Optimal Estimation Framework for Ocean Color Atmospheric Correction and Pixel-level Uncertainty Quantification

AMIR IBRAHIM,[1,*] BRYAN A. FRANZ,[1] ANDREW M. SAYER,[1,2] KIRK KNOBELSPIESSE,[1] MINWEI ZHANG,[1,3] SEAN BAILEY,[1] LACHLAN MCKINNA,[1,4] MENG GAO,[1,5] AND P. JEREMY WERDELL[1]

[1]*NASA Goddard Space Flight Center, Greenbelt, Maryland 20771, USA*
[2]*University of Maryland Baltimore County, Baltimore, Maryland 21250, USA*
[3]*Science Applications International Corporation, Reston, Virginia 20190, USA*
[4]*Go2Q Pyt Ltd., Buderim, QLD, Australia*
[5]*Science Systems and Applications, Inc., Lanham, Maryland 20706, USA*
**amir.ibrahim@nasa.gov*

Abstract: Ocean color remote sensing requires compensation for atmospheric scattering and absorption (aerosol, Rayleigh, and trace gases), referred to as atmospheric correction (AC). AC allows inference of parameters such as spectrally resolved remote sensing reflectance ($R_{rs}(\lambda)$; sr^{-1}) at the ocean surface from the top-of-atmosphere reflectance. Often, the uncertainty of this process is not fully explored. Bayesian inference techniques provide a simultaneous AC and uncertainty assessment via a full posterior distribution of the relevant variables, given the prior distribution of those variables and the radiative transfer (RT) likelihood function. Given uncertainties in the algorithm inputs, the Bayesian framework enables better constraints on the AC process by using the complete spectral information compared to traditional approaches that use only a subset of bands for AC. This paper investigates a Bayesian inference research method (Optimal Estimation, OE) for ocean color AC by simultaneously retrieving atmospheric and ocean properties using all visible and near-infrared spectral bands. The OE algorithm analytically approximates the posterior distribution of parameters based on normality assumptions and provides a potentially viable operational algorithm with a reduced computational expense. We developed a Neural Network (NN) RT forward model look-up-table-based emulator to increase algorithm efficiency further and thus speed up the likelihood computations. We then applied the OE algorithm to synthetic data and observations from the MODerate resolution Imaging Spectroradiometer (MODIS) on NASA's Aqua spacecraft. We compared the $R_{rs}(\lambda)$ retrieval and its uncertainty estimates from the OE method with in-situ validation data from the SeaWiFS Bio-optical Archive and Storage System (SeaBASS) and Aerosol Robotic Network Ocean Color (AERONET-OC) datasets. The OE algorithm improved $R_{rs}(\lambda)$ estimates relative to the NASA standard operational algorithm by improving all statistical metrics at 443, 555, and 667 nm. Unphysical negative $R_{rs}(\lambda)$, which often appear in complex water conditions, was reduced by a factor of 3. The OE-derived pixel-level $R_{rs}(\lambda)$ uncertainty estimates were also assessed relative to in-situ data and were shown to have skill.

1. Introduction

The atmospheric correction (AC) process in ocean color (OC) remote sensing involves separating and removing the atmospheric contributions (aerosol and gas scattering and absorption) and ocean surface signal from the spectral reflectances observed by a satellite radiometer at the top of the atmosphere (TOA) [1–4]. The science of OC aims to quantify and assess the biogeochemical properties of aquatic ecosystems by interpreting their visible water-leaving spectra. These spectral reflectance signals emerging from the water body primarily depend on the inherent optical properties (IOPs; absorption and scattering properties) of the biogeochemical constituents dissolved or suspended within the water column, in combination with the IOPs of seawater itself. These constituents include organic and inorganic hydrosols suspended in seawater, colored dissolved organic matter (CDOM), and photosynthetic pigments within phytoplankton. The primary heritage OC data product is the near-surface concentration of the photosynthetic pigment chlorophyll-a (Chl-a; mg m-3), which provides a convenient and widely-used proxy for phytoplankton biomass [5]. Phytoplankton biomass is an essential component of the Earth's carbon cycle, and producing climate-quality OC data records is generally necessary for Earth climate studies [6–8].

Chlorophyll-a concentrations are typically derived through an empirical relationship based on coincident in-situ observations of Chl-a and the aforementioned water-leaving radiometric signal, namely spectral remote sensing reflectances ($R_{rs}(\lambda)$; sr-1), which are the radiances exiting the water column normalized to downwelling surface irradiance. However, since the atmospheric radiance contribution to the TOA signal is typically between 85-90% of the total, a small uncertainty in the AC can lead to large uncertainties in the ocean radiances and derived OC products [3]. The Rayleigh scattering of the atmosphere is effectively known, based on assumed molecular properties [9], yet it can introduce additional uncertainties in the AC [10]. However, the aerosol signal must be inferred from the satellite observations since the aerosol type and concentration vary spatially and temporally in the atmosphere [11].

Inferring useful information from satellite-derived radiometry is accomplished by solving the inverse problem, which is ill-posed and under-constrained for AC [12,13]. The TOA reflectance of multi-spectral, single viewing sensors such as the Sea-viewing Wide Field-of-view Sensor (SeaWiFS) [14], MODerate resolution Imaging Spectroradiometer (MODIS) [15], and Visible Infrared Imaging Radiometer Suite (VIIRS) [16], contain less information than what is required to find an unambiguous solution to a complex Atmosphere-Ocean (AO) model. To address this issue, future NASA missions will dedicate more advanced instruments to increase the observed information. For example, NASA's Plankton, Aerosol, Cloud, ocean Ecosystem (PACE) mission will host three instruments that will measure the AO system with unprecedented spectral and angular information [17]. The primary instrument is the Ocean Color Instrument (OCI), which is being developed at the Goddard Space Flight Center (GSFC) and is a hyperspectral scanning radiometer that measures the light from 320 to 890 nm at 5-nm spectral resolution and 2.5-nm spectral sampling, and at seven discrete short-wave infrared (SWIR) channels: 940, 1,038, 1,250, 1,378, 1,615, 2,130, and 2,260 nm. The two other instruments, the Hyper-Angular Rainbow Polarimeter 2 (HARP2) and the Spectro-Polarimeter for Exploration (SPEXone), are aimed at studying aerosols and clouds and are multi-angular polarimeters (MAPs) developed and contributed by external partners. To employ OCI's unprecedented hyperspectral capabilities for ocean applications, an accurate AC process with capabilities beyond the current algorithm designed for multispectral sensors is necessary. One such OCI algorithm has been developed, which relies on the proven heritage AC capabilities of the NASA standard algorithm, extended to hyperspectral data and with added capabilities to seamlessly utilize the SWIR channels for AC in coastal and inland waters (e.g., the multiband AC (MBAC) algorithm) [18]. In addition to this OCI-only AC, the MAPs will provide more complex aerosol information to constrain the AC for OCI [19]. Thus, establishing a probabilistic framework that can combine the information from two or three independent instruments with different spatial and spectral resolutions, information content, and measurement uncertainty characteristics using a Bayesian framework is a logical next step to advance the AC performance and the quality of OC retrievals from PACE.

Deterministic (that is, non-stochastic) AC methods have been and are currently being used as the standard processing algorithms for satellite remote sensing of OC [3,4,11,20,21]. These methods maximize the likelihood (i.e., match radiative transfer prediction models to the observations) of the AC parameters such as the aerosol, surface, and ocean optical properties and often do not directly provide an estimate of the uncertainty on these parameters or consider the uncertainty in the algorithm inputs and parameters. The forward likelihood model is parametrized from radiative transfer simulations in a (pre-computed) look-up-table (LUT) for computational efficiency. These LUTs contain the modeled TOA reflectances for a pre-determined set of relevant parameters within a typical range.

NASA's current operational AC algorithm for OC sensors is based on Gordon and Wang [11], with the current implementation detailed in [20]. The algorithm determines and removes atmospheric (i.e., Rayleigh and aerosol) and surface (i.e., whitecaps, glint) reflectances through a LUT search of pre-computed reflectances as derived using vector radiative transfer (VRT) simulations. One LUT contains the spectral TOA Rayleigh reflectance for different geometries and surface wind speeds. The aerosol reflectance LUTs are parametrized for 80 different aerosol optical models representing the range of relative humidity (RH) and fine-mode volume fractions [22]. These models assume a complex refractive index and bimodal effective radius and variance for coarse and fine aerosol particles, determined from Aerosol Robotic Network (AERONET) observations [23,24]. The absorption coefficients of trace gases such as ozone, water vapor, oxygen, and methane are stored in LUTs and applied to compensate for atmospheric path absorption given the gas concentration and an assumed vertical profile. Ancillary information, including relative humidity, ozone and water vapor concentrations, and wind speed, are provided as auxiliary inputs to constrain the inversion.

The aforementioned models are explicitly parameterized to ensure that the inversion is not mathematically ill-posed and act as a constraint to reduce ambiguity and the potential for degenerate solutions. The aerosol optical models are assumed to be non- to weakly- absorbing and to have a fixed vertical profile. With these assumptions, only two pieces of information are needed for the AC: aerosol optical depth (AOD, i.e., loading) and spectral dependence (i.e., from the optical model), both of which can be determined using a pair of near-infrared (NIR) or SWIR wavelengths (dependent on the sensor). However, the presence of strongly absorbing aerosol types confounds this process, and the AC typically produces either underestimated or non-physical negative ocean radiances in the blue part of the spectrum . This is because the algorithm relies on the extrapolation of the model information determined from the longer NIR or SWIR wavelengths (where the ocean is dark) to the visible (where it is not). The spectral information in the longer wavelengths is insufficient to discern absorbing from non-absorbing aerosols, as they differ primarily in the shorter wavelengths and do not have a discriminating signature in the NIR. Thus, the solution can be ambiguous and aerosol absorption cannot be reliably inferred unless the algorithm is constrained by additional external information.

Pixel-level Uncertainty Quantification (UQ) is critical in assessing the fidelity of geophysical retrievals within the Earth system. UQ also allows for identifying issues and limitations in retrieval algorithms due to inherent modeling assumptions, measurement uncertainties, and gaps in knowledge and sources of uncertainties. Traditionally, uncertainties in R_{rs} are based on the reported average discrepancy between the satellite-derived and in-situ R_{rs} [27,28]. UQ has been attempted through various techniques such as Bayesian approaches [29,30], Monte Carlo simulations [31], or analytical error propagation of sensor random noise [32]. A new approach was developed to estimate pixel-level uncertainties for Sentinel-3 Ocean and Land Colour Imager (OLCI) based on an ensemble of neural network atmospheric correction models for coastal waters, showing an estimate of the R_{rs} uncertainty product that is feasible to apply operationally [33].

Because of the ill-posed nature of the problem, Bayesian approaches are well-suited for AC and indeed have been applied widely for aerosol [34,35], cloud [36,37], atmospheric trace gas profiling [38,39], and OC [29,40–42] retrievals. For a given model, the aerosol and ocean properties can be inferred, along with the associated uncertainties, in the form of a posterior distribution. Bayes theorem calculates conditional probabilities and updates a prior belief when new data (evidence) is introduced [43] such that $P(x|\, y_{obs}) \propto P(y_{obs}|x) \times P(x)$, where $P(x|\, y_{obs})$ is the posterior distribution or probability of the variables needed for AC, x, given the observed data, y_{obs}. The posterior distribution is proportional to the likelihood function, $P(y_{obs}$

$|x)$, and the prior probability of the variables $P(x)$. The likelihood function describes the probability of the observed TOA reflectance, y_{obs}, given the variables x. Here, the likelihood function is the forward model based on RT and x are the variables that describe the state of the ocean and atmosphere, such as the aerosol and ocean optical properties or ancillary data. The AC algorithm requires some prior information, $P(x)$, such as the relative humidity, surface pressure, ozone, and water vapor that can, along with their uncertainties, be directly incorporated into the prior (in contrast to a non-Bayesian retrieval where these values are assumed to be true). In the deterministic sense, the likelihood is typically written as $\mathbf{y_{obs}} = \mathbf{F(x)} + \boldsymbol{\epsilon}$, where $\mathbf{F(x)}$ is the forward operator (model), and $\boldsymbol{\epsilon}$ is the uncertainty associated with that model. In Bayesian terminology, the likelihood probability is modeled as a statistical distribution, assumed normal in this case, with mean and variance determined from the forward model.

There are various numerical techniques that approximate Bayes' theorem. The grid approximation is the most straightforward inference engine by approximating the continuous variables, x, on a finite parameters grid. The posterior is calculated by multiplying the likelihood probability and prior probability evaluated at each grid point: a non-iterative brute force approach. The Generalized Nonlinear Retrieval Analysis (GENRA) algorithm for cloud properties retrievals utilizes the grid approximation to retrieve, for example, the posterior of two independent parameters: cloud optical depth and effective radius [44]. Expanding the grid to higher dimensions, however, can be computationally challenging. But, when the dimension is low (e.g., <5), the method is tractable and yields inference results within a reasonable computational time [30]. This manuscript will focus on the normal or quadratic inference approximation, Optimal Estimation (OE), as used in Rodgers's (2000) formalism [45]. This is a widely used inverse algorithm within the atmospheric science community [34,35,37–39,46].

Due to the high computational demand of an OE inference algorithm that fully considers the correlation structure in the observations and model, a fast likelihood function (i.e., forward model) evaluation is necessary. There are several ways to approximate the forward model in the iterative inversion process. A rigorous RT computation is the most accurate; however, it is computationally slow for a complex AO system. The LUT parametrization of the RT, such as NASA's operational tables, are pre-computed and stored for a pre-determined grid of parameters, thus requiring multi-dimensional interpolation for each iteration in the retrieval. The LUT parametrization is accurate and sufficiently fast (for low-dimensional problems) for deterministic inversion using, for example, non-Bayesian methods or low-dimensional OE. However, in the high-dimensional inverse problem of the coupled AO system, the likelihood function based on the associated LUT interpolations becomes computationally costly. We developed a deep Neural Network (NN) with a simple multi-layer perceptron (MLP) architecture that efficiently and accurately emulates the forward RT LUT parameterization to speed up the forward model computations. In this case, the Forward RT NN is a non-linear function approximator of the radiative transfer equation.

Forward model emulators using MLP-NN have been used to speed up the RT computations in modeling solar radiation [47] and satellite sensor simulators [48]. A forward model emulator was also used in inverting geophysical properties using a Gaussian Process model for land surface parameter inference in spectroscopic remote sensing of land and ocean surfaces [50,51], and polarimetric remote sensing of aerosols [52,53]. Note that the NN is not required for a Bayesian retrieval; it is merely a tool adopted to increase the computational efficiency of the analysis. A forward RT NN emulator provides advantages over an inverse RT NN model that estimates geophysical parameters from observations. The forward NN model is easier to train as there is a 1-1 mapping between the geophysical inputs to the RT and the predicted TOA reflectances, avoiding the ill-posed, non-uniqueness, and overfitting problem common with inverse neural networks due to multicollinearity among variables [54]. The Jacobian matrix (see later) is necessary for the iterative inversion scheme. A NN forward model can efficiently provide the Jacobian using the backpropagation chain rule algorithm and, with modern computer languages, Automatic Differentiation (AD) [55–57]. Additionally, the forward RT NN can be used in any iterative or stochastic inversion models that allow uncertainty propagation or estimation, which is more challenging for an inverse NN. However, several studies aimed at assessing the variability in NN weights and their relationship to geophysical parameter uncertainties showed promising results within their application domain [54,58,59].

This work aims to establish an inference framework for the AC that can be potentially applied to global datasets for a wide range of environmental conditions and provide pixel-level uncertainty. The algorithm relies on the simultaneous estimation of the atmospheric parameters (i.e., AOD and fine-mode fraction, and ancillary related parameters), as well as the ocean's inherent optical properties established through the Generalized Inherent Optical Properties (GIOP) model [60] (i.e., absorption coefficients of seawater, phytoplankton and colored dissolved plus detrital matter and backscattering coefficients for seawater and particle matter). It can exploit the information content of all spectral bands available for an instrument. Our OE algorithm finds the optimal solution to the TOA reflectance state vector and estimates the pixel-level uncertainty (i.e., the error covariance matrix) of R_{rs}. The availability of a spectral error covariance matrix can be used as an input for estimating IOP and biogeochemical product uncertainties [61]. The model considers the uncertainty at the TOA due to instrument random noise, ancillary data uncertainty, and the systematic and forward model uncertainty estimated at the Marine Optical BuoY (MOBY) site.

The OE algorithm effectively recasts the standard NASA algorithm approach into a Bayesian framework. The goal of this framework, however, manifests in several ways:

- We aim to assess the performance of the algorithm's retrievals of R_{rs} for a wide range of water conditions and provide validation metrics relative to in-situ data and compared to the NASA standard algorithm.
- We aim to assess the performance of the pixel-level uncertainty of R_{rs} relative to the error between the in-situ data and the satellite retrievals.
- We aim to assess the algorithm's performance on an entire scene retrieval since the algorithm is computationally fast as it relies on the NN model to emulate the forward calculations and provide the Jacobian matrix necessary for the optimization and the uncertainty estimates.

