Correspondence to:

D. McKee,
david.mckee@strath.ac.uk

Optical Constituent Concentrations and Uncertainties Obtained for Case 1 and 2 Waters From a Spectral Deconvolution Model Applied to In Situ IOPs and Radiometry

M. Lo Prejato[1] and **D. McKee[1,2]**

[1]Physics Department, University of Strathclyde, Glasgow, Scotland, [2]Department Arctic and Marine Biology, Faculty for Bioscience, Fisheries and Economy, UiT the Arctic University of Norway, Tromsø, Norway

Abstract A spectral deconvolution model (SDM) for inversion of light absorption, $a(\lambda)$ and backscattering, $b_b(\lambda)$, to estimate concentrations of chlorophyll (CHL), colored dissolved organic material (CDOM) and non-biogenic mineral suspended solids (MSS) in offshore and shelf waters is presented. This approach exploits the spectral information embedded in the ratio $b_b(\lambda)/a(\lambda)$, without the need to know each parameter separately. The model has been applied to in situ inherent optical properties (IOPs), $a(\lambda)$ and $b_b(\lambda)$, and to in situ remote sensing reflectance, $r_{rs}(\lambda)$. CHL, MSS, and CDOM estimates are provided by propagating uncertainties in input IOPs and material-specific IOPs using a bootstrapping approach. Application of the SDM to a data set collected in the Ligurian Sea provides Mean Average Errors (MAE) of <0.7 mg m^{-3} for CHL, <0.02 m^{-1} for CDOM, and <0.2 g m^{-3} for MSS. The SDM is found to perform as well as, or in some cases better than, single parameter algorithms and other semi-analytical algorithms (SAA) for each parameter for the Ligurian Sea data set. The SDM CHL product is tested using the NOMAD, Case 1 dominated, global data set and found to perform consistently with the quasi-analytical algorithm (Lee et al., 2002, https://doi.org/10.1364/ao.41.005755) but with slightly poorer performance than standard OCx algorithms. However, the additional estimates of CDOM and MSS provided by the SDM suggest that the approach may be particularly useful for Case 2 waters. Successful retrieval of constituent concentrations with uncertainties suggests good potential to adapt this technique for satellite remote sensing.

Plain Language Summary A key goal for ocean color sensing is to be able to retrieve concentrations of optically significant constituents such as chlorophyll, colored dissolved organic material and inorganic particles or sediment. It is also important to assess the accuracy of these concentrations. We present a spectral deconvolution model that allows us to determine concentrations of all of these materials and to provide an estimate of associated uncertainties. The model operates on the ratio of backscattering to absorption, which is easily derived from measurements of remote sensing reflectance. This gives us a shortcut to bypass the difficult step of estimating absorption and backscattering separately that other similar approaches require, and lets us more directly use the information contained in remote sensing reflectance signals. We show that the approach works similarly well in clear oceanic waters and in shallow coastal waters in the Ligurian Sea (part of the North Western Mediterranean Sea). The spectral deconvolution model performs as well as other established algorithms for each product for the Ligurian Sea and is competitive when applied to the NOMAD global data set. Simultaneous sensing of all three constituent concentrations with robust uncertainties is a useful development for monitoring optical water quality in natural waters.

1. Introduction

Ocean waters are identified as Case 1 or Case 2 waters depending on the composition of optically significant constituents (OSC), including suspended particles such as phytoplankton and non-biogenic particles and colored dissolved organic material (e.g., Mobley, 1994). Since the original definition (Morel & Prieur, 1977), several well-defined criteria were suggested for classifying these two types of waters (Gordon & Morel, 1983; IOCCG, 2000; Mobley et al., 2004; Morel, 1988). According to the current definition, Case 1 waters are waters where phytoplankton and associated colored dissolved organic material (CDOM) and detritus degradation products are the dominant OSCs and together with water determine all of the IOPs. In contrast, Case 2 waters present highly variable concentrations of organic and inorganic materials, such as terrigenous particles, resuspended sediments and CDOM, whose concentrations do not covary with the phytoplankton concentration. A number

of ocean color algorithms have been successfully developed to retrieve chlorophyll concentrations in Case 1 waters (e.g., O'Reilly et al., 1998; Maritorena et al., 2002). However, algorithms specifically developed for Case 1 waters can return inaccurate results if employed in Case 2 waters, and vice versa. For instance, most of the algorithms developed to estimate chlorophyll concentrations in Case 1 waters use the spectral signal at blue-green wavelengths (O'Reilly et al., 1998), while a variety of models for chlorophyll estimation in Case 2 waters are based on reflectance signals in the red and near-infrared wavebands (Gitelson et al., 2008; Yacobi et al., 2011). Whilst there have been attempts to tune existing blue-green algorithms for regional or water-type variations (e.g., McKee et al., 2007) it remains the case that accurate retrieval of CHL in optically complex coastal waters remains a challenge.

