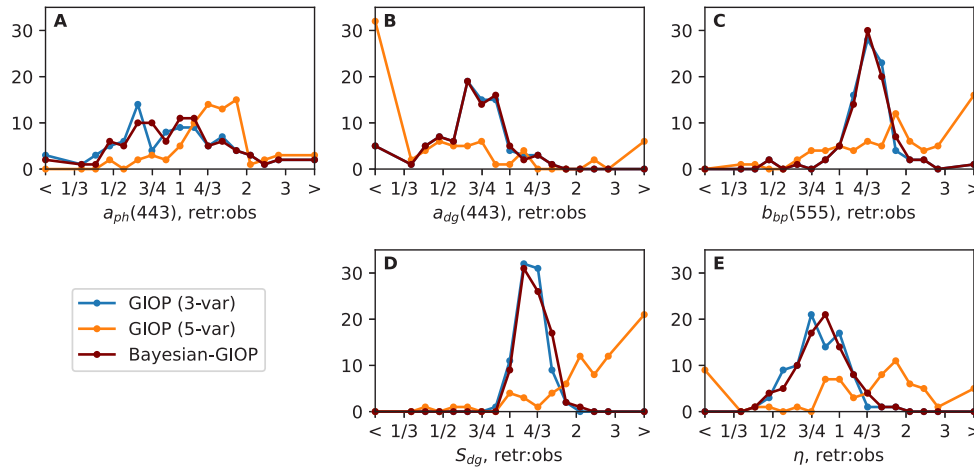
Fig. 4. Best-fit parameters for the 3-variable (blue), 5-variable (orange), and Bayesian (maroon) approach expressed as ratios from the observations. For each histogram, "<" and ">" refer to ratios of less than 1/3 and greater then 3, respectively.

Table 2. Statistics for full sub-set of NOMAD data

variable	metric	3-var	5-var	Bayes
	bias	−9%	+41%	−5%
$a_{ph}(443)$	MAE	59%	55%	53%
	wins	43%	32%	75%
	bias	−34%	−54%	−35%
$a_{dg}(443)$	MAE	59%	214%	60%
	wins	79%	8%	63%
	bias	+35%	+87%	+37%
$b_{bp}(555)$	MAE	43%	112%	44%
	wins	74%	23%	53%
	bias	+24%	+132%	+28%
S_{dg}	MAE	25%	144%	28%
	wins	83%	5%	61%
	bias	−19%	+10%	−16%
η	MAE	32%	104%	30%
	wins	60%	22%	68%

and wins are calculated as the fraction of head-to-head contests where each method provides a more accurate estimate (note that the total percentage in this row is 150%; see Ref. [34]). The 3-variable and Bayes retrievals consistently perform better, in virtually every metric, than the 5-variable retrieval. Between the two of them, the 3-variable approach performs better (smallest bias and MAE and largest percentage of wins) for $b_{bp}(555)$ and S_{dg}, the Bayesian approach performs better for $a_{ph}(443)$ and η, and they are essentially equivalent for $a_{dg}(443)$, with a slight edge for the 3-variable approach in wins.