The structure of this paper is as follows. Section 2 provides details of the physical forward model based on radiative transfer computations for the atmospheric LUTs, with the analytical forward model of the GIOP algorithm described in Supplement 1. We follow that with a discussion on the development of the NN model that serves as the likelihood function for the OE algorithm and the associated NN training process. Section 3 details the assumed uncertainty sources. Section 4 describes the OE algorithm architecture, selection of priors, uncertainty propagation, and derivation of R_{rs} through the AC process. Section 5 describes the validation datasets, including the in-situ SeaBASS and AERONET-OC datasets and satellite imagery from the Moderate Resolution Imaging Spectroradiometer (MODIS), along with matchup statistics and uncertainty validation metrics. In Section 6, we evaluate the performance of the NN model as well as the R_{rs} retrieval from the OE algorithm. The OE algorithm is also evaluated on a real validation dataset and compared with the operational algorithm. Finally, we discuss the results and provide a conclusion in Section 7.

2. Methods

2.1 Forward model

The TOA reflectance is based on a radiative coupling of various components of the atmosphere, ocean, and surface [20]. The forward model relates the retrievable geophysical parameters to the TOA observations measured by the satellite sensor. For a clear (cloud-free) ocean pixel, the TOA reflectance is calculated as follows:

$$\rho_t(\lambda;Geom) = \left(\rho_{path}(\lambda;Geom) + \rho'_w(\lambda;Geom) + \rho'_{surface}(\lambda;Geom)\right) \times T_g(\lambda;Geom). \quad (1)$$

It is a function of $Geom$ (i.e., solar zenith θ_0, sensor zenith θ, and relative azimuth φ), and wavelength, λ; $\rho_{path}(\lambda;Geom)$ is the path reflectance due to scattering and absorption by air molecules (Rayleigh scattering) and aerosols bounded by the sea surface; $\rho'_w(\lambda;Geom)$ is the ocean body reflectance, and $\rho'_{surface}(\lambda;Geom)$ is the reflectance contribution from surface glint and whitecaps, where both $\rho'_w(\lambda;Geom)$ and $\rho'_{surface}(\lambda;Geom)$ are expressed at the TOA after propagation through the atmosphere. $T_g(\lambda;Geom)$ is the two-way absorbing gas transmittance along the solar and sensor zenith. The path reflectance is a summation of two terms, the Rayleigh reflectance and the aerosol reflectance (including the aerosol-Rayleigh interaction):

$$\rho_{path}(\lambda;Geom) = \rho_r(\lambda;Geom) + \rho_a(\lambda;Geom). \quad (2)$$

The $\rho_r(\lambda;Geom)$ term is calculated through the tabulation of VRT simulations. The Rayleigh optical depth is calculated from [9]. Although the path reflectance term is shown in Eqs. (1) and (2) as a function of only wavelengths and geometry, the Rayleigh reflectance is also a function of surface pressure and wind speed. The former is needed to know the total number of air molecules in the atmospheric column. The latter is to account for the interaction of Rayleigh scattering with the wind-roughened sea surface. The surface roughness model is from Cox and Munk (1954), and the effect of pressure variation is modeled by [62].

The second term in Eq. (2) is the aerosol reflectance, calculated through the VRT simulations for each of 80 different bimodal aerosol models from [22], consisting of assumed aerosol microphysical properties for a pre-determined set of 8 near-surface atmospheric RHs and 10 fine-mode volume fractions. The aerosol vertical profile in the atmosphere is taken from [63]. The aerosol reflectance calculations include the effects of multiple scattering and molecule-aerosol interaction within the atmosphere. Note that these simulations also provide the molecule-aerosol diffuse transmittance along the solar and sensor directions, $t_{sol}(\lambda, Geom)$ and $t_{sen}(\lambda,Geom)$, respectively, used later to propagate the water and surface reflectance to the TOA.

$\rho'_w(\lambda;Geom)$ is the ocean reflectance at TOA. The bottom of atmosphere (BOA) ocean reflectance $\rho_w(\lambda;Geom)$ is calculated through a forward model that provides the ocean reflectance as a function of Chl-a, $Geom$, and spectral IOPs. The BOA reflectance contribution is attenuated by the diffuse transmittance of the atmosphere, such that $\rho'_w(\lambda;Geom) = t_{sen}(\lambda,Geom) \times \rho_w(\lambda;Geom)$. The BOA ocean reflectance are generated from an ocean reflectance model (ORM) that derives the above-water remote sensing reflectance, $R_{rs}(\lambda;\ sr^{-1})$, which is converted from nadir geometry to the desired solar and sensor path geometries using the bidirectional reflectance distribution function (f_{brdf}) of [64], and then propagated to the TOA as:

$$\rho'_w(\lambda;Geom) = \pi R_{rs} t_{sol} t_{sen} / f_{brdf}. \quad (3)$$

$R_{rs}(\lambda)$ is modeled using the quasi-single scattering approximation ORM [65] included within the Generalized Inherent Optical Property algorithm framework (GIOP) [60]. Given the IOP data as an input to the GIOP ORM forward model, we can simulate a realistic $R_{rs}(\lambda)$ distribution for various conditions observed by ocean color sensors. Details of the GIOP forward model are provided in Supplement 1.

The surface reflectance, $\rho'_{surface}(\lambda;Geom)$, is the light scattered by the air-sea interface. It has two terms: the direct sun glint reflectance and the whitecap reflectance, both of which are driven by the ocean surface wind speed. It is important to remember that the sky glint reflection was calculated through the VRT model of the Rayleigh signal. However, the direct glint signal is calculated by the two-way attenuation of the direct solar beam that is modulated by the surface glint reflectance, $L_{GN}(\lambda)$, which is modeled using Cox and Munk (1954) wave slope statistics [66]. The TOA direct glint reflectance is then $\pi L_{GN} T_{sol} T_{sen}/\mu_0$, where μ_0 is the cosine of the solar zenith angle, and spectral (and geometric for T) dependency is implied. The whitecap irradiance reflectance at the BOA, $\rho_{wc}(\lambda)$, is based on Koepke (1984) [67] combined with the windspeed-dependent fractional coverage model of Stramksa and Petelski (2003) [68] and the whitecap albedo spectral-dependence in the red and near-infrared from [69]. The BOA irradiance reflectance is then propagated to TOA, similar to the ocean reflectance, as $\rho_{wc} t_{sol}$ t_{sen}, with spectral and geometric dependency implied.

We also account for the main absorbing gases in the atmosphere, including O_3, H_2O, and O_2. The H_2O and O_2 transmittance are based on the HITRAN 2016 line by line (LBL) spectroscopic dataset [70]. Assuming the US standard atmospheric profile, we calculate the LBL transmittance for different column water vapor (CWV) values. Then we apply the instrument spectral response function (SRF) to the LBL transmittances and store them in a LUT. Spectral H_2O transmittance at each $Geom$, T_{wv}, is then interpolated from the LUT for a given slant water vapor (WV) concentration along the path as cwv/μ, where μ is the cosine of the path zenith angle. The O_2 transmittance is calculated similarly for different path lengths of the atmosphere given the observation geometry. The O_3 transmittance is calculated from the O_3 optical depth assuming the Beer-Lambert-Bougier law, where the optical depth is determined from the spectral O_3 absorption coefficient [71] integrated with the sensor SRFs, and the O_3 concentration. H_2O and O_3 concentrations are taken from ancillary sources.

2.2 Neural network forward model

2.2.1 Data generation

In this work, the NN training dataset is derived from NASA's operational atmospheric LUTs. Hence, the TOA reflectance can be represented as

$$\rho_t(\lambda) = \mathbf{F}(RH, O_3, Pr, WS, WV, fmf, \tau_a, a_{ph}, a_{dg}, b_{bp}, \gamma, Chl-a, \theta_0, \varphi, \theta_v), \quad (4)$$

where $\mathbf{F}$ is the atmospheric LUT and ORM forward model operator, λ is sensor (here MODIS Aqua) band center wavelengths within the solar spectrum, RH is the relative humidity in the atmosphere, O_3 is the column ozone concentration in Dobson units, Pr is the atmospheric pressure in mbar, WS is the wind speed in m/s, WV is the column water vapor concentration in cm, fmf is the aerosol volume fine-mode fraction, τ_a is the AOD at 869 nm, a_{ph} is the phytoplankton absorption coefficient at 443nm, a_{dg} is the colored dissolved and detrital matter absorption coefficient at 443nm, b_{bp} is the particulate backscattering coefficient at 443mn, and γ is the slope of the backscattering coefficient.

Table 1. The range of all the parameters used in the NN training.

Variable	*Range*	*Distribution*	*Distribution Log_{10} mean*	*Distribution Log_{10} standard deviation*
$\boldsymbol{\lambda}$ (nm)	412:869	-	-	-
$\boldsymbol{RH}$ (%)	30:95	Uniform	-	-
$\boldsymbol{O_3}$ (DU)	200:500	Uniform	-	-
$\boldsymbol{Pr}$ (mbar)	800:1100	Uniform	-	-
$\boldsymbol{WS}$ (m/s)	0.1:15	Uniform	-	-
$\boldsymbol{WV}$(cm)	0.01:30	Lognormal	0.173	0.53
$\boldsymbol{fmf}$ (unitless)	0:1	Uniform	-	-
$\boldsymbol{\tau_a}$(unitless)	0:0.4	Lognormal	-1.03	0.316
$\boldsymbol{a_{ph}}$ (m^{-1})	0.001:5	Lognormal	-1.5	0.45
$\boldsymbol{a_{dg}}$(m^{-1})	0.001:5	Lognormal	-1.2	0.63
$\boldsymbol{b_{bp}}$(m^{-1})	0.0001:0.1	Lognormal	-2.35	0.44
Chl-a (mg m^{-3})	0.05:50	Lognormal	-0.217	0.724
$\boldsymbol{\gamma}$ (nm^{-1})	0:2	Uniform	-	-
$\boldsymbol{\theta_0}$(°)	5:77	MODISA geometry	-	-
$\boldsymbol{\varphi}$ (°)	0:180	MODISA geometry	-	-
$\boldsymbol{\theta_v}$(°)	0:65	MODISA geometry	-	-

The input parameters were generated for random uniform distribution with data ranges given in Table 1, with a few exceptions. Aerosol optical depth, τ_a which was modeled with a log-normal distribution such that low optical depth cases have greater representation than higher optical depths [29,72,73]. Similarly, we assumed a log-normal distribution for the column water vapor, WV, a_{ph}, a_{dg}, and b_{bp} with the distribution mean and standard deviation reported in Table 1. The ocean IOPs are based on monthly mean (Level-3, L3) climatology products from MODIS Aqua, as distributed by the OB.DAAC, but the range was extended to include more extreme cases as observed in Level-2 (L2) data. The geometric parameters θ_0, φ, and θ_v were all sampled from two MODIS Aqua orbits for a day in the summer and winter seasons, thus covering the entire solar geometry range of the sensor's imaging duty cycle. As expected, the NN training is highly sensitive to the choice of geometries since radiant path geometry is a primary driver for signal variations at TOA. Sampling from observed orbit geometries ensures that the NN training considers only realistic solar and viewing geometry combinations, thus improving performance. However, we did not include covariance between the other parameters, which are all assumed independent.

2.2.2 Training process

We generated spectral TOA reflectance, ρ_t, from the standard algorithm LUT for 16 million different data points. After excluding data points with the normalized sun glint radiance > 0.005, similar to the operational algorithm, we ended with ~ 9 million data points for the training. The training was performed using the open-source machine learning platform Keras-TensorFlow (Keras.io). The NN input layer vector has 15 parameters (Table 1), and the output layer is the TOA reflectance, ρ_t, at 13 MODIS wavelengths from 412 to 869 nm. We found by trial and error that four hidden layers provide a good performance of the NN, with additional layers just adding forward model computational cost in the retrieval with negligible performance improvement. The Rectified Linear Unit (ReLU) activation function was used for the NN hidden layers [74]. We trained the NN with the Adam optimization algorithm for 10,000 epochs and with a batch size of 1,000 [75]. The dataset was split into the training set (85%) and a test set (15%). The mean squared error cost function between the training dataset (i.e., ρ_t) and the predicted values was minimized through the optimization process of the NN weights. We compared the NN performance on the training and test (independent) sets for all the training epochs, showing a continuous decline in the cost function for both training and testing, indicating that the NN did not overfit the training data.

3. Uncertainty sources

It is important to properly account for an instrument's measurement uncertainty when validating the uncertainty estimates of the inferred variables. The measurement uncertainty includes both random and systematic components. The random component (noise) is calculated using the instrument's signal-to-noise ratio (SNR). In this work, we assume the sensor's uncorrelated random noise effects. We calculate the noise-equivalent radiance as follows:

$$NE\Delta L(\lambda) = [C_0(\lambda) + C_1(\lambda) \times L_t(\lambda)] \times S(\lambda), \tag{5}$$

where, $C_0(\lambda)$ and $C_1(\lambda)$ are linear fit coefficients of the noise model from [76] , and $S(\lambda)$ is the spectrally-dependent spatial weight that brings all bands to a common 1 km spatial resolution [18]. The standard deviation of the signal is radiance-dependent and calculated as:

$$\sigma_n(\lambda) = \frac{NE\Delta L(\lambda)}{L_t(\lambda)} . \tag{6}$$

The standard deviation of the radiance is then converted to noise-equivalent reflectance after normalizing by the solar irradiance at a specific solar angle.

Systematic (calibration) errors in measurements are challenging to characterize post-launch due to the lack of an accurate absolute calibration apparatus on-orbit. Typically, the systematic uncertainty is correlated between bands. The Marine Optical BuoY (MOBY) site, off the coast of Lanai, Hawaii, is the system vicarious calibration site for all NASA-supported ocean color missions. NOAA has continuously operated MOBY since 1996 as the in-situ calibration source for vicarious calibration and a source of high-quality R_{rs} data [77,78]. There are 523 co-located, coincident MODIS Aqua-MOBY matchups, to date, of which a smaller fraction are used for the system vicarious calibration to derive the gain corrections at the TOA. Our approach relies on estimating the total uncertainty between the observed and predicted TOA reflectance in this work. Similar to [78], we calculate the predicted TOA reflectance by propagating the in-situ MOBY R_{rs} to the TOA while simultaneously solving for the aerosol properties.

The residual uncertainly between the observed and predicted TOA reflectance represents the total uncertainty at TOA defined below:

$$S_t = S_n + S_a + S_w + S_b, \tag{7}$$

where these terms represent error covariance matrices, with subscripts t for the total uncertainty, n for random noise, a is for ancillary data uncertainty, w is for the in-water component from MOBY, and b for the uncertainty due to instrumental systematic artifacts as well as the forward model uncertainty (e.g., RT simplifications). It is valid to sum these terms assuming that each is independent. The terms S_n and S_a are known given the SNR model and the ancillary data uncertainty. The uncertainty in MOBY R_{rs} observations is not well known for all conditions, but is expected to be a few percent [79]; thus, we assume it is negligible in this work as a first approximation. The term S_b can then be estimated and taken as a measure of systematic and forward model uncertainty used in the retrieval process.

4. Optimal Estimation (OE)

Optimal Estimation finds the most probable values of the unknown parameters in Table 1 by minimizing a cost function that incorporates the likelihood function, priors, and uncertainties. The likelihood and priors are assumed to be normal distributions, characterized by a one sigma width and correlations for all measurement pairs. The cost function near the solution is typically the weighted sum of squared differences between the forward model and the measurements, plus a similar weighted squared difference between the state and prior knowledge of the state. For non-linear problems such as the radiative transfer in the AO system, an iterative constrained optimization is used to minimize the cost function. Also, for simplicity, a conjugate Gaussian distribution of the error covariance matrices is assumed, and, therefore, the computationally intensive sampling of the distributions is unnecessary. Note strictly that these are "uncertainty" rather than "error" covariance matrices, as in this case, the true value is not known (uncertainty is a measure of dispersion, and error is a departure from the truth) [12]. However, we use the common "error covariance" terminology for convenience (JCGM, 2008). OE involves determining the maximum a posteriori (MAP) solution, which is a single point estimate of the approximately normal distribution at the mode of the posterior, obtained by minimizing the negative log posterior (known as the cost function, χ^2):

$$-2\log_e \mathrm{P}\left(\mathbf{x} \,\middle|\, \mathbf{y_{obs}}, \mathbf{x}_a\right) = [\mathbf{y}-\mathbf{F}(\mathbf{x},\mathbf{b})]^{\mathrm{T}}\mathbf{S}_e^{-1}[\mathbf{y}-\mathbf{F}(\mathbf{x},\mathbf{b})] + [\mathbf{x}-\mathbf{x}_a]^{\mathrm{T}}\mathbf{S}_a^{-1}[\mathbf{x}-\mathbf{x}_a]. \quad (8)$$

For this study, the forward model, $\mathbf{F}(\mathbf{x},\mathbf{b})$, is the forward radiative transfer calculated for a given state vector $\mathbf{x}$ and, while $\mathbf{b}$ represents the parameters that are used as an input to the forward model, but not part of the state vector. $\mathbf{y_{obs}}$ is a vector that contains the spectral observed TOA reflectance, while $\mathbf{x_a}$ is the prior state vector (knowledge of the state vector $\mathbf{x}$ before measurements). $\mathbf{S}_e$ is the measurement error covariance matrix, and $\mathbf{S}_a$ is the prior error covariance matrix. The diagonal elements of these matrices are the variances, while the off-diagonal elements represent the correlated standard uncertainties in the state variables. Both matrices need to be positive semi-definite (i.e., non-negative). The forward model parameters (state vector) are:

$$\mathbf{x} = \left[RH, O_3, Pr, WS, WV, fmf, \tau_a, a_{ph}, a_{dg}, b_{bp}\right]. \quad (9)$$

The state vector $\mathbf{x}$ in Eq. (9) includes the ancillary data as retrievable parameters, which is different from many other approaches that either assume they are known perfectly or that they are known imperfectly with some uncertainty (in which case this uncertainty is typically propagated to TOA and included in $\mathbf{S}_e$). Suppose the uncertainty of the ancillary data is known or assumed. In that case, it is logical to have them as part of the state vector $\mathbf{x}$ since the ancillary data do influence the observations. Meanwhile, the non-retrievable parameters $\mathbf{b}$ include Chl-a and γ.