A long term goal for ocean color remote sensing is to be able to retrieve concentrations of the chemical proxies for each of the OSCs: chlorophyll—CHL (mg m^{-3}), non-biogenic mineral suspended solids—MSS (g m^{-3}) and colored dissolved organic material—CDOM (m^{-1}), together with estimates of uncertainties for each (Organelli et al., 2016). Each OSC contributes to the inherent optical properties of the water (IOPs) with absorption, $a(\lambda)$ (m^{-1}), and backscattering, $b_b(\lambda)$ (m^{-1}), being particularly important for remote sensing applications. Forward modeling from OSCs via IOPs to ocean color reflectance signals and other apparent optical properties (AOPs) is well-established, with radiative transfer models such as Hydrolight (Numerical Optics Ltd) largely limited by the quality of input data (OSCs and/or IOPs). However, the inverse problem remains a challenge, particularly in optically complex coastal waters where the non-covariance of each OSC is particularly problematic. It is well established that the performance of standard and even Case 2 focused CHL algorithms struggle to reach mission targets in coastal waters (Hooker et al., 1992), and Brewin et al. (2015) demonstrated that performance of existing semi-analytical algorithms generally produced inferior quality CHL estimates than standard empirical algorithms.

Various different inversion approaches have been adopted to estimate constituent concentrations from marine AOPs (e.g., Werdell et al., 2018). In most open ocean waters, considered Case 1 waters (Morel & Prieur, 1977), phytoplankton and co-varying CDOM dominate the non-water signal and retrieval of CHL is often achievable using standard blue-green reflectance ratio algorithms to within mission targets (e.g., ±30%, IOCCG, 2019). However, the performance of these algorithms deteriorates dramatically in optically complex coastal waters because the main OSCs are often widely variable and poorly correlated. The optical variability in coastal regions is influenced by various biogeochemical processes (Cherukuru et al., 2014 and references therein) and sediments and CDOM can significantly influence total IOPs and therefore impact on the performance of empirical reflectance ratio algorithms designed for clear, Case 1 waters (e.g., Fan & Warner, 2014; McKee et al., 2007).

Successful retrieval of CHL in optically complex waters will generally require understanding of, or at least accommodation of, the contributions from other OSCs. This can be achieved using semi-analytical inversion algorithms (SAA). In such algorithms inverting the remote sensing reflectance above or below the water, $R_{rs}(\lambda)$ or $r_{rs}(\lambda)$ respectively (both sr^{-1}), to IOPs relies on approximate solution of the radiative transfer equation (RTE), the algorithm's analytical part, and a method for partitioning total IOPs into their components coupled with estimations of IOP component spectral shapes (the algorithm's empirical part).

Roesler and Perry (1995) presented one of the earliest examples of a SAA to obtain unknown IOPs from AOPs. Another early example is the linear matrix inversion (LMI) technique suggested by Hoge and Lyon (1996). Both approaches (and the subsequent SAA algorithms) extract absorption and scattering properties of oceanic constituents, from surface spectral reflectance measurements, by merging of robust relationships between r_{rs} and the absorption and backscattering coefficients (e.g., Gordon et al., 1988). Hoge and Lyon's algorithm, in particular, using the spectral shapes of the partitioned IOP spectrum, determined by fixed empirical parameters, and reflectances at three wavelengths (412, 490, and 555 nm), derives three unknown IOPs, (b_{bp}, a_{ph}, and a_{dg}) algebraically and simultaneously. Moreover, despite the great variability of water mass types, the LMI technique minimize the errors in IOP spectral models by using a selection of covarying wavelengths, which reduces the impact of the high variability of other wavelengths. Therefore, the linear matrix nature of Hoge and Lyon's model suggests an easy retrieval of the total absorption and, after hyper-spectral data was tested, the usage of the fewest number of co-varying wavelengths (i.e., 3) achieves the best result and ensures a good efficiency in processing satellite images. The SDM presented in this paper shares with Hoge and Lyon's model the same linear matrix nature and the same need to consider only a few wavelengths, with the advantage of directly retrieving the water constituent concentrations through a set of observed SIOPs. It should be noted that in the early stages of the estimation of constituent concentrations from marine AOPs, sensitivity analyses have been frequently used to assess
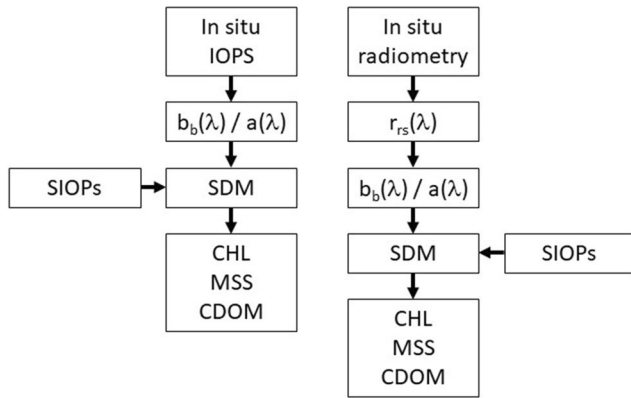
Figure 1. The SDM model can operate on either directly observed in situ IOPs or on values of $b_b(\lambda)/a(\lambda)$ derived from in situ radiometry. In both cases the SDM uses data in the form of $b_b(\lambda)/a(\lambda)$ as input along with SIOPs. Uncertainties in optical inputs and SIOPs are propagated by bootstrapping to produce distributions of estimated CHL, MSS, and CDOM. Final estimates of OSCs consist of medians and 95% confidence intervals for the bootstrapped distributions.

how models affect retrievals (Hoge & Lyon, 1996; Roesler & Perry, 1995). However, early SAAs produced inverted products that were reported without associated uncertainties. The first general approach for quantifying uncertainties in inverted parameters, based on uncertainties in the reflectance data and the model itself, was developed by Wang et al. (2005) and was then modified by Boss and Roesler (Chap 8, IOCCG, 2006) to be applied to remote sensing reflectance (R_{rs}). The SDM proposed in this study has the additional advantage of using a set of SIOPs with associated uncertainties. This feature further enables estimation of uncertainties for the final products of the SDM which in this case are the concentrations of the main optical constituents.