To illustrate the spatial information gained from the Bayesian approach, each of the three GIOP implementations is applied to a scene from MODIS Aqua described in Section 4 (Fig. 5). The Bayesian version does not substantially change the magnitude or spatial pattern of any of the
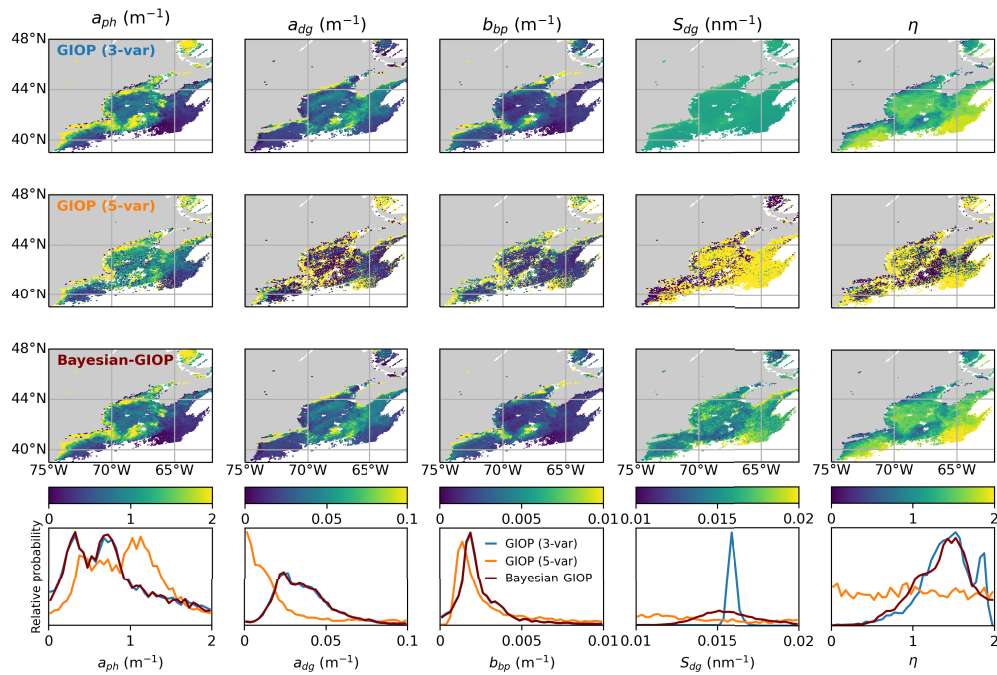
Fig. 5. Example from a Level 2 MODIS-Aqua overpass scene for the Gulf of Maine. First three rows give the GIOP retrieval of the five different parameters using the 3-variable, 5-variable, and Bayesian methods described in the text. Histograms on the bottom show the distribution of each parameter within the kernel for the 3-variable (blue), 5-variable (orange), and Bayesian (maroon) results.

magnitude variables $a_{ph}(443)$, $a_{dg}(443)$, or $b_{bp}(555)$ with respect to the 3-variable version. The range of η is somewhat increased, with more variability at smaller scales, and the range of S_{dg} is both substantially larger and at larger scales predicts opposite spatial trends, with, for example, substantially larger values in the southeast region where the 3-variable version calculates slightly a slightly smaller S_{dg}. The 5-variable version, by contrast, predicts on average significantly higher $a_{ph}(443)$, substantially lower average $a_{dg}(443)$, an overall decrease in $b_{bp}(555)$, and a much larger range of S_{dg} and η than either of the other algorithms. The over-fit 5-variable version also exhibits considerably more spatial spikes in the data than the other versions.

6. Discussion

Use of the Bayesian, as opposed to the standard 3-variable SAA, primarily affects the two shape parameters, S_{dg} and η, most notably leading to a broader range of variability in S_{dg}. This expanded variability is supported by historical *in situ* measurements, which suggest variation of S_{dg} from 0.012 to 0.017 nm^{-1} in the Gulf of Maine [35]. Bayesian-GIOP therefore provides a potential for stronger testable hypotheses as to the variation in composition of both phytoplankton communities (since η is retrieved rather than calculated) and CDOM than the 3-variable retrieval.

Higher spectral resolution may also be expected to increase the accuracy of all of the GIOP approaches here. In particular, the 5-variable approach resulted in highly inaccurate retrievals of the shape parameters S_{dg} and η, which were ill-constrained using only multispectral data. Large inaccuracies in S_{dg} and η then propagate into retrievals of the associated magnitude parameters $a_{dg}(443)$ and $b_{bp}(555)$. The added spectral resolution of PACE, a satellite which will provide 2.5 nm spectral steps and 5 nm resolution throughout the wavelength range 340–890 nm [36]

may lead to more highly constrained solutions for these shape parameters. However, even in this context the Bayesian approach will be useful, as this method provides more flexible bounds to allowable parameter values than many other algorithms, which typically apply either free ($-\infty$ to ∞) or hard boundaries. Furthermore, the increase in spectral resolution will open the door for additional fitted parameters such as multiple different types of CDOM [37] and phytoplankton [38–41]. Past work has shown that a Bayesian implementation of GIOP shows skill in deciphering multiple phytoplankton species [19]. In the future, an expanded GIOP framework encompassing a wider number and range of optical constituents, perhaps varying with region or season, may become both possible and necessary; such a framework would benefit from a Bayesian approach.