The iterative process to find a solution to the state vector, $\mathbf{x}$, follows the modified Gauss-Newton optimization method by Levenberg-Marquardt (LM) [80,81] . We used the Python library SciPy which implements the least-squares algorithm. The LM algorithm is very efficient and provides a high convergence rate. Once a solution is found, we can estimate the error covariance matrix at the estimated parameters. This is calculated using error propagation through the Jacobian matrix, $\hat{\mathbf{K}}$, expressed as:

$$\hat{\mathbf{S}} = \left(\hat{\mathbf{K}}^T\mathbf{S}_e^{-1}\hat{\mathbf{K}} + \mathbf{S}_a^{-1}\right)^{-1}, \quad (10)$$

where $\hat{\mathbf{K}}$ is the partial first derivative of the forward function with respect to the state vector (i.e., $\partial\mathbf{F}/\partial\mathbf{x}$).

This $\hat{\mathbf{S}}$ term is the retrieval uncertainty of the state vector parameters and combines uncertainty introduced by the measurements with the a priori constraints ($\mathbf{S_a}$, see later).

The OE technique described here is based on the normal distribution approximation of prior, likelihood, and measurement uncertainty that may cause problems [12]. The LM algorithm may converge to a local rather than the global minimum when the posterior is multi-modal. This can occur for high-dimensional retrievals that are not properly constrained. The retrieval uncertainty may also be over-or under-estimated if the forward model is highly nonlinear near the solution, which is typically only an issue for poorly constrained parameters.

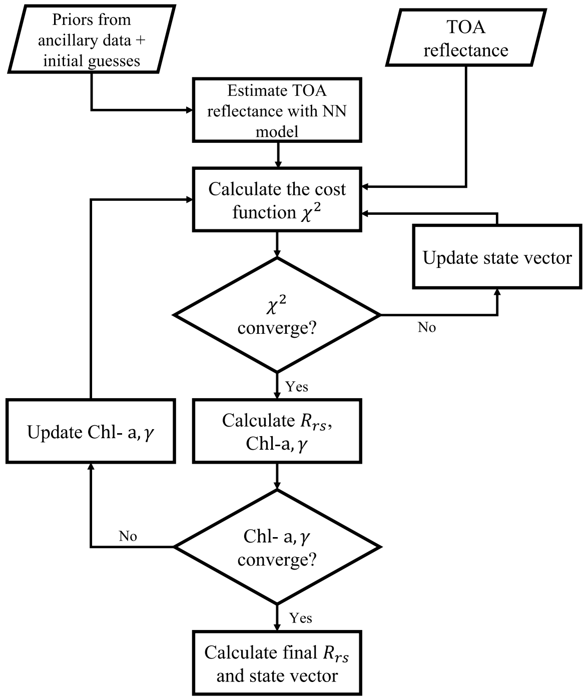

Fig. 1. Flow diagram of the OE algorithm

Figure 1 shows the OE algorithm flow diagram. The required inputs include the TOA reflectance observed from MODIS, the ancillary data, and a priori information about the atmospheric and oceanic state. The prior distribution describes our current knowledge of the parameters of interest, and its mean is used as the first guess in the iterative inversion. Typically there are three types of priors: non-informative such as unbounded uniform distribution; weakly informative, such as bounded uniform or normal distribution with large variance; and informative such as normal with small variance. When non-informative priors are used, the prior does not affect the posterior, and the inference is then identical to the estimate of the likelihood. We used a normal prior distribution with no correlation between parameters in our analysis. RH,O_3, Pr,WS, and WV are obtained from ancillary data sources (National Centers for Environmental Prediction; NCEP). Assuming the mean is known, the uncertainty (standard deviation of the normal distribution) is assumed to be 1 mbar for Pr, 1 m/s for WS, 5% of the mean for RH, 1% of the mean for O_3, and 10% of the mean for WV [82,83]. The fmf,τ_a, a_{ph}, a_{dg}, and b_{bp} priors are assumed weakly informative normal with mean values obtained from the 4-km MODIS Aqua climatology obtained from the OB.DAAC and with a large standard deviation of 10. The standard deviation is much larger than the range of data, but the priors are bounded within their physical values in the inversion. The values for γ and Chl-a were used as the first guess and are obtained from climatology data as well. Given the latitude and longitude of each observation, we interpolate to the nearest neighbor of the global L3 image.

With the initial values of the input parameters, the TOA reflectance is calculated by evaluating the NN forward likelihood model, and a χ^2 value is calculated (Eq. 8). The algorithm iteratively updates the state vector until it converges. In the next step, we calculate the R_{rs} by performing the AC outlined in the following section. Neither Chl-a nor the backscattering slope is part of the state vector. Including them creates a highly ill-posed problem. The spectral backscattering requires simultaneously solving for its shape and magnitude. To avoid this problem, but provide a calculation of both, we utilize empirical relationships to estimate γ and Chl-a from R_{rs}. We use the OCx algorithm for Chl-a [84], and we use the Quasi Analytical

Algorithm (QAA) for γ [85]. The Chl-a and γ are iteratively adjusted and the OE AC correction is repeated until they converge (i.e., they change by >2%), with a maximum of 10 iterations (typically 2-3 are needed). It is important to note that we assume the uncertainty from Chl-a and γ do not propagate into the R_{rs} uncertainty since Chl-a would only impact the uncertainty in the BRDF correction, while γ does not play a role in the R_{rs} uncertainty estimate other than constraining the AC.

4.1 Remote sensing reflectance (R_{rs}) retrieval

Our approach to the AC is a two-step one. First, the OE algorithm estimates the atmosphere-related parameters in the previous section. Second, the inferred parameters are ingested into a proper atmospheric correction similar to the operational algorithm. That involves removing the atmospheric and glint signal from TOA observations and compensating for the atmospheric diffuse and direct transmittance once these properties are inferred. We start by relating R_{rs} to the normalized water-leaving radiance (for simpler notation, λ is not included in the following equations),

$$R_{rs} = \frac{L_{wn}}{F_0},\ (sr^{-1}), \tag{11}$$

where L_{wn} is the normalized water-leaving radiance after the BRDF correction and F_0 is the extraterrestrial solar irradiance at 1 astronomical unit. L_{wn} is connected to TOA observations by

$$L_{wn} = \frac{f_{brdf}\, tL_w}{t_{sen}\, t_{sol}\, \mu_0\, f_{sol}}, \tag{12}$$

where tL_w is the water-leaving radiance measured at TOA, t_{sen} and t_{sol} represent the diffuse transmittance along the viewing and solar direction, respectively, μ_0 is the cosine of the solar zenith angle, f_{sol} is the earth-sun distance correction factor, and f_{brdf} is the BRDF correction factor:

$$tL_w = \frac{F_0\mu_0}{\pi} \times \left[\frac{\rho_t}{T_{gsol}\, T_{gsen}} - \rho_{path+surf}\right]. \tag{13}$$

ρ_t is the observed TOA reflectance. T_{gsol} and T_{gsen} represent the gas transmittance (ozone and water vapor in this case) along the solar and viewing directions, respectively, $\rho_{path+surf}$ is the TOA reflectance with a black ocean that includes only the reflectance from Rayleigh, aerosols, glint, and white caps reflectance. The dark ocean TOA reflectance is calculated using LUTs such that:

$$\rho_{path+surf} = \mathbf{F}_{\boldsymbol{a}}(Pr, WS, RH, fmf, \tau_a, \theta_0, \varphi, \theta_0). \tag{14}$$

The diffuse transmittance of the atmosphere needs to be calculated and is simply estimated from the LUTs:

$$t_{sol} = \mathbf{F}_{\boldsymbol{tsol}}(Pr, RH, fmf, \tau_a, \theta_0), \tag{15}$$

and

$$t_{sen} = \mathbf{F}_{\boldsymbol{tsen}}(Pr, RH, fmf, \tau_a, \theta_v). \tag{16}$$

The above equations can therefore be used to solve for R_{rs}

$$R_{rs} = \frac{f_{brdf}}{\pi\ t_{sen}\, t_{sol}} \times \left[\frac{\rho_t}{T_{gsol}\, T_{gsen}} - \rho_{path+surf}\right]. \tag{17}$$

To estimate the uncertainty in the R_{rs} estimate, we can easily propagate the uncertainties from the inferred parameters through the above equations step-by-step. In the OE method, we can calculate the Jacobian matrices of $\mathbf{F}_{\boldsymbol{a}}$, $\mathbf{F}_{\boldsymbol{tsol}}$, and $\mathbf{F}_{\boldsymbol{tsen}}$ which are denoted as $\mathbf{K}_{a}$, $\mathbf{K}_{tsol}$, and $\mathbf{K}_{tsen}$, respectively. We can then simplify the estimate of the R_{rs} as follows:

$$R_{rs}(\lambda) = \mathbf{F}_{\mathbf{AC}}(\rho_t(\lambda), RH, O3, Pr, WS, WV, FMF, \tau_a, \theta_0, \varphi, \theta_v), \tag{18}$$

where $\mathbf{F}_{\mathbf{AC}}$ is the Atmospheric Correction function. Using the chain rule, we can efficiently calculate its Jacobian matrix, $\mathbf{K}_{\mathbf{AC}}$, to estimate the error covariance matrix of the remote sensing reflectance, $\mathbf{S}_{\boldsymbol{Rrs}}$, as follows:

$$\mathbf{S}_{\boldsymbol{Rrs}} = \mathbf{K}_{\mathbf{AC}}{}^{\mathbf{T}}\hat{\mathbf{S}}\,\mathbf{K}_{\mathbf{AC}} + \mathbf{K}_{\mathbf{TOA}}{}^{\mathbf{T}}\mathbf{S}_{\mathbf{e}}\,\mathbf{K}_{\mathbf{TOA}}. \tag{19}$$

The second term in Eq. (19) accounts for propagating sensor noise to R_{rs} directly where $\mathbf{K}_{\mathbf{TOA}}$ is the Jacobian of R_{rs} with respect to the TOA reflectance. This method is a two-step approach, where both terms on the right-hand side of Eq. (19) are assumed to be independent.

5. Validation data

5.1 In-situ radiometry

The in-situ R_{rs} data were obtained from the NASA SeaBASS database (seabass.gsfc.nasa.gov) includes above and in-water radiometry as well as retrievals from AERONET-OC (Version 2.0, Level 2.0) sites (aeronet.gsfc.nasa.gov) [86,87]. The AERONET-OC sites shown in Figure 2, marked in red circles, are primarily located in coastal water near land. We used Version 2.0 for consistency with the latest validation statistics used in the operational algorithm of the SeaWiFS Data Analysis System (SeaDAS) and applied Level 2.0 quality filtering to ensure the highest quality data. A complete list of the locations and characteristics of the AERONET-OC sites are found on the AERONET-OC webpage and in [86]. The SeaBASS data points are marked in blue circles shown in Figure 2, including samples in open ocean conditions. Accordingly, the data exhibits a large dynamic range of R_{rs}. Full details on the R_{rs} dynamic range for all datasets are available on the SeaBASS web page.

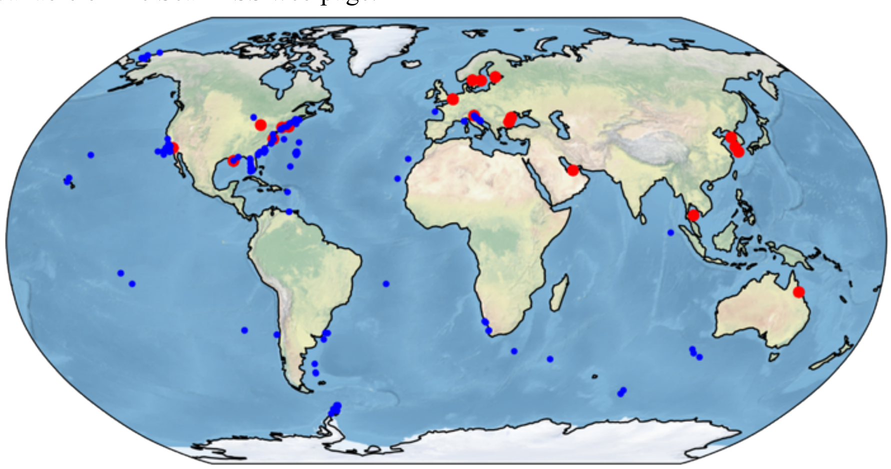

Fig. 2. Map of the SeaBASS (blue circles) and AERONET-OC (red circles) sites used in the validation.

5.2 MODIS Aqua

TOA reflectance data from MODIS onboard the Aqua satellite (MODIS-A) were used in this study to validate R_{rs} matchups. MODIS-A level-1A (L1A) data were obtained from NASA's OB.DAAC and processed to level-1B (L1B) after georeferencing. Satellite match-ups coincident with the in situ validation dataset were identified following [88]. Satellite measurements are derived from a box of pixels (i.e., 5 km ×5 km) centered on the location of the in situ measurement. The satellite value is defined as the filtered mean of unflagged pixels in the box, and the spatial homogeneity and other quality criteria at the validation point are evaluated. Since in-situ data are rarely collected at the precise moment when a satellite views its location, we allow a time window threshold of ±3-hours around the ground truth observations. The length of that window is a compromise between being short enough to minimize differences due to temporal variability in the ocean and being long enough to create a sufficient volume of successful match-ups with satellite observations. The L1B file was then processed to L2 using the SeaDAS standard algorithms to obtain geophysical products as well as the TOA reflectance after applying the Ocean Biology Processing Group (OBPG) calibrations of reprocessing R.2018 (e.g., polarization correction and vicarious calibration) [78]. The standard L2 products were stored and used for the validation comparisons. Since the vicarious calibration is an AC-specific procedure, we removed the vicarious gains from the TOA reflectance by dividing the standard algorithm gains for the OE L2 processing. We then use the modified TOA reflectance in the OE algorithm, as shown in Figure 1.

5.2 Statistical metrics

When comparing satellite-derived R_{rs} with the in-situ value, we use several metrics, primarily mean bias, δ, and the mean absolute error (MAE or $|\delta|$) both of which are routinely used to assess model skill in SeaBASS [89]. We also calculated the root mean squared errors (RMSE) (Δ) and the Pearson and Spearman squared, R^2 correlation as well as the centered (bias-corrected) MAE $|\delta|_c$ and RMSE Δ_c and the mean absolute relative error, $|\psi|_m$. We adopt the IOCCG report 18 [28] notation assuming the satellite observations are denoted $x_{i=1,N}$ and in-situ denoted $y_{i=1,N}$ and the following metrics are:

$$\delta = \frac{1}{N} \times \sum_{i=1}^{N} y_i - x_i, \tag{20}$$

$$|\delta| = \frac{1}{N} \times \sum_{i=1}^{N} |y_i - x_i|, \tag{21}$$

$$\Delta = \sqrt{\frac{1}{N} \times \sum_{i=1}^{N} (y_i - x_i)^2}, \tag{22}$$

$$|\psi|_m = 100 \times \frac{1}{N} \times \sum_{i=1}^{N} \frac{|y_i - x_i|}{x_i}. \tag{23}$$

The centered statistics $|\delta|_c$ and Δ_c simply involve removing the average bias between y_i and x_i, thus showing the algorithm performance without any potential bias either in the algorithm or the in-situ data.