Classical spectral deconvolution methods (SDMs) are SAAs which first determine total absorption and backscattering coefficients from AOPs, usually from $R_{rs}(\lambda)$, and then further partition these values into contributions from specific optical components (phytoplankton, detrital and dissolved matter, e.g., Lee et al., 2002; Smyth et al., 2006) to subsequently estimate water constituent concentrations. Commonly, the relationship used to link $R_{rs}(\lambda)$ to $r_{rs}(\lambda)$ is the formula developed by Lee et al. (2002), $r_{rs}(\lambda) = R_{rs}(\lambda)/[0.52 + 1.7R_{rs}(\lambda)]$, while the relationship between $r_{rs}(\lambda)$ and IOPs follows the approach of Gordon et al. (1988), with $r_{rs}(\lambda)$ as function of the ratio of the backscattering coefficient to the sum of absorption and backscattering coefficients, $b_b(\lambda)/[a(\lambda) + b_b(\lambda)]$. Other classical SDMs, such as the quasi-analytical algorithm (QAA) developed by Lee et al. (2002), do not directly determine constituent concentrations but rather focus on retrieving absorption coefficients of phytoplankton pigments and gelbstoff. It is commonplace to test SDM performance through combinations of simulated and field-measured data sets. However, SDMs are not usually coupled with procedures for direct propagation of measurement uncertainties, even though uncertainty propagation has been analyzed in some cases, notably Lee et al. (2010).

Recently, Ramirez-Perez et al. (2018) proposed a spectral deconvolution model for inversion of light absorption and attenuation data (with scattering derived by subtraction) in optically complex shelf seas, designed for use with in situ AC-9/s submersible spectrophotometer data. The simple formulation of the Ramirez-Perez et al. (2018) SDM facilitated propagation of uncertainties from input IOPs to output constituent concentrations. Moreover, this SDM presents an attractive opportunity to be adapted for remote sensing applications simply by replacing scattering with backscattering. The SDM presented here has been extended to the use of backscattering data, with the distinct advantage of extracting the spectral information contained in the ratio $b_b(\lambda)/a(\lambda)$. This is particularly important for remote sensing applications as it is relatively easy to determine $b_b(\lambda)/a(\lambda)$ from remote sensing reflectance data, and significantly harder to determine $a(\lambda)$ and $b_b(\lambda)$ separately. Thus while the SDM can be used with separate measurements of $a(\lambda)$ and $b_b(\lambda)$ for example, from in situ IOP measurements, it can also be used with estimates of $b_b(\lambda)/a(\lambda)$ derived from remote sensing reflectance measurements (here from in situ radiometry profiling). Importantly, Lo Prejato et al. (2020) have shown that $R_{rs}(\lambda)$ and $r_{rs}(\lambda)$ are both equally well represented as functions of either $b_b(\lambda)/a(\lambda)$ or $b_b(\lambda)/[a(\lambda) + b_b(\lambda)]$ in both Case 1 and Case 2 waters. For this application, the $b_b(\lambda)/a(\lambda)$ version provides the simplest route to a SDM and contains all of the spectral information provided by remote sensing reflectance. The SDM presented here can be applied to (a) in situ measurements of total $a(\lambda)$ and $b_b(\lambda)$, and (b) to remote sensing reflectances measured above or below the sea surface, $R_{rs}(\lambda)$ and $r_{rs}(\lambda)$ respectively (see Figure 1 later). In all cases, concentrations of CHL, MSS, and CDOM can be obtained by inversion of input optical data and uncertainty estimates can be provided by propagating input measurement uncertainties through the SDM.

As for all SAA methods, this inversion procedure requires knowledge of spectral optical properties of OSCs, here the material-specific inherent optical properties (SIOPs). These have been estimated through linear regression of partitioned IOPs against associated constituent concentrations, determined from natural samples, following the partitioning methodology applied by Ramirez-Perez et al. (2018). Using linear relationships gives the advantage of solving the SDM as a system of linear equations, meaning that we always find a single best solution (in a linear least-squares sense), which is not guaranteed with non-linear inversion approaches (which find the local minimum near a given initial guess, but may not successfully find the global minimum). Uncertainties in both

IOPs and SIOPs have been estimated from real sample data and used in a bootstrapping approach to quantify uncertainties associated with results.

The SDM presented here attempts to invert 3 unknowns only (CHL, MSS, and CDOM), therefore a minimum of 3 spectral bands is required to solve the inversion, while a higher number of spectral signals leads to an over-constrained system. The information carried by all the available bands together could be redundant when solving for 3 unknowns only, but it is expected that the uncertainty in the retrieved unknowns will be much lower (Devred et al., 2013) because the use of more wavelengths allows better constraint of the model, resulting in a smaller uncertainty. On the other hand, increasing the number of constraints, when affected by significant uncertainty, could introduce a bias in the final result. Therefore, the performance of the SDM has been tested with a synthetic data set of absorption and backscattering coefficients in order to establish an optimal combination of spectral information and number of bootstrap iterations to successfully retrieve the required OSCs.