However, higher spectral resolution will not lead to more accurate IOP retrievals if the underlying algorithm is inaccurate. In the sub-set of NOMAD data used here, we calculated the expected $R_{rs}(\lambda)$ using Eqs. (1–6) and compared to observed $R_{rs}(\lambda)$. Nearly half (38/86, or 44%) of the modeled $R_{rs}(\lambda)$ had greater than 25% MAE with respect to observations. This is concerning, and points to additional work that needs to be done on the underlying optical assumptions within any SAA configuration. For example, the assumption that a single factor controls the spectral slope of CDOM [42,43] or particulate backscatter [44] may cause significant retrieval errors. An increase in spectral resolution from observations, such as will be possible from PACE, may allow for more realistic spectral shapes for both $a_{dg}(\lambda)$ and $b_{bp}(\lambda)$.

7. Conclusions

This manuscript demonstrates a method to use prior knowledge of key optical parameters as constraints in a Bayesian implementation of a common remote sensing inversion algorithm, GIOP. The Bayesian approach allows elements related to the spectral shape of absorption and backscattering IOPs to be retrieved as part of the fitted algorithm, rather than calculated according to empirical formulae. Use of Bayesian-GIOP therefore alleviates a current issue with the GIOP algorithm, which is that in the standard formulation fitting shape parameters results in retrieved values that are either nonsensical or are completely determined by specified boundary constraints, which currently serve as a cruder version of a Bayesian prior. This approach is also expected to generalize well to future GIOP-like algorithms involving a larger number of IOPs, which will dramatically improve the fidelity of data products from future, hyperspectral missions.

Disclosures. The authors declare no conflicts of interest.

Data Availability. NOMAD data used in this paper are presented by Ref. [30], further processed by Ref. [31], and available at [45] . The MODIS-Aqua satellite scene is a product of the Ocean Biology Processing Group within the Ocean Ecology Laboratory located at NASA Goddard Space Flight Center and is available at [46]. Code supporting this manuscript is available at [47]. The Bayesian Levenberg-Marquardt optimization routine is currently being implemented within NASA's GIOP framework. It is due to be released as part of the Ocean Color Software Suite (OCSSW) in 2023.

References

1. C. R. McClain, "A decade of satellite ocean color observations," Annu. Rev. Mar. Sci **1**(1), 19–42 (2009).
2. J. E. O'Reilly, S. Maritorena, B. G. Mitchell, D. A. Siegel, K. L. Carder, S. A. Garver, M. Kahru, and C. McClain, "Ocean color chlorophyll algorithms for SeaWiFS," J. Geophys. Res.: Oceans **103**(C11), 24937–24953 (1998).
3. C. Hu, Z. Lee, and B. Franz, "Chlorophyll algorithms for oligotrophic oceans: A novel approach based on three-band reflectance difference," J. Geophys. Res.: Oceans **117**(C1), 1 (2012).
4. D. Stramski, R. A. Reynolds, M. Babin, S. Kaczmarek, M. R. Lewis, R. Röttgers, A. Sciandra, M. Stramska, M. Twardowski, B. Franz, and H. Claustre, "Relationships between the surface concentration of particulate organic carbon and optical properties in the eastern South Pacific and eastern Atlantic Oceans," Biogeosciences **5**(1), 171–201 (2008).
5. H. R. Gordon, G. C. Boynton, W. M. Balch, S. B. Groom, D. S. Harbour, and T. J. Smyth, "Retrieval of coccolithophore calcite concentration from SeaWiFS imagery," Geophys. Res. Lett. **28**(8), 1587–1590 (2001).

6. W. Balch, H. R. Gordon, B. Bowler, D. Drapeau, and E. Booth, "Calcium carbonate measurements in the surface global ocean based on moderate-resolution imaging spectroradiometer data," J. Geophys. Res.: Oceans **110**(C7), C07001 (2005).

7. J. E. O'Reilly and P. J. Werdell, "Chlorophyll algorithms for ocean color sensors-OC4, OC5 & OC6," Remote. Sens. Environ. **229**, 32–47 (2019).

8. P. J. Werdell, B. A. Franz, S. W. Bailey, G. C. Feldman, E. Boss, V. E. Brando, M. Dowell, T. Hirata, S. J. Lavender, Z. Lee, H. Loisel, S. Maritorena, F. Mélin, T. S. Moore, T. J. Smyth, D. Antoine, E. Devred, O. H. F. d'Andon, and A. Mangin, "Generalized ocean color inversion model for retrieving marine inherent optical properties," Appl. Opt. **52**(10), 2019–2037 (2013).

9. A. V. Vähätalo and R. G. Wetzel, "Photochemical and microbial decomposition of chromophoric dissolved organic matter during long (months–years) exposures," Mar. Chem. **89**(1-4), 313–326 (2004).