5.2 Uncertainty validation

Our assumption to account for all sources of uncertainties at TOA relies on the MOBY vicarious calibration to be representative of the global oceans. To validate this assumption, we provide a closure analysis by comparing the satellite-derived Rrs and their associated uncertainties to the in-situ measurements. Since we derive the pixel-level uncertainty, we can use a statistical ensemble method to compare the derived uncertainty to the error between the satellite-derived R_{rs} and the in-situ R_{rs}. The uncertainty estimated by OE is a normal distribution with a standard deviation obtained through the analytical error propagation technique. Meanwhile, the error defined as the difference between the retrieved and in-situ truth R_{rs} is an instantaneous realization of that uncertainty distribution. Thus, a direct pixel-level comparison between pixel-level uncertainty and retrieval errors is irrelevant. A more appropriate approach to compare the two quantities is to calculate the normalized error distribution Δ_N following the approach of [90,91], where

$$\Delta_N = \frac{\Delta_s}{\sqrt{u_{sat}^2 + u_{ref}^2}}. \tag{24}$$

Δ_s is the error (i.e., difference) between the satellite-derived and in-situ R_{rs}. u_{sat}^2 is the variance in the pixel-level uncertainty derived from the OE algorithm, while u_{ref}^2 is the variance in the in-situ measurements. In an ideal scenario, where all sources of uncertainties are accounted for in the satellite and in-situ data, with perfect error propagation and with uncertainties following a normal distribution, the normalized error distribution should follow a normal distribution with a zero mean (i.e., no bias) and with 1 variance (or standard deviation). It is possible to examine the normality of the normalized error by plotting the cumulative distribution function (CDF). This provides an assessment of the average comparison between the total error and the OE-provided uncertainties. It is also possible to extend this analysis to assess the variability of the error to uncertainty relationship across the dynamic range of errors which would require stratifying the errors and comparing the 68th percentile of the error to the mean of the uncertainty within a bin (i.e., dividing the data by the expected error into equally populated bins) [91]. The choice of the number of bins depends on the available data volume in order to have a representative sample within each. In our analysis, we choose not to bin the data and provide a comparison between the 68th percentile of the error $\overline{\Delta_s}$, and the mean of the uncertainty, $\overline{u_{sat}}$.

6. Results

6.1 Neural network performance

Our initial analysis of the NN prediction error on the testing dataset indicates that the error varies systematically with radiant path geometry. Figure 3 shows the percent prediction error histogram of the independent dataset (i.e., the 15% of the dataset reserved for testing) for three visible wavelengths (443, 547, and 678 nm) and three NIR wavelengths (748, 859, and 869 nm).

The percent error is calculated as follows:

$$\%\ error = 100 \times (\rho_t^{NN}(\lambda) - \rho_t^{LUT}(\lambda))/\rho_t^{LUT}(\lambda),. \qquad (25)$$

where $\rho_t^{NN}(\lambda)$ is the TOA reflectance calculated by the forward NN model, and $\rho_t^{LUT}(\lambda)$ is the TOA reflectance calculated from the atmosphere-ocean RT-based LUT model. We also calculate the mean absolute error, $|\delta|$, where y_i is retrieved data (i.e., $\rho_t^{NN}(\lambda)$), x_i is the truth (i.e., $\rho_t^{LUT}(\lambda)$), and N is the number of data points (approximately 1.3 million).

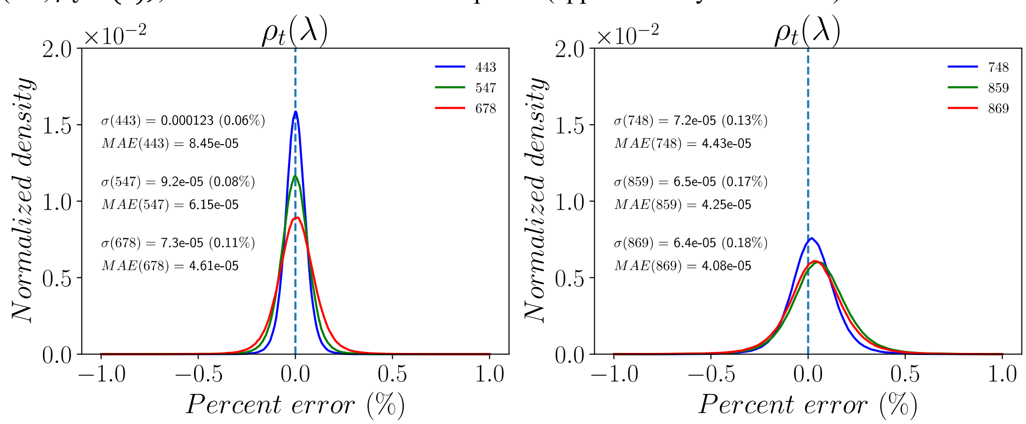

Fig. 3. Histograms of the percent error between the NN derived TOA reflectance and the LUT using an independent validation data set for 443, 547, and 678 nm (right panel), and 748, 859, and 869 nm (left panel). Errors are mostly smaller than 0.2% in reflectance. σ is the standard deviation of the absolute error, while the value in parenthesis is for the percent error. MAE is the mean absolute error.

Figure 3 shows a larger percent error at longer wavelengths, with a slight bias at 859 and 869 nm. Overall, the performance of the NN is excellent with an error < 0.2% for 82% of the testing cases in the worst-case scenario and <0.06% in the best case, similar to the instrument's radiometric noise and within the bounds of the vicarious calibration uncertainty [78]. We parametrized the NN model uncertainty, σ_{NN}, as a function of the geometry to account for the forward model uncertainty needed in the inference process. The NN was trained with the AC LUTs, which were calculated with a coarse grid that can cause interpolation errors. However, in [18], we showed that the LUT interpolation error is the smallest fraction of the total uncertainty. Therefore, this forward model uncertainty here is a fraction of the total forward model uncertainty, which is unknown and likely systematic because of the simplification of the physics (i.e., not accounting for absorbing aerosols and not including other unknown unknowns) [12].

6.2 Synthetic data analysis

Out of the NN test dataset, we extracted 10,000 cases of TOA reflectances and the 'truth' geophysical parameters used in the OE algorithm. Before passing to the algorithm, we added random and systematic radiometric uncertainty to the TOA reflectance derived in section 3 and to the ancillary data input as a prior. The input data set spanned a wide range of environmental conditions and geometries with statistical samples representing the NN training and testing data. In Figure 4, rather than showing a scatter plot, we show the scatter density histogram plot for each retrieved parameter of the OE algorithm. The color bar indicates the normalized density of the data frequency. The plot shows the difference (error) between the retrieved data and the truth. Thus a perfect retrieval would show a zero error on the y-axis. We choose the x-axis that is relevant to the AC process. Since the AC and the TOA reflectance strongly depend on the AOD, dependence between the AC parameters and the R_{rs} is expected. The black dashed lines are the mean of the difference between the retrieved and the truth, while the red dashed lines indicate the +/- standard deviation around the mean of the difference. A bias between the retrieval and the truth would manifest in the black dashed line deviating away from zero. Above zero, the retrieval is overestimated and vice versa for underestimated retrieval. A larger spread between the retrieval and truth would lead to a more significant deviation of the red dashed lines away from the mean black dashed lines.

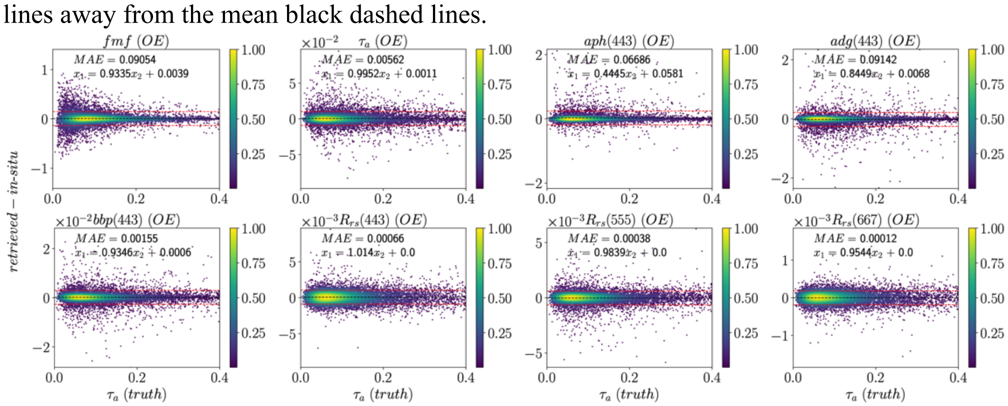

Fig. 4: Scatters density histogram of the synthetic data retrievals using the OE algorithm. The color bar indicates the data normalized density ranging from 0 to 1. MAE is the mean absolute error between the retrieved and truth, while $x_1 = ax_2 + b$ is the regression line between retrieved and truth with a being the slope and b is the bias.

Figure 4 shows the retrieval performance for two parameters related to the AC (*fmf,* τ_a) and three ocean-related parameters from the GIOP model a_{ph}, a_{dg},b_{bp} all at 443 nm. The R_{rs} at 443, 555, and 665 nm were calculated after performing the AC by removing the atmospheric signal contribution from the TOA. There is a negligible bias in the retrieval for all parameters, particularly for R_{rs} with no dependence on the τ_a. The fmf error shows a slight dependence on τ_a at low values, where the uncertainty is increased at low AOD. The MAE ($|\delta|$) of R_{rs} is 0.00066, 0.00038, 0.00012 for 443, 555, and 665nm, respectively, showing that the absolute magnitude of uncertainty is higher at shorter wavelengths consistent with what is observed based on real data validation statistics [27].

To evaluate the retrieval uncertainty for each case, we calculate the CDF of the normalized error distribution, Δ_N, for each retrieval parameter. For a perfect retrieval and uncertainty estimate, the calculated normalized error would agree with the ideal case across the normalized error range. Figure 5 shows the CDF of Δ_N in red compared to the ideal case of a standard normal in black. When the red curve is within the grey shaded region, the uncertainty is underestimated and overestimated when the curve is in the white region. Overall, there is a good agreement for all parameters except b_{bp}, where the uncertainty is underestimated.

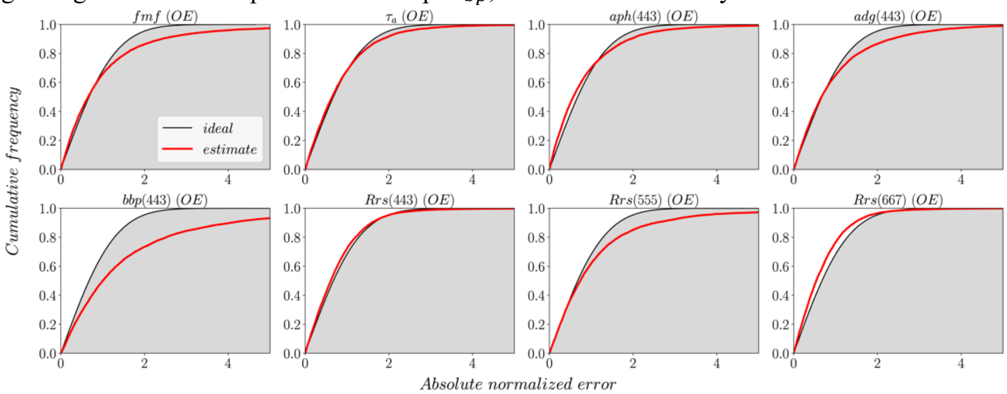

Fig. 5. CDF plot of the absolute normalized error, Δ_N, for all retrieval parameters of the synthetic dataset. The estimated CDF from the OE algorithm is shown in red, and the ideal CDF for a standard normal is shown in black. The grey shaded region shows where the uncertainty is underestimated.

In Table 2, we compare the mean uncertainty estimate $\overline{u_{sat}}$ of the retrieval as compared to the 68th percentile of the retrieval error, $\overline{\Delta_s}$.

Table 2. $\overline{\Delta_s}$ is the 68th percentile of the error between the truth and the retrieval and $\overline{u_{sat}}$ is the mean uncertainty for each parameter.

	fmf	τ_a	a_{ph}	a_{dg}	b_{bp}	$R_{rs}(443)$	$R_{rs}(555)$	$R_{rs}(667)$
$\overline{\Delta_s}$	0.00865	0.0054	0.0373	0.0564	0.00140	0.00076	0.000449	0.00015
$\overline{u_{sat}}$	0.00633	0.0050	0.0372	0.0460	0.00063	0.00078	0.000363	0.00016

The results in Table 2 complement Figure 5, indicating a good agreement between the two results, except for underestimation of b_{bp} uncertainty.

6.3 In-situ validation

6.3.1 SeaBASS

The SeaBASS dataset provides an overall assessment of the OE algorithm in a wide range of water conditions. Figure 6 shows the error between the R_{rs} retrieval and the in-situ truth for three wavelengths at 443, 555, and 667nm. The first row is for the OE algorithm, while the second row is for the operational retrieval using the SeaDAS/l2gen L2 processing software. The matchup analysis shows a lower MAE for the three OE algorithm bands than the operational one. There is no apparent correlation with τ_a, as a primary source of AC errors in all cases.

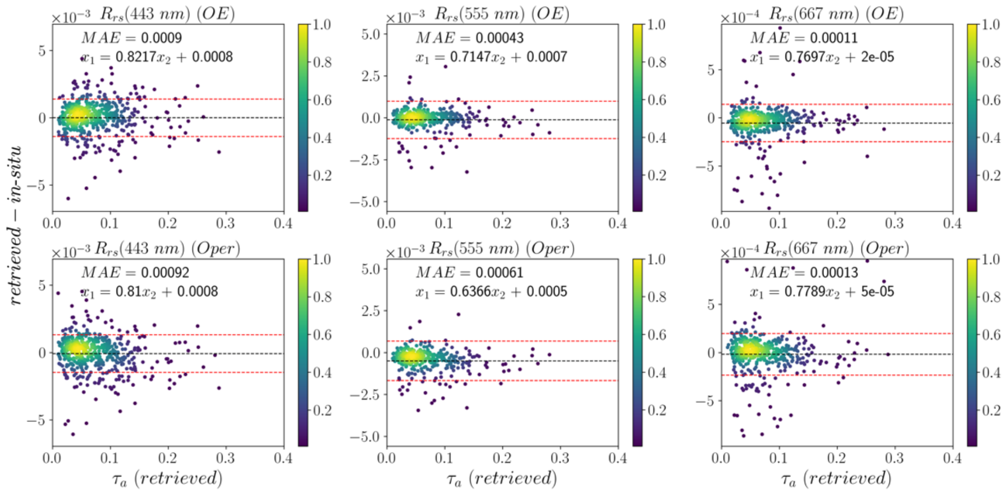

Fig. 6. Scatter density histograms of the SeaBASS data retrievals using the OE algorithm. The first row is for the OE algorithm, and the second row is for the operational algorithm. The color bar indicates the data normalized density ranging from 0 to 1. MAE is the mean absolute error between the retrieved and truth, while $x_1 = ax_2 + b$ is the regression line between retrieved and truth with a being the slope and b is the bias.

For a quantitative assessment, we provide in Table 3 all metrics and validation statistics of the matchups for R_{rs}. In Table 3, the numbers highlighted in bold are for the OE algorithm, while the non-bolded numbers in parentheses are for NASA's current operational AC algorithm.

Table 3. Matchups statistics for the SeaBASS dataset. OE statistics are in bold font-weight, while NASA's operational AC are in normal font and in parentheses. N^- is the number of negative R_{rs} retrievals.

	N (N^-)	δ	$\lvert\psi\rvert_m$ (%)	$\lvert\delta\rvert$	$\lvert\delta\rvert_c$	Δ	Δ_c	R^2 (Pearson)	R^2 (Spearman)
R_{rs}(443)	589	**-1×10^{-6}**	**22.0**	**9×10^{-4}**	**9×10^{-4}**	**1.39×10^{-3}**	**1.39×10^{-3}**	**0.69**	**0.73**
	(**0**, 1)	(-8.1×10^{-3})	(23.7)	(9.2×10^{-4})	(9.2×10^{-4})	(1.39×10^{-3})	(1.40×10^{-3})	(0.69)	(0.71)
R_{rs} (555)	438	**-1.3×10^{-4}**	**13.7**	**4.3×10^{-4}**	**4.5×10^{-4}**	**1.12×10^{-3}**	**1.14×10^{-3}**	**0.76**	**0.73**
	(**0**, 0)	(-5×10^{-4})	(18.6)	(6.1×10^{-4})	(1×10^{-3})	(1.28×10^{-3})	(1.54×10^{-3})	(0.76)	(0.66)
R_{rs} (667)	490	**-5.5×10^{-5}**	**43.4**	**1.1×10^{-4}**	**1.4×10^{-4}**	**2.02×10^{-4}**	**2.23×10^{-4}**	**0.8**	**0.3**
	(**1**, 14)	(-2×10^{-5})	(63.1)	(1.3×10^{-4})	(1.3×10^{-4})	(2.17×10^{-4})	(2.19×10^{-4})	(0.75)	(0.23)

The mean bias, δ, between the in-situ and retrieved R_{rs} is smaller for the OE algorithm, while the $|\psi|_m$ and $|\delta|$ is reduced for all bands. The improvement at 443nm is marginal at 1.7%, however, the $|\psi|_m$ is reduced by 4.9% and 19.7% for the OE algorithm for 555 and 667nm, respectively. We also calculate the centered statistics after removing the mean bias showing consistent results where the OE algorithm outperforms the operational algorithm. The Spearman correlation is improved for the OE algorithm. In contrast, the Pearson correlation shows no improvement except for 667 nm, possibly due to spurious outliers that the metric can be sensitive to.