The inversion procedure developed in this study is primarily intended for use in Case 2 waters where the ability to simultaneously derive CHL, MSS, and CDOM has most potential impact. However as Brewin et al. (2015) note, there is a qualitative requirement to produce algorithms that "facilitate seamless merging of Case-1 (open-ocean) and Case-2 (coastal optically-complex) waters." It is therefore imperative to establish SDM performance in both Case 1 and 2 waters. As part of that analysis, the SDM can be reformulated as a Case 1 model where the contribution of MSS is assumed to be zero, and the potential negative impact of applying a Case 2 model in Case 1 waters can be assessed. The SDM presented here has the advantage of being applicable to: in situ IOP data, $a(\lambda)$ and $b_b(\lambda)$ directly measured in the field; in situ radiometric data, by retrieving $b_b(\lambda)/a(\lambda)$ from $r_{rs}(\lambda)$; and remotely sensed reflectance data, by retrieving $b_b(\lambda)/a(\lambda)$ from $R_{rs}(\lambda)$, with the IOP-reflectance relationships needed for the last two options available from Lo Prejato et al. (2020). This represents a comprehensive approach to determining key OSCs from in situ and remotely sensed optical measurements that will, we hope, facilitate consistent estimation of OSCs in situ and from space. Importantly, the method is based on simple bio-optical models and standard propagation of measurement uncertainties, making it easy to understand performance across different water types.

In this paper we build on the earlier work presented by Ramirez-Perez et al. (2018) by converting the SDM to operate on the ratio of backscattering to absorption which is very easily related to remote sensing reflectance (Lo Prejato et al., 2020). This formulation of the SDM is tested using a synthetic data set encompassing a very broad range of OSC concentration combinations and using field data collected in the Ligurian Sea covering deep Case 1 and shallow Case 2 waters. Importantly, the SDM is tested using first in situ IOPs and then in situ radiometry as inputs, illustrating the flexibility of the approach. Following the approach presented in Ramirez-Perez et al. (2018), measurement uncertainties in both SIOPs and optical measurements are easily propagated through the SDM using bootstrapping to provide the user with simultaneous estimates of CHL, MSS, and CDOM and uncertainties in each. Performance of the SDM approach is finally compared with existing single parameter and SAA algorithms for both the Ligurian Sea data set and the NOMAD global data set.

2. Theory

2.1. Spectral Deconvolution Model Development

In order to invert IOP measurements and estimate the concentrations of the major OSCs, the approach presented here is built on (a) a bio-optical model relating optical and biogeochemical data, and (b) non-parametric empirical SIOPs, able to describe the spectral shape of all the partial IOPs employed in the model. The bio-optical model presented in this study extends the model developed by Ramirez-Perez et al. (2018) by including backscattering and is the same bio-optical model used in Lo Prejato et al. (2020). Taking advantage of the additive characteristic of IOPs and a practical methodology for partitioning them, each total IOP has been written as sum of partial IOPs, with each one depending on linear contributions of relevant OSCs:

$$a(\lambda) = a_{ph}^*(\lambda)\,\text{CHL} + a_{bdet}^*(\lambda)\,\text{CHL} + a_{ndet}^*(\lambda)\,\text{MSS} + a_{cdom}^*(\lambda)\text{CDOM} + a_w(\lambda) \tag{1}$$

$$b(\lambda) = b_{ph}^*(\lambda)\,\text{CHL} + b_{ndet}^*(\lambda)\,\text{MSS} + b_w(\lambda) \tag{2}$$

$$c(\lambda) = a(\lambda) + b(\lambda) \tag{3}$$

$$b_b(\lambda) = b_{bph}^*(\lambda)\,\text{CHL} + b_{bndet}^*(\lambda)\text{MSS} + b_{bw}(\lambda) \tag{4}$$

Each subscript represents an OSC: phytoplankton, ph, biogenic detritus, bdet, non-biogenic detritus, ndet, colored dissolved organic material, cdom, and water, w. The asterisks denote the constituent-specific SIOPs that is, the rate of absorption, scattering or backscattering per unit concentration of constituent. The biogenic and non-biogenic detrital particulate absorption (a_{bdet} and a_{ndet}) are assumed to co-vary with CHL and MSS respectively, while particulate scattering and backscattering are partitioned into algal and non-algal components which also vary with CHL and MSS respectively. The algal components of scattering and backscattering incorporate biogenic detrital components that are assumed to covary with CHL and are currently impossible to effectively partition from in situ observations. We note that this simple bio-optical model assumes linear relationships between IOPs and constituents and that alternative and possibly more robust approaches would be possible. For example, we have previously developed a set of non-linear relationships between IOPs and constituents for this data set (Bengil et al., 2016). The linear version of the bio-optical model is preferred in this instance as it allows formation of a linear system of equations for which robust solutions are more readily obtained than would be the case for a non-linear set of equations.