10. J. B. Clark, P. Neale, M. Tzortziou, F. Cao, and R. R. Hood, "A mechanistic model of photochemical transformation and degradation of colored dissolved organic matter," Mar. Chem. **214**, 103666 (2019).

11. T. Kostadinov, D. Siegel, and S. Maritorena, "Retrieval of the particle size distribution from satellite ocean color observations," J. Geophys. Res.: Oceans **114**(C9), C09015 (2009).

12. W. H. Slade and E. Boss, "Spectral attenuation and backscattering as indicators of average particle size," Appl. Opt. **54**(24), 7264–7277 (2015).

13. H. Loisel, J.-M. Nicolas, A. Sciandra, D. Stramski, and A. Poteau, "Spectral dependency of optical backscattering by marine particles from satellite remote sensing of the global ocean," J. Geophys. Res.: Oceans **111**(C9), C09024 (2006).

14. T. Kostadinov, D. Siegel, and S. Maritorena, "Global variability of phytoplankton functional types from space: Assessment via the particle size distribution," Biogeosciences **7**(10), 3239–3257 (2010).

15. A. L. Whitmire, W. S. Pegau, L. Karp-Boss, E. Boss, and T. J. Cowles, "Spectral backscattering properties of marine phytoplankton cultures," Opt. Express **18**(14), 15073–15093 (2010).

16. P. J. Werdell, L. I. McKinna, E. Boss, S. G. Ackleson, S. E. Craig, W. W. Gregg, Z. Lee, S. Maritorena, C. S. Roesler, C. S. Rousseaux, D. Stramski, J. M. Sullivan, M. S. Twardowski, M. Tzortziou, and Z. Xiaodong, "An overview of approaches and challenges for retrieving marine inherent optical properties from ocean color remote sensing," Prog. Oceanogr. **160**, 186–212 (2018).

17. B. Cael, K. Bisson, E. Boss, and Z. K. Erickson, "How many independent quantities can be extracted from ocean color?" Limnology and Oceanography Letters (2023).

18. C. D. Rodgers, *Inverse methods for atmospheric sounding: Theory and practice*, vol. 2 (World scientific, 2000).

19. 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).

20. R. M. Pope and E. S. Fry, "Absorption spectrum (380–700 nm) of pure water. II. integrating cavity measurements," Appl. Opt. **36**(33), 8710–8723 (1997).

21. 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).

22. A. Bricaud, A. Morel, and L. Prieur, "Absorption by dissolved organic matter of the sea (yellow substance) in the UV and visible domains," Limnol. Oceanogr. **26**(1), 43–53 (1981).

23. 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).

24. A. M. Ciotti and A. Bricaud, "Retrievals of a size parameter for phytoplankton and spectral light absorption by colored detrital matter from water-leaving radiances at seawifs channels in a continental shelf region off brazil," Limnol. Oceanogr.: Methods **4**(7), 237–253 (2006).

25. X. Zhang, L. Hu, and M.-X. He, "Scattering by pure seawater: Effect of salinity," Opt. Express **17**(7), 5698–5710 (2009).

26. H. R. Gordon, M. R. Lewis, S. D. McLean, M. S. Twardowski, S. A. Freeman, K. J. Voss, and G. C. Boynton, "Spectra of particulate backscattering in natural waters," Opt. Express **17**(18), 16192–16208 (2009).

27. 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.: Atmos. **93**(D9), 10909–10924 (1988).

28. 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).

29. A. Ibrahim, B. A. Franz, A. M. Sayer, K. Knobelspiesse, M. Zhang, S. W. Bailey, L. I. McKinna, M. Gao, and P. J. Werdell, "Optimal estimation framework for ocean color atmospheric correction and pixel-level uncertainty quantification," Appl. Opt. **61**(22), 6453–6475 (2022).

30. P. J. Werdell and S. W. Bailey, "An improved in-situ bio-optical data set for ocean color algorithm development and satellite data product validation," Remote. Sens. Environ. **98**(1), 122–140 (2005).

31. A. Valente, S. Sathyendranath, and V. Brotas, *et al.*, "A compilation of global bio-optical in situ data for ocean-colour satellite applications," Earth Syst. Sci. Data **8**(1), 235–252 (2016).

32. L. I. McKinna, I. Cetinić, A. P. Chase, and P. J. Werdell, "Approach for propagating radiometric data uncertainties through NASA ocean color algorithms," Front. Earth Sci. **7**, 176 (2019).