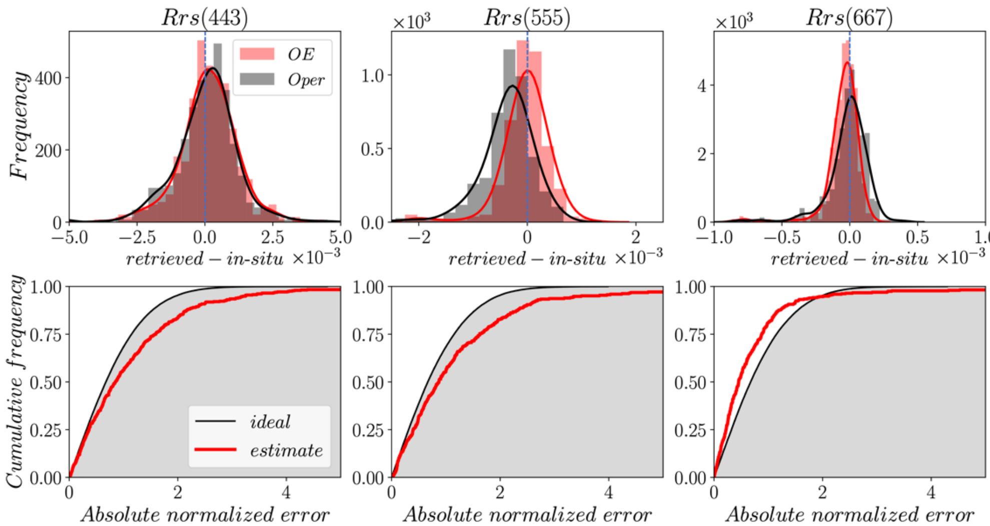

Fig. 7. Top row is a histogram of the difference between the retrieved and in-situ R_{rs} at 443, 555, and 667nm, respectively for the OE algorithm in red, and the operational algorithm in black. The bottom row is the CDF of the absolute normalized error Δ_N for R_{rs} at the same three bands, where the red curve is estimated from the OE algorithm, and the black curve is the ideal case for a standard normal.

Figure 7 shows the histogram of the error as the difference between R_{rs} in-situ and retrieved at 443, 555, and 667 nm, respectively in the top row. The error is mostly centered around zero, and the histogram follows a normal distribution indicating that a random process likely dominates the error. The mode of the R_{rs} histogram is closer to zero for the OE algorithm than for the operational algorithm, for the green and red bands, while the distributions look very similar for the blue band. The bottom row of Figure 7 is the CDF of the absolute normalized error Δ_N. These results indicate that R_{rs} uncertainty is underestimated relative to the in-situ matchup errors for the blue and green bands with higher underestimation for larger errors. At 667 nm, the uncertainty was overestimated for lower errors and vice versa for higher errors. It is important to note that the matchup errors would implicitly include other sources of errors such as in-situ data uncertainties, adjacency effects, and temporal and spatial mismatch.

Table 4. $\overline{\Delta_s}$ is the 68th percentile of the error between the truth and the retrieval and $\overline{u_{sat}}$ is the mean uncertainty for SeaBASS R_{rs} matchups.

	$R_{rs}(443)$	$R_{rs}(555)$	$R_{rs}(667)$
$\overline{\Delta_s}$	0.00091	0.00030	0.00008
$\overline{u_{sat}}$	0.00071	0.00025	0.00012

We also compare the mean uncertainty estimate $\overline{u_{sat}}$ of the retrieval as compared to the 68th percentile of the retrieval error, $\overline{\Delta_s}$ for the SeaBASS matchups in Table 4. There is a good agreement between the two metrics, however, the uncertainty is underestimated slightly at 443 and 555 nm and overestimated for 667 nm.

6.3.2 AERONET-OC

We extended our matchup validation analysis to the AERONET-OC coastal water sites. The plot in Figure 8 shows the retrieval error for R_{rs} at 443, 555, 667nm versus the retrieved τ_a. The results for the OE algorithm show a smaller MAE but comparable spread to the operational algorithm. The plot shows little dependence of the error on the retrieved τ_a indicating no aerosol-dependent bias in the AC.

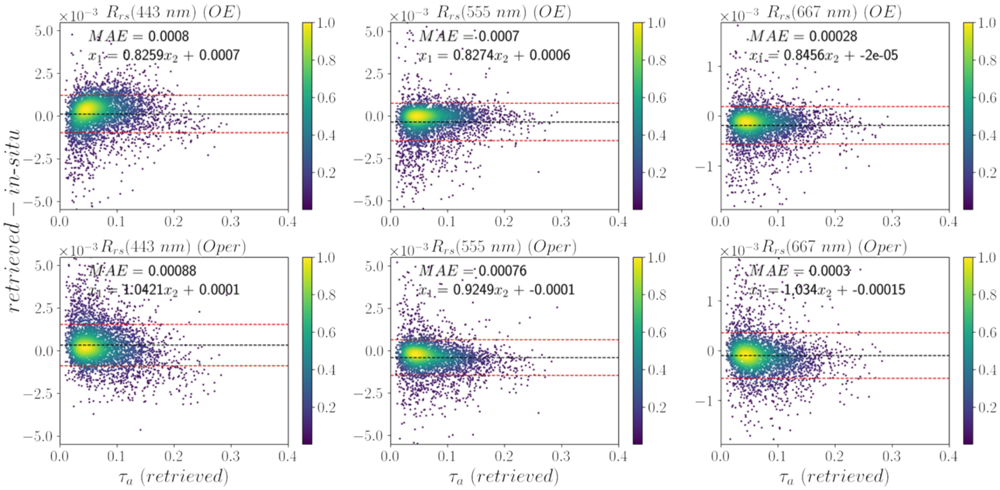

Fig. 8. Same as Fig. 6, but for the AERONET-OC dataset.

The detailed matchup statistics in Table 5 show better metrics for OE, with a smaller mean bias except for 667 nm and lower $|\psi|_m$ and $|\delta|$ across all bands. $|\psi|_m$ was reduced by 1.5, 1.9, and 6.6% for 443, 555, and 667nm, respectively. The centered metrics $|\delta|_c$and Δ_c are consistently better for the OE algorithm, except for the red band, while the correlations are improved for all bands, except for the Pearson metric at 443 nm.

Table 5. Matchup statistics for the AERONET-OC dataset. OE statistics are in bold font-weight, while operational are in normal font. N$^-$ is the number of negative R_{rs} retrievals.

	N (N$^-$)	δ	$\lvert\psi\rvert_m$ (%)	$\lvert\delta\rvert$	$\lvert\delta\rvert_c$	Δ	Δ_c	R^2 (Pearson)	R^2 (Spearman)
Rrs(443)	4300	**1.1×10^{-4}**	**37.7**	**8×10^{-4}**	**8.3×10^{-4}**	**1.10×10^{-3}**	**1.12×10^{-3}**	**0.83**	**0.82**
	(**16,** 47)	(3.2×10^{-4})	(39.2)	(8.8×10^{-4})	(1×10^{-3})	(1.25×10^{-3})	(1.37×10^{-3})	(0.84)	(0.78)
Rrs(555)	3746	**-3.5×10^{-4}**	**12.0**	**7×10^{-4}**	**8.6×10^{-4}**	**1.16×10^{-3}**	**1.31×10^{-3}**	**0.92**	**0.92**
	(**0**, 0)	(-4.1×10^{-4})	(14.1)	(7.6×10^{-4})	(1×10^{-3})	(1.14×10^{-3})	(1.34×10^{-3})	(0.91)	(0.90)
Rrs(667)	3815	**-1.8×10^{-4}**	**30.8**	**2.7×10^{-4}**	**4.1×10^{-4}**	**4.19×10^{-4}**	**5.31×10^{-4}**	**0.87**	**0.82**
	(**13**, 157)	(-1×10^{-4})	(37.4)	(3×10^{-4})	(3.5×10^{-4})	(4.68×10^{-4})	(4.98×10^{-4})	(0.85)	(0.77)

Similar to the SeaBASS analysis, in Figure 9, we show the histogram of the matchup errors and the CDF of the absolute normalized error Δ_N in the top row. The histogram shows a distribution similar to normal for both algorithms. The modes of the distributions are consistently close to zero, where at 555 nm, the OE algorithm is closer to zero than the operational algorithm. In the bottom row, the CDF comparison indicates a close agreement between the estimated uncertainty from the OE algorithm and the matchup errors for 443 nm. The 555 and 667 nm uncertainty are underestimated. This indicates that we did not account for all sources of errors in the algorithm.

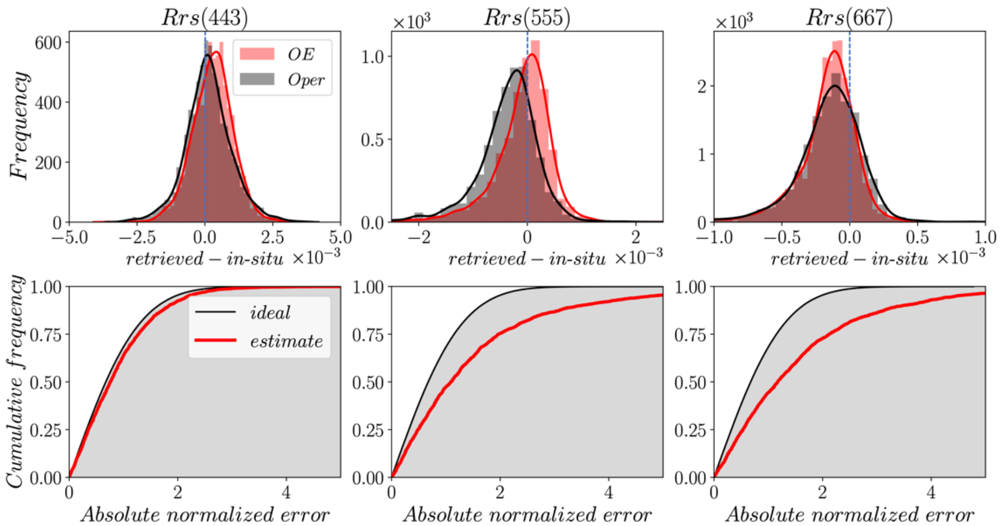

Fig. 9. Same as Fig. 7, but for the AERONET-OC matchups.

Stratification of data by location provides more insight into the performance of the OE algorithm for different environmental conditions. Figure 10 presents results at a more granular level by showing site-by-site CDFs of Δ_N for the 9 sites with at least 250 matchups, plus the remaining sites pooled together (labeled “Others”). The overall performance shows a good agreement in the uncertainty estimate at 443 nm for all sites with slight underestimation. For 550 and 667 nm, the underestimation of the uncertainty is more significant, particularly for MVCO (a highly productive region) and Palgrunden (an inland site). The best agreement was for the Helsinki site, followed by Gustav, both characterized by their high CDOM concentrations [86]. Although Venise provides the most significant volume of data, the uncertainty was underestimated in the green and red bands.

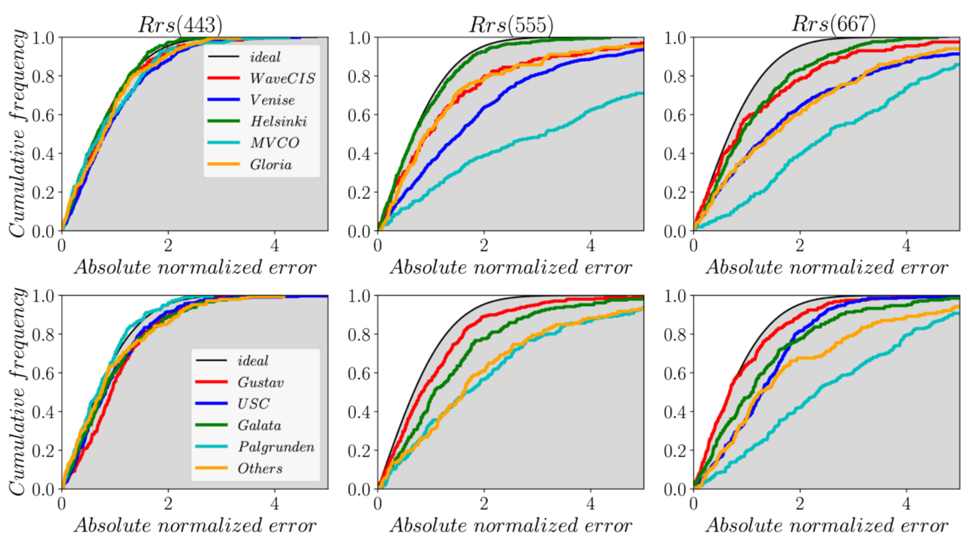

Fig. 10. Same as Fig. 9, but stratified for different AERONET-OC sites.

It is important to note that Δ_N shows a combined effect of retrieval bias and scatter; thus, a highly biased retrieval that is not captured in the uncertainty estimate would lead to a significant over-or underestimation of the normalized error.

We also provide (Table 6) the comparison between the 68[th] percentile of the error, $\overline{\Delta_s}$, and the mean uncertainty from the retrieval, $\overline{u_{sat}}$ for all sites and a breakdown of the 5 best-sampled sites. Similar to Figure 10, there is a good agreement for all sites at 443 nm, however, the uncertainty is underestimated for the green and red bands.

Table 6. $\overline{\Delta_s}$ is the 68[th] percentile of the error between the truth and the retrieval and $\overline{u_{sat}}$ is the mean uncertainty for AERONET-OC R_{rs} matchups.

	$R_{rs}(443)$		$R_{rs}(555)$		$R_{rs}(667)$	
Site	$\overline{\Delta_s}$	$\overline{u_{sat}}$	$\overline{\Delta_s}$	$\overline{u_{sat}}$	$\overline{\Delta_s}$	$\overline{u_{sat}}$
All sites	0.000814	0.000724	0.000483	0.000257	0.000245	0.000137
Venise	0.000910	0.000720	0.000546	0.000241	0.000252	0.000125
Helsinki	0.000683	0.000721	0.000290	0.000272	0.000192	0.000150
MVCO	0.000846	0.000741	0.001150	0.000267	0.000425	0.000140
Gloria	0.000794	0.000726	0.000393	0.000253	0.000280	0.000134
Gustav	0.000854	0.000720	0.000297	0.000262	0.000154	0.000146

6.3.3 MODIS Aqua imagery analysis

Figures 11 and 12 show results of the OE and NASA operational algorithms from a MODIS-Aqua image over the eastern coast of the United States and extending into the Atlantic Ocean on September 21[st], 2010. The scene includes a wide range of water conditions, including coastal waters such as the Chesapeake Bay region and open ocean low Chl-a regions further away from the coast. Figure 11 shows a true-color composite, highlighting four pixels (labeled A-D), representing different water conditions based on the Chl-a and the fitting residual χ^2 of the OE algorithm.

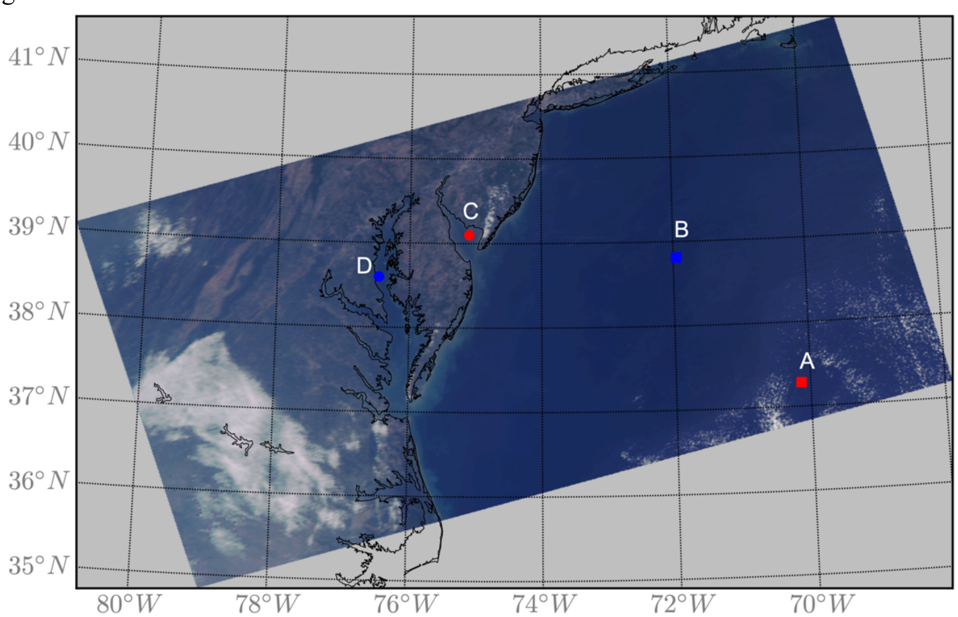

Fig. 11. True-color image composite of MODIS-Aqua from TOA reflectance over the eastern coast of the United States from September, 21[st], 2010.