The aim of this exercise is to develop a SDM that can ultimately be applied to remote sensing reflectance data (in situ and from space) as well as to in situ measurements of IOPs. A central feature of our approach is that the information contained within either $R_{rs}(\lambda)$ or $r_{rs}(\lambda)$ is retained within the ratio $b_b(\lambda)/a(\lambda)$. Thus it makes sense to develop a SDM based on this ratio. A key benefit of this approach is that it does not require further separation into $a(\lambda)$ and $b_b(\lambda)$ separately (which is non-trivial), but can operate directly on $b_b(\lambda)/a(\lambda)$. To this end Equations 1 and 4 are combined to give

$$\frac{b_b(\lambda)}{a(\lambda)} = \frac{b_{bpph}^*(\lambda)\,CHL + b_{bpndet}^*(\lambda)MSS + b_{bw}(\lambda)}{a_{ph}^*(\lambda)\,CHL + a_{bdet}^*(\lambda)\,CHL + a_{ndet}^*(\lambda)\,MSS + a_{cdom}^*(\lambda)CDOM + a_w(\lambda)} \quad (5)$$

This can be rearranged to give

$$\left[[a_{ph}^*(\lambda) + a_{bdet}^*(\lambda)]\left[\frac{b_b}{a}\right](\lambda) - b_{bpph}^*(\lambda)\right] CHL$$

$$+ \left[a_{ndet}^*(\lambda)\left[\frac{b_b}{a}\right](\lambda) - b_{bpndet}^*(\lambda)\right] MSS$$

$$+ a_{cdom}^*(\lambda)\left[\frac{b_b}{a}\right](\lambda)CDOM$$

$$= b_{bw}(\lambda) - a_w(\lambda)\left[\frac{b_b}{a}\right](\lambda) \quad (6)$$

Absorption and backscattering values by pure water (a_w and b_{bw} respectively) are available from Pope and Fry (1997) and Smith and Baker (1981). Spectrally resolved SIOPs need to be determined for each OSC (see below). CHL, MSS, and CDOM are the 3 unknowns and can be obtained by solving the linear system of equations that Equation 6 subsequently forms (see below).

Equation 6 provides the Case 2 SDM to find CHL, MSS, and CDOM for optically complex coastal waters. However, for oceanic (Case 1) waters, the SDM can be reduced to two variables (CHL and CDOM) and becomes

$$\left[[a_{ph}^*(\lambda) + a_{bdet}^*(\lambda)]\left[\frac{b_b}{a}\right](\lambda) - b_{bpph}^*(\lambda)\right] CHL$$

$$+ a_{cdom}^*(\lambda)\left[\frac{b_b}{a}\right](\lambda)CDOM$$

$$= b_{bw}(\lambda) - a_w(\lambda)\left[\frac{b_b}{a}\right](\lambda) \quad (7)$$

This Case 1 version is presented in order to further assess the impact of using the Case 2 version in Case 1 waters. In practice, it may be difficult to automatically partition data into Case 1 and 2 variants, in which case the expectation is that the Case 2 version would be the default option. This is tested and discussed in more detail below.

The SDM presented here operates on the ratio of $b_b(\lambda)/a(\lambda)$. This makes it particularly accessible for use with radiometric data, including remote sensing data, where this ratio can be obtained from measurements of either $R_{rs}(\lambda)$ or $r_{rs}(\lambda)$ using relationships given in Lo Prejato et al. (2020). It is significantly easier to derive the ratio of

$b_b(\lambda)/a(\lambda)$ than to find $a(\lambda)$ and $b_b(\lambda)$ separately. Lo Prejato et al. (2020) demonstrated that there is no advantage to using the $b_b(\lambda)/[(a(\lambda) + b_b(\lambda)]$ form to relate IOPs to remote sensing reflectance data and in this case the $b_b(\lambda)/a(\lambda)$ form is easier to manipulate in Equations 6 and 7. Figure 1 shows the workflows for operation of the SDM from either in situ IOP measurements or in situ radiometric measurements.

2.2. Solving the SDM and Uncertainty Propagation

Equation 6, with three unknowns, can be written for each available waveband, generating a system of n linear equations. In order to solve the system for three unknowns a minimum of three equations is needed, while with $n > 3$ the system becomes over-determined. In this study, the Matlab® (MathWorks Inc.) *mldivide* operator has been chosen because it can solve any linear equation system giving the best least-squares solution. The *mldivide* operator solves over-determined systems automatically by QR decomposition (also known as a QU factorization), applied to both square and rectangular matrices, providing the number of equations is greater than the number of unknowns. It does not, however, ensure non-negative solutions. We note that other studies have used singular value decomposition to solve over-determined systems (e.g., Brando et al., 2012). Both approaches appear to work similarly well for this type of problem.

As recently reminded by the IOCCG (International Ocean-Colour Coordinating Group) in their 18th Report (IOCCG, 2019), *"there is no reason to trust a measurement value without an associated uncertainty"*. Following this principle, uncertainties have been established for all measured values used as model input, and are propagated through the model using bootstrapping in order to estimate the uncertainty for retrieved OSC concentrations. Bootstrapping is a Monte Carlo method for estimating uncertainties in model-derived products (Davison & Hinkley, 1997). According to this procedure, input data are perturbed B times, thus returning an ensemble of B potential solutions. Perturbations are informed by the distribution of uncertainties for each input parameter. By randomly allocating an uncertainty estimate to each input value, it is possible to generate output distributions which are representative of the combined effect of uncertainties in all of the input variables. In this study, the median of this distribution of solutions is considered as the optimal solution, with an associated uncertainty given by 1.96 times the standard deviation (σ_{boot}) of the bootstrap distribution (95% confidence interval, 95%CI).