33. M. Zhang, A. Ibrahim, B. A. Franz, Z. Ahmad, and A. M. Sayer, "Estimating pixel-level uncertainty in ocean color retrievals from modis," Opt. Express **30**(17), 31415–31438 (2022).

34. B. N. Seegers, R. P. Stumpf, B. A. Schaeffer, K. A. Loftin, and P. J. Werdell, "Performance metrics for the assessment of satellite data products: An ocean color case study," Opt. Express **26**(6), 7404–7422 (2018).

35. E. J. D'Sa, R. G. Steward, A. Vodacek, N. V. Blough, and D. Phinney, "Determining optical absorption of colored dissolved organic matter in seawater with a liquid capillary waveguide," Limnol. Oceanogr. **44**(4), 1142–1148 (1999).

36. P. J. Werdell, M. J. Behrenfeld, P. S. Bontempi, E. Boss, B. Cairns, G. T. Davis, B. A. Franz, U. B. Gliese, E. T. Gorman, O. Hasekamp, K. D. Knobelspiesse, A. Mannino, J. V. Martins, C. R. McClain, G. Meister, and L. A. Remer, "The Plankton, Aerosol, Cloud, ocean Ecosystem mission: Status, science, advances," Bull. Am. Meteorol. Soc. **100**(9), 1775–1794 (2019).

37. D. Siegel, S. Maritorena, N. Nelson, D. Hansell, and M. Lorenzi-Kayser, "Global distribution and dynamics of colored dissolved and detrital organic materials," J. Geophys. Res.: Oceans **107**(C12), 21-1–21-14 (2002).

38. A. Bracher, H. A. Bouman, and R. J. Brewin, *et al.*, "Obtaining phytoplankton diversity from ocean color: A scientific roadmap for future development," Front. Mar. Sci. **4**, 55 (2017).

39. A. Chase, E. Boss, I. Cetinić, and W. Slade, "Estimation of phytoplankton accessory pigments from hyperspectral reflectance spectra: Toward a global algorithm," J. Geophys. Res.: Oceans **122**(12), 9725–9743 (2017).

40. C. B. Mouw, N. J. Hardman-Mountford, S. Alvain, A. Bracher, R. J. Brewin, A. Bricaud, A. M. Ciotti, E. Devred, A. Fujiwara, T. Hirata, T. Hirawake, T. S. Kostadinov, S. Roy, and J. Uitz, "A consumer's guide to satellite remote sensing of multiple phytoplankton groups in the global ocean," Front. Mar. Sci. **4**, 41 (2017).

41. J. A. Schulien, A. Della Penna, P. Gaube, A. P. Chase, N. Haëntjens, J. R. Graff, J. W. Hair, C. A. Hostetler, A. J. Scarino, E. S. Boss, L. Karp-Boss, and M. J. Behrenfeld, "Shifts in phytoplankton community structure across an anticyclonic eddy revealed from high spectral resolution lidar scattering measurements," Front. Mar. Sci. **7**, 493 (2020).

42. J. R. Helms, A. Stubbins, J. D. Ritchie, E. C. Minor, D. J. Kieber, and K. Mopper, "Absorption spectral slopes and slope ratios as indicators of molecular weight, source, and photobleaching of chromophoric dissolved organic matter," Limnol. Oceanogr. **53**(3), 955–969 (2008).

43. B. Cael and E. Boss, "Simplified model of spectral absorption by non-algal particles and dissolved organic materials in aquatic environments," Opt. Express **25**(21), 25486–25491 (2017).

44. W. H. Slade, E. Boss, and C. Russo, "Effects of particle aggregation and disaggregation on their inherent optical properties," Opt. Express **19**(9), 7945–7959 (2011).

45. A. Valente, S. Sathyendranath, and V. Brotas, *et al.*, "A compilation of global bio-optical in situ data for ocean-colour satellite applications," PANGEA, (2015) https://doi.org/10.1594/PANGAEA.854832

46. NASA Ocean Biology Processing Group, Aqua MODerate Resolution Imaging Spectroradiometer (MODIS) Level-2 Ocean Color Data, 2018 Re-processing, *NASA Ocean Biology Distributed Active Archive Center*, [Accessed on 2021/04/15] https://oceancolor.gsfc.nasa.gov/cgi/browse.pl

47. Z. K. Erickson, Code repository, Zenodo, (2023). https://doi.org/10.5281/zenodo.7869394