Figure 12 shows the L2 image of R_{rs} at 443, 555, and 667 nm. Spatial patterns and magnitudes are similar for both the OE (top row) and operational (middle) algorithms, particularly in waters further away from the coast. However, there are differences in coastal waters in regions where $R_{rs}(555)$ are high, indicative of optically complex conditions with high particulate backscattering. The OE algorithm does not perform a cloud screening step similar to the operational approach. However, this caused artifacts in the OE R_{rs} retrieval around cloud edges, which can be mitigated by using additional cloud screening and masking approaches, such as limiting the retrieval with extremely high fitting error χ^2.

In the third row of Figure 12, we show the pixel-level uncertainty produced by the OE algorithm for the three bands. On average, the magnitude of the uncertainty is higher for the blue bands than the green/red bands; it is mainly affected by the atmospheric correction of the aerosol optical depth and fine mode fraction, which are (in this scene) spatially smooth. The fourth row shows the R_{rs} relative uncertainty (%). These images show a more pronounced spatial structure; for example, in the optically-complex Chesapeake Bay waters, the water column's absorption coefficient is so significant that R_{rs} is small in the blue bands, thus the relative uncertainty is substantial. This is reversed in the bright blue waters further from the coast, where the uncertainty is smaller (5-10%). Similarly, in coastal waters, the R_{rs} in the green band is relatively large, so the uncertainty is smaller than in low Chl-a waters. This is also consistent for the red bands, where the low Chl-a conditions show very significant uncertainties (>50% and in some cases >100%), however, this is expected since the R_{rs} is near zero in the red band.

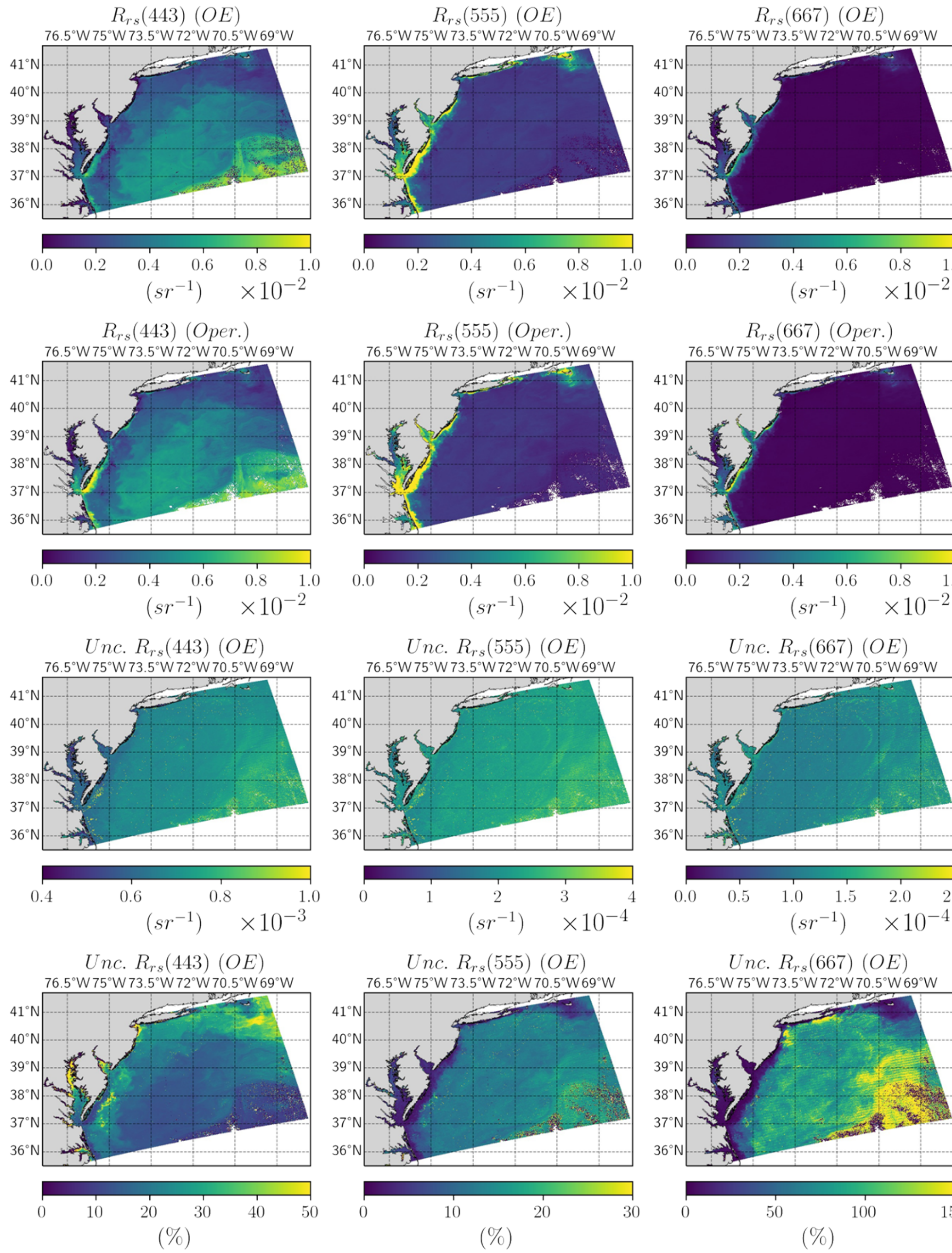

Fig. 12. MODIS-A image of R_{rs} retrieval at 443, 555, 667nm. The top row is the OE algorithm retrieval, and the middle is the operational algorithm. The third row is the absolute uncertainty estimated from the OE algorithm. The last row is for the relative percent uncertainty.

Figure 13 shows the Chl-a retrieval using the OC3 band ratio algorithm after performing the AC to retrieve R_{rs} [84]. The spatial distribution of Chl-a exhibits the typical spatial pattern in that region with high Chl-a values in the Chesapeake Bay (and its estuaries), Delaware Bay, Albemarle Sound, and low Chl-a values in offshore waters of the mid-Atlantic Bight. τ_a and fmf spatial distributions are smooth and do not show artifacts, particularly in very bright waters, where the non-negligible water-leaving radiance in the longer wavelengths can be erroneously attributed as an aerosol signal. However, there is a slight artifact near the mouth of the Chesapeake Bay and adjacent to the southeast of the Delmarva Peninsula between 37° and 38° N in a region where R_{rs} values are relatively high. It is not clear if these are finer aerosol values or retrieval errors due to the high water reflectance signal; however, some of these artifacts are reflected as a higher uncertainty in R_{rs} and χ_n^2, as shown in Figures 12 and 13, respectively. Note that χ_n^2 is the normalized χ^2 where it is divided by the number of bands used in the fitting, such that the theoretical χ_n^2 should have the mode close to 1.

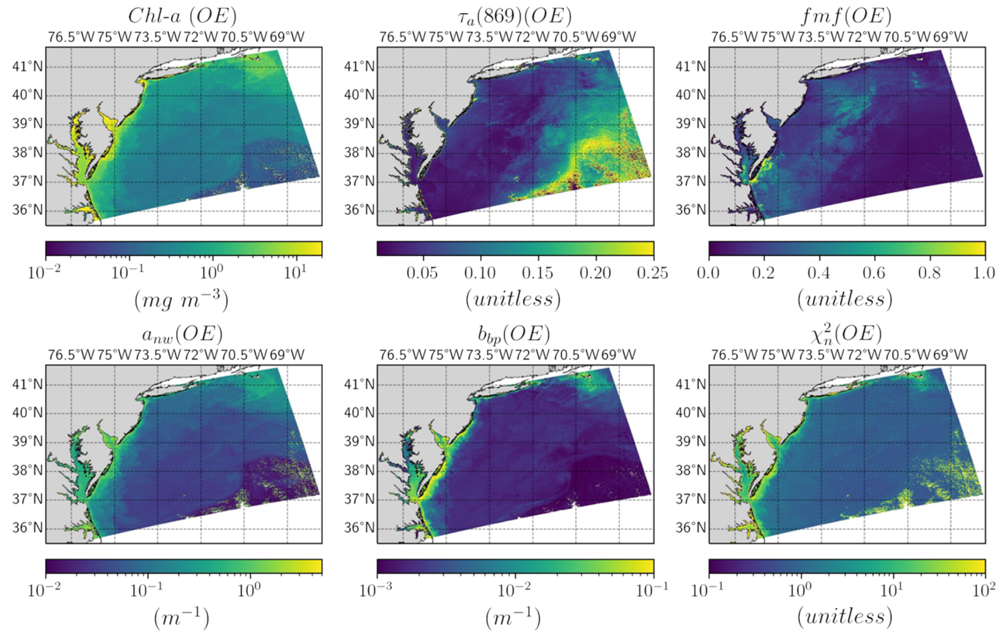

Fig. 13. The top row is the OE algorithm retrieval from MODIS-A of $\mathrm{Chl-a}$, $\tau_a(869)$, fmf. The bottom row is the OE algorithm retrieval of a_{nw} and b_{bp} both at 443nm and χ_n^2.

As intermediate products of the OE algorithm, the absorption and scattering coefficients of the GIOP model can be retrieved. The ocean non-water absorption, a_{nw}, and particulate backscattering, b_{bp}, coefficients at 443 nm are shown in Figure 13 as well. The focus of this algorithm is on the AC. Thus, any detailed evaluation of the IOPs retrieved we consider to be beyond the scope of this manuscript. However, we show the spatial distribution of the IOPs since the OE algorithm relies on a realistic estimate of the surface reflectance to better constrain the AC process by utilizing more bands, including the visible bands. Both IOPs show realistic spatial distributions with relatively high values in coastal waters, particularly within the Chesapeake Bay, which is typically dominated by high CDOM absorption. Both coefficients are smaller further away from the coast, indicating less presence of absorbing and scattering matter in the open ocean.

The final panel of Figure 13 shows χ_n^2, a good metric to indicate the performance of both the forward model and the assumed TOA uncertainty estimate. χ_n^2 values around 1 show a good match between the residual of the forward model at the solution and the uncertainty of the signal; higher χ_n^2 values mean underfitting the forward model and vice versa for lower χ_n^2. Interestingly, in most of the scene, χ_n^2 is close to 1, particularly in pixels away from the coast. However, in coastal waters, χ_n^2 values are, for example, higher than 5, indicating either more difficulty fitting the observations with the forward model or underestimating the assumed measurement/forward model uncertainty. This is expected as coastal waters are significantly more challenging to model with only three parameters, while the atmosphere could also be more complex in these regions (i.e., absorbing aerosols).

Lastly, Figure 14 compared R_{rs} from the OE and operational algorithms at the four locations A-D in Figure 11. Cases A and B represent low Chl-a conditions with values of 0.11 and 0.2 mg m^{-3}, and χ_n^2 of 0.6 and 0.69, respectively. There is an excellent agreement between both retrievals as expected due to the simplicity of the environmental conditions in these waters. This demonstrates that the OE algorithm does provide viable R_{rs} estimates, and low χ_n^2 indicates a good fit of the forward model to the measurements. Since the OE algorithm provides the uncertainty estimate, we also show 1 and 2 standard deviations of R_{rs} estimated by the OE algorithm.

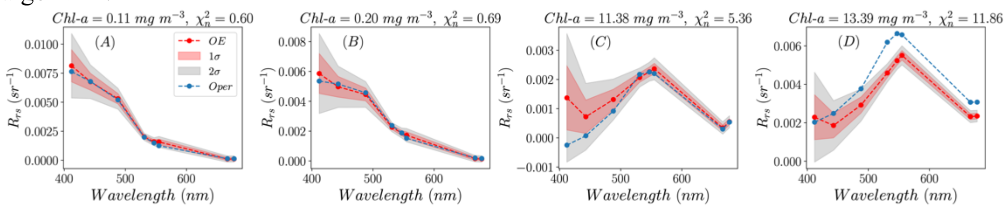

Fig. 14. Spectral R_{rs} retrieval using the OE algorithm (red dashed lines) vs the Operational algorithm (blue dashed lines) for 4 different cases (locations). The 1σ and 2σ envelope of the R_{rs} uncertainty estimated using the OE algorithm is shown in red and grey shading, respectively. The 4 different cases (indicated in Fig. 11) highlight R_{rs} at different water conditions, from low to high Chl-a and χ_n^2.

Figure 14 panels C and D are the R_{rs} retrievals for the coastal sites with Chl-a values of 11.38 and 13.39 mg m^{-3} and χ_n^2 of 5.36 and 11.86, respectively. There is a good agreement between the two algorithms for the green-red bands with larger deviation in the shorter bands for case C. Furthermore, in case C, the OE is higher than the operational algorithm, where R_{rs} is (unphysically) negative for the operational retrieval. The agreement between both retrievals is mostly within one standard deviation of the OE algorithm, except for the 412 nm band, where the OE retrieval appears unrealistically high, likely due to not applying the vicarious calibration gain (which would have reduced TOA reflectance at 412 nm by approximately 2%, which is significant). The last case (D) is from inland Chesapeake Bay waters that are typically highly absorbing (high CDOM concentration) and highly scattering due to sediment discharge from several estuaries in the region. Case D shows the worst mismatch between both algorithms. The OE R_{rs} is lower across the whole spectrum than the operational algorithm, except for 412 nm, where (similar to case C) it may be overestimated. Cases C and especially D show a high χ_n^2, indicating the forward model likely is not fully capturing the radiative conditions of the atmosphere and the ocean. Although it is challenging to conclude which algorithm provides a more correct retrieval in this case, our previous in-situ matchup analysis indicates that the OE algorithm performs better than the operational algorithm overall.

7. Discussion and conclusion

In this paper, we have developed a framework based on the optimal estimation algorithm as presented in Rodgers 2000, which relies on Bayes' theorem to find the optimal solution to the atmospheric correction problem given a representative model of the atmosphere-ocean system and prior information on the state of that system. The advantage of this framework are as follows:

- The ability to calculate pixel-level uncertainty estimates and fully consider the covariance of the uncertainty in the system. Since the algorithm propagates the error covariance, rather than just the diagonal elements of the covariance (i.e., without correlation), it is possible to fully account for the correlation in the R_{rs} uncertainty when further propagating the uncertainty in subsequent products such as IOPs and Chl-a [61].

- Improved computational speed and differentiability through the NN forward model approach. The algorithm has been accelerated using a NN model that can accurately perform the forward calculations necessary for the iterative approach to find the optimal solution. The NN replaces the LUT interpolation of the AC and the analytical ORM, and also provides the Jacobian matrix needed for the optimization and error propagation.
- Potential for better utilization of the information-rich multi-angle polarimeter instruments for the PACE mission to improve the AC of the Ocean Color Instrument (OCI). OE can utilize prior information from external sources such as ancillary data sources. This knowledge about the state of the AO system can be fed into OE, improving and better constraining the AC problem.
- Because of the speed, differentiability of the algorithm, and its ability to process the full dynamic range of atmospheric and oceanic conditions, and it is operationally capable.
- The algorithm is flexible in its band set configurations since a spectral weight is assigned to the cost function, similar to the multi-band AC (MBAC) algorithm [18]. This allows for the use of information from across the spectrum (i.e., using NIR and/or SWIR only, or using the entire spectrum, including the UV).

Although this work demonstrates an improved framework for the AC problem, there are limitations. This OE framework is a research algorithm and has not been thoroughly tested on large-scale global data. Also, the OE algorithm requires an accurate uncertainty model of the TOA reflectance with a reasonable spectral dependence that influences the cost function. To the best of our knowledge, there has been no standardized approach to model the TOA uncertainty post-launch, including the covariance in uncertainties. In our work, we attempted to estimate the TOA uncertainty using MOBY matchups generated during system vicarious calibration, assuming that the most significant portion of the uncertainty budget is the instrument's systematic and forward modeling uncertainty. The assumption that the uncertainty estimates at MOBY can be applied to the global ocean is strong but may not be valid for the coastal AERONET-OC dataset, as evident from the underestimation of uncertainty relative to the error, particularly for 550 and 667nm.