Uncertainties assigned to the input total IOPs were wavelength-independent since they were estimated through a residual analysis of the geometric mean regression (GMR) between two signals of the same instrument, closely spaced spectrally (McKee et al., 2009—see Section 3.2) Uncertainties in the ratio $b_b(\lambda)/a(\lambda)$, when this ratio is derived from a IOP-reflectance relationship applied to $r_{rs}(\lambda)$ data, were estimated through bootstrapping of wavelength-dependant uncertainties in $r_{rs}(\lambda)$ data (see Section 3.2). Uncertainties associated with SIOPs were wavelength-dependant and quantified using 95% confidence interval (95%CI) of linear regressions slopes between partitioned IOPs and related constituent concentrations (McKee & Cunningham, 2006; McKee et al., 2014—see Section 3.3). Uncertainties were assumed to be normally distributed. For each iteration of the bootstrap process, uncertainty in each input parameter was randomly assigned within the constraints of the uncertainty estimates established for that parameter.

During testing it was noticed that the quantity and quality of spectral information considered by the SDM affected the final result. Bootstrap estimators cannot be assumed to be unbiased, with the bias potentially depending on the nature of the model representation or it can originate from a conjunction of skewness of the sampling distribution and small sample size. Moreover, the recommended number of bootstrap iterations (B) may vary depending on the case study and the statistical parameter one wants to infer (Mooney & Duval, 1993). Therefore the SDM was tested with a synthetic data set using different numbers of bootstrap iterations and subgroups of spectral bands in order to assess the behavior of both the bias and random uncertainty, and to determine optimal combinations of both.

3. Methods

3.1. Field Sampling

The present study is based on data collected in the Ligurian Sea (LS), North Western Mediterranean Basin, on board *NRV Alliance* in March 2009 (Figure 2). Data from this cruise (BP09) have been published previously (Bengil et al., 2016; McKee et al., 2013, 2014; Ramirez-Perez et al., 2018). The data presented here includes
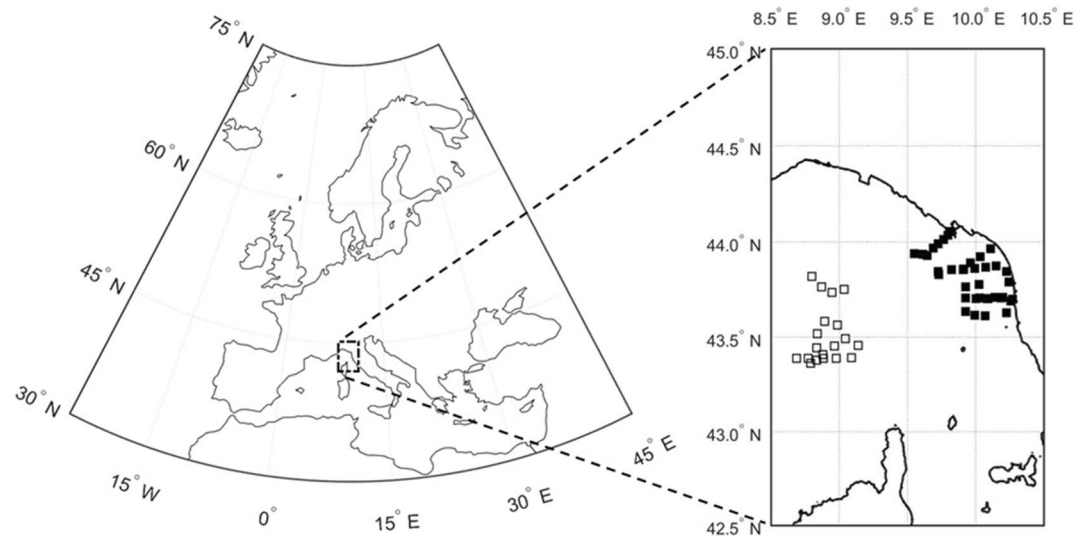
Figure 2. In the left map, the small dashed rectangle shows the study area, located in the North Western Mediterranean Sea. On the right, the positions of the sampling stations (BP09 cruise), in the Ligurian Sea, classified as offshore (empty squares) and onshore (filled squares). Reproduced from Lo Prejato et al. (2020), their Figure 1.