In the synthetic data analysis, we found that the uncertainty estimate, compared to the truth, is slightly underestimated on average. The ratio between $\overline{\Delta_s}$ and $\overline{u_{sat}}$ of 1 indicates a perfect uncertainty estimate, and for a ratio >1, it indicates underestimation in the OE uncertainty, while <1 means overestimation. For the AC parameters, the ratio was 1.36 and 1.08 for fmf and τ_a, respectively. The uncertainty in the IOPs showed an excellent agreement for a_{ph} with a ratio of 1.002 and a good agreement for a_{dg} with a ratio of 1.22, however, the uncertainty was severely underestimated for b_{bp} where the ratio is 2.22. On the other hand, it is important to note that the focus of this paper is to improve the estimate of R_{rs} and its associated uncertainty. The ratios between $\overline{\Delta_s}$ and $\overline{u_{sat}}$ for R_{rs} at 443, 555, and 667 nm are 0.98, 1.23, and 0.94, respectively, showing a slight overestimation in the blue and red bands and underestimation in the green bands.

We tested the OE algorithm and its uncertainty estimation technique using the SeaBASS dataset, encompassing a large dynamic range of water conditions spanning coastal to open waters. While the overall validation statistics showed an improvement for the OE algorithm relative to the operational one, the improvement was not significant, where $|\boldsymbol{\psi}|_{\boldsymbol{m}}$ was reduced by 1.7, 5, and 19.7% for 443, 555, and 667 nm, respectively, likely because both algorithms forward models rely on the same aerosol microphysical assumptions [22]. This is expected since a large portion of the uncertainty is likely from the modeling assumptions and the inherent limitations in the validation process that would apply to any newly developed algorithm. However, the OE algorithm shows an improved bias in the retrieval with fewer negative R_{rs}

retrievals, particularly for 667 nm, where the error is the most reduced. This improvement is likely due to an improved AC and not an effect of lack of the vicarious calibration since the standard vicarious gain does increase the TOA reflectance by approximately 1%. The ratio between $\overline{\Delta_s}$ and $\overline{u_{sat}}$ (1.27, 1.2, for 443 and 555 nm, respectively) indicates underestimated uncertainties at those wavelengths, while the ratio of 0.67 at 667 nm indicates an overestimate. These ratios show a relatively good agreement, given that we are not fully considering the uncertainty in the in-situ data and other error sources. Large outliers would significantly impact the analysis for small signals in the red. However, there is no clear explanation for why 667 nm uncertainty is overestimated, other than the retrieval error for the SeaBASS dataset is significantly smaller than that at MOBY (where the uncertainty at TOA is calculated).

By extending this analysis to the AERONET-OC sites, we stratified the dataset by different locations. This is due to the large variability of environmental conditions, proximity to land, and water conditions [86]. Since the AERONET-OC sites are predominantly coastal, the validation process is expected to be more challenging. Similar to the SeaBASS dataset, all statistical metrics show an improvement in the matchups using the OE algorithm relative to the operational algorithm with a reduction in bias and improvement in error metrics. The matchups showed a significant decrease in negative R_{rs} retrievals, particularly for 443 nm, where it is reduced nearly three times and 12 times for 667 nm. This is a remarkable improvement and shows that the simultaneous AO retrieval process using multiple bands for the AC provides a valuable advantage over using only NIR bands in coastal waters. Moving to validating uncertainties, in the case of using all available data, the agreement between $\overline{\Delta_s}$ and $\overline{u_{sat}}$ is good for 443 nm with a ratio of 1.12, showing a slight underestimation. However, for 555 and 667 nm, the ratio of 1.88 and 1.78 shows a significant underestimation of the uncertainty. This underestimation happens at all sites, with the worst two performing sites being MVCO and Palgrunden. Both are characterized by low aerosol loadings and smaller fine mode fractions than average. MVCO showed a higher median wind speed (4.6 m/s compared to the median of all cases of 2.5 m/s). Palgrunden is also a high latitude site where the solar angle is typically larger than 40°. Some of these environmental conditions can impact the assessment of the uncertainty validation due to underestimating the TOA reflectance uncertainty characterized at MOBY and retrieval bias. At 443 nm, all sites showed a good agreement with a ratio that ranges from 1.26 to 0.94 (Helsinki being the only site showing slight overestimation). Helsinki and Gustav also showed the best agreement for 550 and 665nm; however, they are underestimated.

There are a few theoretical and practical reasons that could explain the underestimation of the satellite-derived uncertainty:

- The OE algorithm relies on the assumption of a Gaussian posterior distribution, where the variance of the distribution should capture the uncertainty estimate within one standard deviation. This is not necessarily true for the atmosphere-ocean system, as demonstrated using the grid approximation Bayesian inference method in [30], which showed that the full posterior uncertainty is typically larger than the standard deviation of a normal distribution.
- The error propagation relies on the estimate of the Jacobian matrix (i.e., the first derivative). This approximation would not hold for a highly nonlinear relationship between the observations and the state parameters. This issue manifests in the optimization procedure that relies on the derivative of the cost function, which could lead to a local minima leading to a biased inversion.
- The absolute normalized error metric requires complete knowledge of the uncertainty in the in-situ data for each measurement. This encompasses instrument calibration and radiometry knowledge, the effect of environmental conditions on the measurements uncertainty, and spatio-temporal mismatch with the satellite retrieval. This is consistent with the findings of Zibordi et al., 2022 [92], which found that when

assuming 5% uncertainty in the satellite-derived water-leaving radiance, the absolute normalized error metric consistently shows an underestimation of the uncertainty. They attributed that to the overly optimistic 5% uncertainty typically set as a gold standard for ocean color requirements. Additionally, ignoring complex spatio-temporal uncertainties does play a significant role in the underestimation as well as possible biases either in in-situ data or satellite retrievals due to, for example, land adjacency effects [93].

- We assumed ancillary data uncertainty based on fixed absolute and relative uncertainties that do not vary with space or time. Recent work has shown that the uncertainty varies geographically and could have a significant impact on the R_{rs} retrieval, particularly due to relative humidity and windspeed uncertainty which has a large impact on the aerosol quantification [27].
- We assumed that the uncertainty of the TOA observations estimated at the MOBY site is representative of the global oceans. However, in coastal sites, the forward modeling errors are likely larger than in the open ocean due to more complexity in the atmosphere and ocean optical properties, such as the presence of strongly absorbing aerosols and errors in the BRDF correction.

Finally, in our analysis here of the OE algorithm performance, we did not apply the standard vicarious gains that are otherwise applied to the input TOA radiances when operating the standard NASA AC algorithm. Our justification is that these vicarious gains are tuned for the standard algorithm, which relies on the black-pixel (NIR bands) assumption for the AC rather than utilizing the entire visible spectrum as the OE algorithm does. As the OE approach relies on all measurements simultaneously, it is less sensitive to measurement uncertainties than NIR-based AC algorithms that typically use only two bands (unless there is a large systematic bias in the observations). For MODIS-A, the standard vicarious gains are mostly close to 1, except for water vapor bands near 645 and 869 nm (likely due to systematic uncertainty in the water vapor correction) and 412 nm. The gain coefficient at 412 nm reduces the TOA reflectance by ~2%, which is significant (and likely instrument-specific) and would be realized as a large bias in R_{rs} for the OE algorithm. We noticed that R_{rs} at that band was consistently overestimated relative to the operational algorithm, partially explaining less negative R_{rs} at that band. However, negative R_{rs} at all other bands are significantly reduced, likely because of the better constraint on the surface properties using the GIOP forward model. Future work will implement a vicarious calibration procedure for the OE algorithm. Therefore, this research algorithm's performance can only improve beyond what is presented here. That includes improving the aerosol modeling, the RT accuracy, and the bio-optical modeling of the ocean. Our future work plan includes the following steps:

- Further investigate the impact of the prior information either from models or other external sources on the reduction of R_{rs} uncertainty.
- Assess the performance of the full error covariance matrix estimated from the OE algorithm.
- Develop and apply a system vicarious calibration (SVC) procedure for the OE algorithm.
- Develop an operational implementation of the OE algorithm for the PACE mission, to fully exploit the combined capabilities of the OCI sensor and MAPs for ocean color retrievals.

In summary, this work presents a practical recasting of ocean color AC within a Bayesian framework. It demonstrates slightly better quantitative retrieval performance than the current standard approach, as well as quantitatively relevant pixel-level uncertainty estimates that have been missing until now. The OE framework can be applied to current and heritage ocean color sensors. Looking to the future, the Bayesian approach would allow the OCI instrument on

PACE, for example, to utilize retrieval products from its companion instruments, the information-rich MAPs, as informative priors to further constrain the AC process for OCI. In a general sense, the OE framework provides a pathway to take advantage of complementary instruments on the same satellite platform or atmospheric measurements from ancillary sources to improve the quality of satellite ocean color retrievals.

Funding. This research was funded by the National Aeronautics and Space Administration (NASA) Ocean Biology and Biogeochemistry Program via an award under the solicitation of The Science of Terra, Aqua, and Suomi NPP (NNH20ZDA001N), and Plankton, Aerosol, Cloud, ocean Ecosystem (PACE) mission Project Science.

Acknowledgments. We acknowledge the MOBY team and all of the AERONET-OC PIs for their support to make their data available. We also thank the SeaBASS team for archiving and distributing the validation data.

Disclosures. The authors declare no conflicts of interest.

Data availability. Data underlying the results presented in this paper are not publicly available at this time but may be obtained from the authors upon reasonable request

Supplemental document. See Supplement 1 for supporting content.

References

1. H. R. Gordon, "Atmospheric correction of ocean color imagery in the Earth Observing System era," J. Geophys. Res. **102**, 17081–17106 (1997).
2. H. R. Gordon and D. K. Clark, "Clear water radiances for atmospheric correction of coastal zone color scanner imagery," Appl. Opt., AO **20**, 4175–4180 (1981).
3. IOCCG, "IOCCG (2010). Atmospheric Correction for Remotely-Sensed Ocean-Colour Products," (2010).
4. Y. J. Kaufman, D. Tanré, H. R. Gordon, T. Nakajima, J. Lenoble, R. Frouin, H. Grassl, B. M. Herman, M. D. King, and P. M. Teillet, "Passive remote sensing of tropospheric aerosol and atmospheric correction for the aerosol effect," J. Geophys. Res. **102**, 16815–16830 (1997).
5. M. J. Behrenfeld and P. G. Falkowski, "Photosynthetic rates derived from satellite-based chlorophyll concentration," Limnology and Oceanography **42**, 1–20 (1997).
6. M. J. Behrenfeld, R. T. O'Malley, D. A. Siegel, C. R. McClain, J. L. Sarmiento, G. C. Feldman, A. J. Milligan, P. G. Falkowski, R. M. Letelier, and E. S. Boss, "Climate-driven trends in contemporary ocean productivity," Nature **444**, 752–755 (2006).
7. S. W. Chisholm, "Oceanography: Stirring times in the Southern Ocean," Nature **407**, 685–687 (2000).
8. B. Franz, M. J. Behrenfeld, D. Siegel, and S. Signorini, "Global ocean phytoplankton [in State of the Climate in 2015]," Bulletin of the American Meteorological Society **97**, S87–S89 (2016).
9. B. A. Bodhaine, N. B. Wood, E. G. Dutton, and J. R. Slusser, "On Rayleigh Optical Depth Calculations," J. Atmos. Oceanic Technol. **16**, 1854–1861 (1999).
10. A. Gilerson, E. Herrera-Estrella, R. Foster, J. Agagliate, C. Hu, A. Ibrahim, and B. Franz, "Determining the Primary Sources of Uncertainty in Retrieval of Marine Remote Sensing Reflectance From Satellite Ocean Color Sensors," Frontiers in Remote Sensing **3**, (2022).
11. H. R. Gordon and M. Wang, "Retrieval of water-leaving radiance and aerosol optical thickness over the oceans with SeaWiFS: a preliminary algorithm," Applied Optics **33**, 443 (1994).
12. A. C. Povey and R. G. Grainger, "Known and unknown unknowns: uncertainty estimation in satellite remote sensing," Atmospheric Measurement Techniques **8**, 4699–4718 (2015).

13. R. J. Frouin, B. A. Franz, A. Ibrahim, K. Knobelspiesse, Z. Ahmad, B. Cairns, J. Chowdhary, H. M. Dierssen, J. Tan, O. Dubovik, X. Huang, A. B. Davis, O. Kalashnikova, D. R. Thompson, L. A. Remer, E. Boss, O. Coddington, P.-Y. Deschamps, B.-C. Gao, L. Gross, O. Hasekamp, A. Omar, B. Pelletier, D. Ramon, F. Steinmetz, and P.-W. Zhai, "Atmospheric Correction of Satellite Ocean-Color Imagery During the PACE Era," Front. Earth Sci. **7**, (2019).
14. C. R. McClain, G. C. Feldman, and S. B. Hooker, "An overview of the SeaWiFS project and strategies for producing a climate research quality global ocean bio-optical time series," Deep Sea Research Part II: Topical Studies in Oceanography **51**, 5–42 (2004).
15. W. E. Esaias, M. R. Abbott, I. Barton, O. B. Brown, J. W. Campbell, K. L. Carder, D. K. Clark, R. H. Evans, F. E. Hoge, H. R. Gordon, W. M. Balch, R. Letelier, and P. J. Minnett, "An overview of MODIS capabilities for ocean science observations," IEEE Transactions on Geoscience and Remote Sensing **36**, 1250–1265 (1998).
16. R. E. Murphy, W. L. Barnes, A. I. Lyapustin, J. Privette, C. Welsch, F. DeLuccia, H. Swenson, C. F. Schueler, P. E. Ardanuy, and P. S. M. Kealy, "Using VIIRS to provide data continuity with MODIS," in *IGARSS 2001. Scanning the Present and Resolving the Future. Proceedings. IEEE 2001 International Geoscience and Remote Sensing Symposium (Cat. No.01CH37217)* (2001), Vol. 3, pp. 1212–1214 vol.3.
17. 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," Bulletin of the American Meteorological Society **100**, 1775–1794 (2019).
18. A. Ibrahim, B. A. Franz, Z. Ahmad, and S. W. Bailey, "Multiband Atmospheric Correction Algorithm for Ocean Color Retrievals," Front. Earth Sci. **7**, (2019).
19. M. Gao, P.-W. Zhai, B. A. Franz, K. Knobelspiesse, A. Ibrahim, B. Cairns, S. E. Craig, G. Fu, O. Hasekamp, Y. Hu, and P. J. Werdell, "Inversion of multiangular polarimetric measurements from the ACEPOL campaign: an application of improving aerosol property and hyperspectral ocean color retrievals," Atmospheric Measurement Techniques **13**, 3939–3956 (2020).
20. C. Mobley, J. Werdell, B. Franz, Z. Ahmad, and S. Bailey, *Atmospheric Correction for Satellite Ocean Color Radiometry* (2016).
21. F. Steinmetz, P.-Y. Deschamps, and D. Ramon, "Atmospheric correction in presence of sun glint: application to MERIS," Opt. Express, OE **19**, 9783–9800 (2011).
22. Z. Ahmad, B. A. Franz, C. R. McClain, E. J. Kwiatkowska, J. Werdell, E. P. Shettle, and B. N. Holben, "New aerosol models for the retrieval of aerosol optical thickness and normalized water-leaving radiances from the SeaWiFS and MODIS sensors over coastal regions and open oceans," Applied Optics **49**, 5545 (2010).
23. B. N. Holben, T. F. Eck, I. Slutsker, D. Tanré, J. P. Buis, A. Setzer, E. Vermote, J. A. Reagan, Y. J. Kaufman, T. Nakajima, F. Lavenu, I. Jankowiak, and A. Smirnov, "AERONET—A Federated Instrument Network and Data Archive for Aerosol Characterization," Remote Sensing of Environment **66**, 1–16 (1998).
24. O. Dubovik and M. D. King, "A flexible inversion algorithm for retrieval of aerosol optical properties from Sun and sky radiance measurements," Journal of Geophysical Research: Atmospheres **105**, 20673–20696 (2000).
25. H. R. Gordon, T. Du, and T. Zhang, "Remote sensing of ocean color and aerosol properties: resolving the issue of aerosol absorption," Appl. Opt., AO **36**, 8670–8684 (1997).
26. S. E. Schollaert, J. A. Yoder, J. E. O'Reilly, and D. L. Westphal, "Influence of dust and sulfate aerosols on ocean color spectra and chlorophyll a concentrations derived from SeaWiFS off the U.S. east coast," J. Geophys. Res. **108**, 3191 (2003).

27. F. Mélin, "From Validation Statistics to Uncertainty Estimates: Application to VIIRS Ocean Color Radiometric Products at European Coastal Locations," Frontiers in Marine Science **8**, (2021).

28. F. Mélin and IOCCG, *Uncertainties in Ocean Colour Remote Sensing* (International Ocean Colour Coordinating Group (IOCCG), 2019).

29. R. Frouin and B. Pelletier, "Bayesian methodology for inverting satellite ocean-color data," Remote Sensing of Environment **159**, 332–360 (2015).