11 stations in deep, clear oceanic waters (considered Case 1) and 23 stations in shallow, turbid coastal waters (considered Case 2), in proximity to the plumes of the river Arno and other minor rivers. In the Arno river mouth, concentrations of chlorophyll are frequently higher than 1 mg m^{-3}, while in the study area, spring blooms of variable intensity occur typically from the end of February until the beginning of May (Bosc et al., 2004). The coastal data set is considered to be Case2 due to the presence of suspended inorganic sediment (from the Arno plume and resuspended from the seabed) and CDOM associated with freshwater inputs. During the BP09 cruise, a spring bloom was recorded offshore and the ranges of concentration of in water constituents were: 0.29–3.31 mg m^{-3} for chlorophyll; 0.13–3.77 g m^{-3} for total suspended solids and 0.012–0.190 m^{-1} for colored dissolved organic matter.
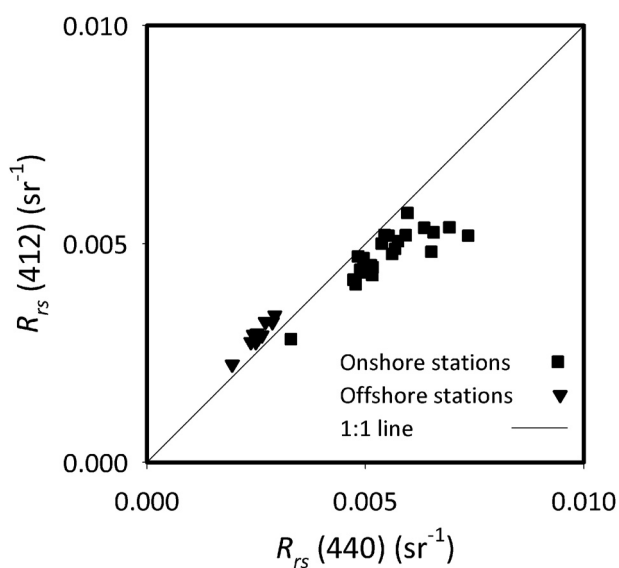

Figure 3. Classification of Case 1 and Case 2 waters based on reflectance data (Matsushita et al., 2012). Offshore waters verify the criteria for Case 1 waters: $R_{rs}(412) \geq R_{rs}(443)$. Spectral signal at 440 nm (instead of 443 nm) has been used to apply the criterion.

Ramirez-Perez et al. (2018) previously showed that offshore, deep oceanic waters had a lower particulate backscattering ratio and concentration of total suspended solids compared to the shallow coastal waters where non-biogenic material was abundant. According to Matsushita et al. (2012), $R_{rs}(412) \geq R_{rs}(443)$ occurs in Case 1 waters, and vice versa in Case 2 waters. Figure 3 shows the application of this criterion to the Ligurian Sea data set: the signal at 440 nm has been used in place of the 443 waveband and the reflectance data measured below the surface, $r_{rs}(\lambda, 0^-)$, have been converted to remote sensing reflectance in air, $R_{rs}(\lambda, 0^+)$ through the relationship $R_{rs}(\lambda, 0^+) = 0.518\, r_{rs}(\lambda, 0^-)/[1-1.562 r_{rs}(\lambda, 0^-)]$ (Lee et al., 1998). The two geographically separated subsets of offshore and onshore stations conform to the Matsushita et al. classification for Case 1 and Case 2 waters. We note that this classification scheme may not be universally robust but appears to perform to expectations for these waters.

3.2. In Situ Optical Measurements

Non-water absorption $a_n(\lambda)$ and attenuation $c_n(\lambda)$ coefficients were measured in situ with a 25 cm path length AC-9 reflective tube attenuation and absorption meter (WETLabs Inc.) operating at 9 wavebands centered at 412, 440, 488, 510, 532, 555, 650, 676, and 715 nm. IOP measurements were corrected for salinity and temperature using data from a Seabird SBE 19 plus CTD. AC-9 absorption data was corrected for scattering errors using the proportional
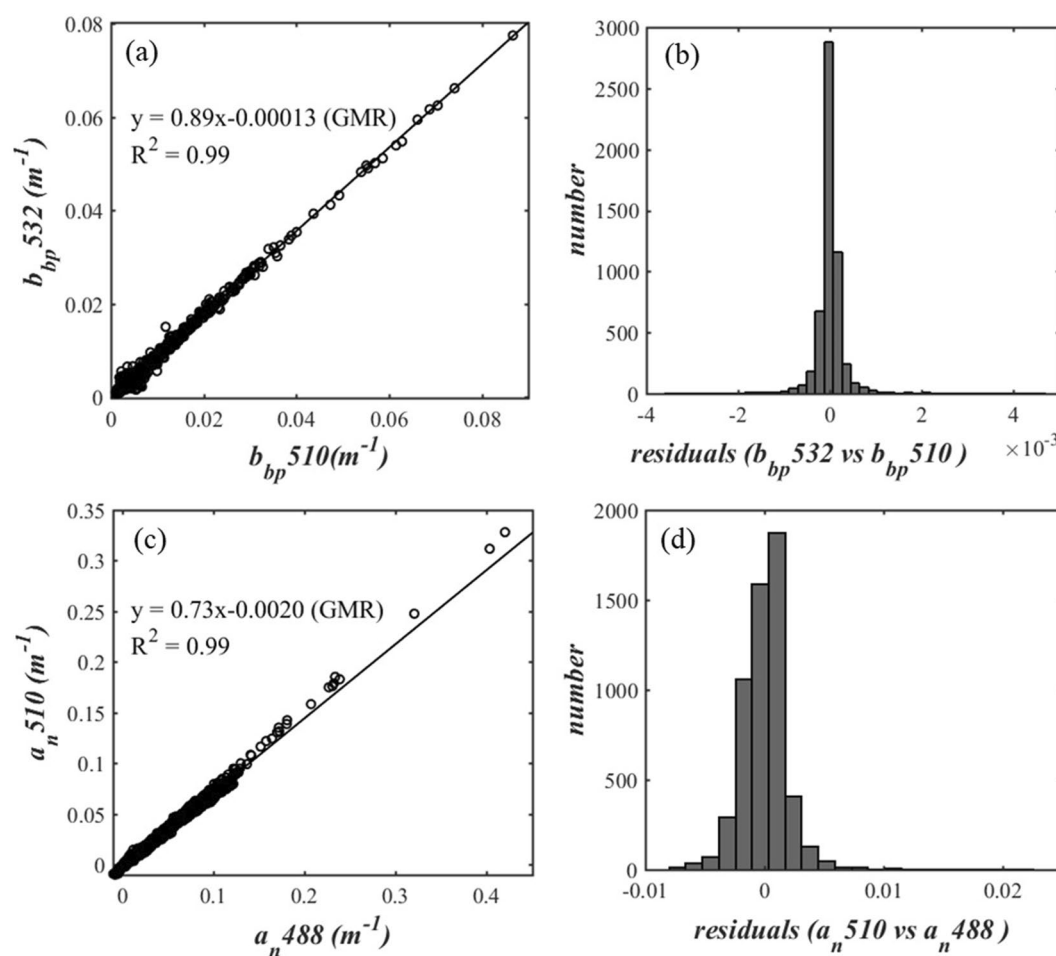
Figure 4. Measurement uncertainties estimated for both backscattering and absorption data from GMR and 95th percentile, according to McKee et al. (2009): (a) GMR of particulate backscattering (b_{bp}) signals at two close wavelengths, 510 and 532 nm and (b) related residual histogram. (c) GMR of $a_n(510)$ against $a_n(488)$ and (d) residual histogram.