30. K. Knobelspiesse, A. Ibrahim, B. Franz, S. Bailey, R. Levy, Z. Ahmad, J. Gales, M. Gao, M. Garay, S. Anderson, and O. Kalashnikova, "Analysis of simultaneous aerosol and ocean glint retrieval using multi-angle observations," Atmospheric Measurement Techniques Discussions 1–32 (2020).

31. B. Franz and E. Karaköylü, "Estimating uncertainty in the retrieval of water-leaving reflectance from spaceborne ocean color sensors: effect of instrument noise," in (2016).

32. D. B. Gillis, J. H. Bowles, M. J. Montes, and W. J. Moses, "Propagation of sensor noise in oceanic hyperspectral remote sensing," Opt. Express, OE **26**, A818–A831 (2018).

33. T. Schroeder, M. Schaale, J. Lovell, and D. Blondeau-Patissier, "An ensemble neural network atmospheric correction for Sentinel-3 OLCI over coastal waters providing inherent model uncertainty estimation and sensor noise propagation," Remote Sensing of Environment **270**, 112848 (2022).

34. G. E. Thomas, C. A. Poulsen, A. M. Sayer, S. H. Marsh, S. M. Dean, E. Carboni, R. Siddans, R. G. Grainger, and B. N. Lawrence, "The GRAPE aerosol retrieval algorithm," Atmospheric Measurement Techniques **2**, 679–701 (2009).

35. A. Lipponen, T. Mielonen, M. R. A. Pitkänen, R. C. Levy, V. R. Sawyer, S. Romakkaniemi, V. Kolehmainen, and A. Arola, "Bayesian aerosol retrieval algorithm for MODIS AOD retrieval over land," Atmospheric Measurement Techniques **11**, 1529–1547 (2018).

36. C. A. Poulsen, R. Siddans, G. E. Thomas, A. M. Sayer, R. G. Grainger, E. Campmany, S. M. Dean, C. Arnold, and P. D. Watts, "Cloud retrievals from satellite data using optimal estimation: evaluation and application to ATSR," Atmospheric Measurement Techniques **5**, 1889–1910 (2012).

37. A. K. Heidinger, "Rapid Daytime Estimation of Cloud Properties over a Large Area from Radiance Distributions," Journal of Atmospheric and Oceanic Technology **20**, 1237–1250 (2003).

38. C. D. Rodgers, "Retrieval of atmospheric temperature and composition from remote measurements of thermal radiation," Reviews of Geophysics **14**, 609–624 (1976).

39. S. Turquety, J. Hadji-Lazaro, C. Clerbaux, D. A. Hauglustaine, S. A. Clough, V. Cassé, P. Schlüssel, and G. Mégie, "Operational trace gas retrieval algorithm for the Infrared Atmospheric Sounding Interferometer," Journal of Geophysical Research: Atmospheres **109**, (2004).

40. C. Shi and T. Nakajima, "Simultaneous determination of aerosol optical thickness and water-leaving radiance from multispectral measurements in coastal waters," Atmospheric Chemistry and Physics **18**, 3865–3884 (2018).

41. C. Shi, M. Hashimoto, and T. Nakajima, "Remote sensing of aerosol properties from multi-wavelength and multi-pixel information over the ocean," Atmospheric Chemistry and Physics **19**, 2461–2475 (2019).

42. D. R. Thompson, K. Cawse-Nicholson, Z. Erickson, C. G. Fichot, C. Frankenberg, B.-C. Gao, M. M. Gierach, R. O. Green, D. Jensen, V. Natraj, and A. Thompson, "A unified approach to estimate land and water reflectances with uncertainties for coastal imaging spectroscopy," Remote Sensing of Environment **231**, 111198 (2019).

43. T. Bayes and null Price, "LII. An essay towards solving a problem in the doctrine of chances. By the late Rev. Mr. Bayes, F. R. S. communicated by Mr. Price, in a letter to

John Canton, A. M. F. R. S," Philosophical Transactions of the Royal Society of London **53**, 370–418 (1763).
44. T. Vukicevic, O. Coddington, and P. Pilewskie, "Characterizing the retrieval of cloud properties from optical remote sensing," Journal of Geophysical Research: Atmospheres **115**, (2010).
45. C. D. Rodgers, *Inverse Methods for Atmospheric Sounding*, Atmospheric, Oceanic and Planetary Physics (World Scientific Publishing, 2000), Vol. Volume 2.
46. A. M. Sayer, N. C. Hsu, C. Bettenhausen, J. Lee, J. Redemann, B. Schmid, and Y. Shinozuka, "Extending “Deep Blue” aerosol retrieval coverage to cases of absorbing aerosols above clouds: Sensitivity analysis and first case studies," Journal of Geophysical Research: Atmospheres **121**, 4830–4854 (2016).
47. H. Takenaka, T. Y. Nakajima, A. Higurashi, A. Higuchi, T. Takamura, R. T. Pinker, and T. Nakajima, "Estimation of solar radiation using a neural network based on radiative transfer," Journal of Geophysical Research: Atmospheres **116**, (2011).
48. P. Castellanos and A. da Silva, "A Neural Network Correction to the Scalar Approximation in Radiative Transfer," Journal of Atmospheric and Oceanic Technology **36**, 819–832 (2019).
49. J. L. Gómez-Dans, P. E. Lewis, and M. Disney, "Efficient Emulation of Radiative Transfer Codes Using Gaussian Processes and Application to Land Surface Parameter Inferences," Remote Sensing **8**, 119 (2016).
50. J. Brajard, C. Jamet, C. Moulin, and S. Thiria, "Use of a neuro-variational inversion for retrieving oceanic and atmospheric constituents from satellite ocean colour sensor: Application to absorbing aerosols," Neural Networks **19**, 178–185 (2006).
51. B. D. Bue, D. R. Thompson, S. Deshpande, M. Eastwood, R. O. Green, V. Natraj, T. Mullen, and M. Parente, "Neural network radiative transfer for imaging spectroscopy," Atmospheric Measurement Techniques **12**, 2567–2578 (2019).
52. C. Fan, G. Fu, A. Di Noia, M. Smit, J. H. H. Rietjens, R. A. Ferrare, S. Burton, Z. Li, and O. P. Hasekamp, "Use of A Neural Network-Based Ocean Body Radiative Transfer Model for Aerosol Retrievals from Multi-Angle Polarimetric Measurements," Remote Sensing **11**, 2877 (2019).
53. M. Gao, B. A. Franz, K. Knobelspiesse, P.-W. Zhai, V. Martins, S. Burton, B. Cairns, R. Ferrare, J. Gales, O. Hasekamp, Y. Hu, A. Ibrahim, B. McBride, A. Puthukkudy, P. J. Werdell, and X. Xu, "Efficient multi-angle polarimetric inversion of aerosols and ocean color powered by a deep neural network forward model," Atmospheric Measurement Techniques **14**, 4083–4110 (2021).
54. F. Aires, C. Prigent, and W. B. Rossow, "Neural network uncertainty assessment using Bayesian statistics with application to remote sensing: 2. Output errors," Journal of Geophysical Research: Atmospheres **109**, (2004).
55. A. G. Baydin, B. A. Pearlmutter, A. A. Radul, and J. M. Siskind, "Automatic differentiation in machine learning: a survey," arXiv:1502.05767 [cs, stat] (2018).
56. M. Gao, K. Knobelspiesse, B. A. Franz, P.-W. Zhai, V. Martins, S. P. Burton, B. Cairns, R. Ferrare, M. A. Fenn, O. Hasekamp, Y. Hu, A. Ibrahim, A. M. Sayer, P. J. Werdell, and X. Xu, "Adaptive Data Screening for Multi-Angle Polarimetric Aerosol and Ocean Color Remote Sensing Accelerated by Deep Learning," Frontiers in Remote Sensing **2**, (2021).
57. D. E. Rumelhart and J. L. McClelland, "Learning Internal Representations by Error Propagation," in *Parallel Distributed Processing: Explorations in the Microstructure of Cognition: Foundations* (MIT Press, 1987), pp. 318–362.
58. S. E. Craig and E. Karaköylü, "Bayesian Models for Deriving Biogeochemical Information from Satellite Ocean Color," (2019).
59. N. Pahlevan, B. Smith, J. Schalles, C. Binding, Z. Cao, R. Ma, K. Alikas, K. Kangro, D. Gurlin, N. Hà, B. Matsushita, W. Moses, S. Greb, M. K. Lehmann, M. Ondrusek, N.

Oppelt, 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).

60. 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., AO **52**, 2019–2037 (2013).

61. L. I. W. McKinna, I. Cetinić, A. P. Chase, and P. J. Werdell, "Approach for Propagating Radiometric Data Uncertainties Through NASA Ocean Color Algorithms," Frontiers in Earth Science **7**, (2019).

62. M. Wang, "A refinement for the Rayleigh radiance computation with variation of the atmospheric pressure," International Journal of Remote Sensing **26**, 5651–5663 (2005).

63. E. P. Shettle and R. W. Fenn, *Models for the Aerosols of the Lower Atmosphere and the Effects of Humidity Variations on Their Optical Properties* (1979).

64. A. Morel, D. Antoine, and B. Gentili, "Bidirectional reflectance of oceanic waters: accounting for Raman emission and varying particle scattering phase function," Appl. Opt., AO **41**, 6289–6306 (2002).

65. 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**, 10909–10924 (1988).

66. C. Cox and W. Munk, "Measurement of the Roughness of the Sea Surface from Photographs of the Sun's Glitter," J. Opt. Soc. Am., JOSA **44**, 838–850 (1954).

67. P. Koepke, "Effective reflectance of oceanic whitecaps," Appl. Opt., AO **23**, 1816–1824 (1984).

68. M. Stramska and T. Petelski, "Observations of oceanic whitecaps in the north polar waters of the Atlantic," J. Geophys. Res. **108**, 3086 (2003).

69. R. Frouin, M. Schwindling, and P.-Y. Deschamps, "Spectral reflectance of sea foam in the visible and near-infrared: In situ measurements and remote sensing implications," J. Geophys. Res. **101**, 14361–14371 (1996).

70. I. E. Gordon, L. S. Rothman, C. Hill, R. V. Kochanov, Y. Tan, P. F. Bernath, M. Birk, V. Boudon, A. Campargue, K. V. Chance, B. J. Drouin, J.-M. Flaud, R. R. Gamache, J. T. Hodges, D. Jacquemart, V. I. Perevalov, A. Perrin, K. P. Shine, M.-A. H. Smith, J. Tennyson, G. C. Toon, H. Tran, V. G. Tyuterev, A. Barbe, A. G. Császár, V. M. Devi, T. Furtenbacher, J. J. Harrison, J.-M. Hartmann, A. Jolly, T. J. Johnson, T. Karman, I. Kleiner, A. A. Kyuberis, J. Loos, O. M. Lyulin, S. T. Massie, S. N. Mikhailenko, N. Moazzen-Ahmadi, H. S. P. Müller, O. V. Naumenko, A. V. Nikitin, O. L. Polyansky, M. Rey, M. Rotger, S. W. Sharpe, K. Sung, E. Starikova, S. A. Tashkun, J. V. Auwera, G. Wagner, J. Wilzewski, P. Wcisło, S. Yu, and E. J. Zak, "The HITRAN2016 molecular spectroscopic database," Journal of Quantitative Spectroscopy and Radiative Transfer **203**, 3–69 (2017).

71. J. Brion, A. Chakir, J. Charbonnier, D. Daumont, C. Parisse, and J. Malicet, "Absorption Spectra Measurements for the Ozone Molecule in the 350–830 nm Region," Journal of Atmospheric Chemistry **30**, 291–299 (1998).

72. K. D. Knobelspiesse, C. Pietras, G. S. Fargion, M. Wang, R. Frouin, M. A. Miller, A. Subramaniam, and W. M. Balch, "Maritime aerosol optical thickness measured by handheld sun photometers," Remote Sensing of Environment **93**, 87–106 (2004).

73. A. M. Sayer and K. D. Knobelspiesse, "How should we aggregate data? Methods accounting for the numerical distributions, with an assessment of aerosol optical depth," Atmospheric Chemistry and Physics **19**, 15023–15048 (2019).

74. V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," in *Proceedings of the 27th International Conference on International Conference on Machine Learning*, ICML'10 (Omnipress, 2010), pp. 807–814.
75. D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," arXiv:1412.6980 [cs] (2017).
76. X. Xiong, J. Sun, X. Xie, W. L. Barnes, and V. V. Salomonson, "On-Orbit Calibration and Performance of Aqua MODIS Reflective Solar Bands," IEEE Transactions on Geoscience and Remote Sensing **48**, 535–546 (2010).
77. D. K. Clark, H. R. Gordon, K. J. Voss, Y. Ge, W. Broenkow, and C. Trees, "Validation of atmospheric correction over the oceans," J. Geophys. Res. **102**, 17209–17217 (1997).
78. B. A. Franz, S. W. Bailey, P. J. Werdell, and C. R. McClain, "Sensor-independent approach to the vicarious calibration of satellite ocean color radiometry," Applied Optics **46**, 5068 (2007).
79. S. W. Brown, S. J. Flora, M. E. Feinholz, M. A. Yarbrough, T. Houlihan, D. Peters, Y. S. Kim, J. L. Mueller, B. C. Johnson, and D. K. Clark, "The marine optical buoy (MOBY) radiometric calibration and uncertainty budget for ocean color satellite sensor vicarious calibration," in *Sensors, Systems, and Next-Generation Satellites XI* (SPIE, 2007), Vol. 6744, pp. 433–444.
80. K. Levenberg, "A method for the solution of certain non-linear problems in least squares," Quart. Appl. Math. **2**, 164–168 (1944).
81. D. W. Marquardt, "An Algorithm for Least-Squares Estimation of Nonlinear Parameters," Journal of the Society for Industrial and Applied Mathematics **11**, 431–441 (1963).
82. S. R. Smith, D. M. Legler, and K. V. Verzone, "Quantifying Uncertainties in NCEP Reanalyses Using High-Quality Research Vessel Observations," Journal of Climate **14**, 4062–4072 (2001).
83. R. Gelaro, W. McCarty, M. J. Suárez, R. Todling, A. Molod, L. Takacs, C. A. Randles, A. Darmenov, M. G. Bosilovich, R. Reichle, K. Wargan, L. Coy, R. Cullather, C. Draper, S. Akella, V. Buchard, A. Conaty, A. M. da Silva, W. Gu, G.-K. Kim, R. Koster, R. Lucchesi, D. Merkova, J. E. Nielsen, G. Partyka, S. Pawson, W. Putman, M. Rienecker, S. D. Schubert, M. Sienkiewicz, and B. Zhao, "The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2)," J. Climate **30**, 5419–5454 (2017).
84. 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. **103**, 24937–24953 (1998).
85. 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., AO **41**, 5755–5772 (2002).
86. G. Zibordi, B. N. Holben, M. Talone, D. D'Alimonte, I. Slutsker, D. M. Giles, and M. G. Sorokin, "Advances in the Ocean Color Component of the Aerosol Robotic Network (AERONET-OC)," Journal of Atmospheric and Oceanic Technology **38**, 725–746 (2021).
87. G. Zibordi, F. Mélin, J.-F. Berthon, B. Holben, I. Slutsker, D. Giles, D. D'Alimonte, D. Vandemark, H. Feng, G. Schuster, B. E. Fabbri, S. Kaitala, and J. Seppälä, "AERONET-OC: A Network for the Validation of Ocean Color Primary Products," J. Atmos. Oceanic Technol. **26**, 1634–1651 (2009).
88. S. W. Bailey and P. J. Werdell, "A multi-sensor approach for the on-orbit validation of ocean color satellite data products," Remote Sensing of Environment **102**, 12–23 (2006).

89. 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, OE **26**, 7404–7422 (2018).
90. C. J. Merchant, F. Paul, T. Popp, M. Ablain, S. Bontemps, P. Defourny, R. Hollmann, T. Lavergne, A. Laeng, G. de Leeuw, J. Mittaz, C. Poulsen, A. C. Povey, M. Reuter, S. Sathyendranath, S. Sandven, V. F. Sofieva, and W. Wagner, "Uncertainty information in climate data records from Earth observation," Earth System Science Data **9**, 511–527 (2017).
91. A. M. Sayer, Y. Govaerts, P. Kolmonen, A. Lipponen, M. Luffarelli, T. Mielonen, F. Patadia, T. Popp, A. C. Povey, K. Stebel, and M. L. Witek, "A review and framework for the evaluation of pixel-level uncertainty estimates in satellite aerosol remote sensing," Atmospheric Measurement Techniques **13**, 373–404 (2020).
92. G. Zibordi, M. Talone, and F. Mélin, "Uncertainty Estimate of Satellite-Derived Normalized Water-Leaving Radiance," IEEE Geoscience and Remote Sensing Letters **19**, 1–5 (2022).
93. B. Bulgarelli and G. Zibordi, "On the detectability of adjacency effects in ocean color remote sensing of mid-latitude coastal environments by SeaWiFS, MODIS-A, MERIS, OLCI, OLI and MSI," Remote Sensing of Environment **209**, 423–438 (2018).