method (Zaneveld et al., 1994). A WETLabs BB-9 was mounted alongside the AC-9 to measure backscattering coefficients at 412, 440, 510, 532, 595, 660, 676, and 715 nm. Backscattering data were subsequently interpolated to match AC-9 wavelengths and corrected for path length absorption using proportionally corrected AC-9 absorption data. The particulate volume scattering function (VSF) was derived from the total VSF measured by the BB-9 at an effective scattering angle of 117°, after subtraction of water VSF obtained from Morel (1974), while the particulate backscattering coefficients, $b_{bp}(\lambda)$, were estimated from measurements of $\beta_p(117°,\lambda)$ using a χ-factor of 1.1 (Boss & Pegau, 2001). The total backscattering coefficient, $b_b(\lambda)$ was then calculated as the sum of $b_{bp}(\lambda)$ and the backscattering from pure water, $b_{bw}(\lambda)$ derived from Smith and Baker (1981).

Random measurement uncertainties were estimated for both absorption and backscattering data using the approach of McKee et al. (2009). This method is based on residual analysis of the GMR between two signals of the same instrument, closely spaced spectrally. The approach is based on the principle that if the spectral difference of the IOP is stable (i.e., natural variability is minimal) the residual variability between each set of data can be attributed to random measurement uncertainties. The magnitude of the measurement uncertainty range for each parameter can be estimated by taking the 95th percentile of the absolute distribution, broadly equivalent to 95% confidence interval (CI) of a traditional linear regression. The resulting distributions are likely to slightly overestimate the random measurement uncertainty (some of the variability will be due to natural variations in spectral dependence), but the procedure provides reasonable ballpark figures.

Figure 4a shows particulate backscattering data at two wavelengths (532 and 510 nm) plotted against one another for the entire data set. The GMR coefficient of determination is close to unity and the very small offset suggest

that the residual variability (Figure 4b) can be reasonably referred to random uncertainties, with the 95th percentile of the resulting distribution giving a magnitude of $\pm6.1 \times 10^{-4}$ m^{-1}. The same approach has been applied to non-water absorption data at 488 and 510 nm (Figures 4c and 4d), giving ±0.0036 m^{-1} for the residual measurement uncertainty. This value is consistent with the operational uncertainty for an AC-9 that has been calibrated and corrected with optimal accuracy (IOCCG Protocol Series, 2018). The assumption that spectral dependence of $a_n(\lambda)$ is small and uniform for a small wavelength difference probably does not hold as well as it does for $b_{bp}(\lambda)$, but the approach does seem to provide reasonable estimates of random measurement uncertainty for both instruments. In practice there may be other systematic errors that remain unaccounted for in this analysis, but we are not aware of any means to determine them.

Upwards radiance, $L_u(z,\lambda)$ and downwards irradiance, $E_d(z,\lambda)$, were measured using a free-falling HyperPro profiler (Satlantic Inc.) configured with hyperspectral sensors. Downwards above-surface irradiance data, $E_s(t,\lambda)$, were also collected by a reference sensor positioned on the ship's superstructure. The profiling radiometer was deployed in a multi-cast mode to sample the surface layer multiple times at each station. All underwater radiometric data were corrected for changes in solar elevation during the cast sequence and to minimize the impact of wave-induced effects (Zibordi et al., 2004), broadly following Sanjuan Calzado et al. (2011). Only data from the top 3 m were used during the correction routine. This depth was considered adequate to establish good linear fits to log-transformed data and shallow enough to avoid most changes in water column structure and to have reasonable signals in blue and red channels. $L_u(z, \lambda)$ and $E_d(z,\lambda)$ were extrapolated to the surface after applying a linear regression to the ensemble of sub-cast profiles of log-transformed data, concurrently removing outliers associated with wave focusing effects. Uncertainties in $L_u(0^-, \lambda)$ and $E_d(0^-,\lambda)$ just below the sea surface were estimated as 95% confidence intervals (CI) of the intercepts of the best fit lines (Figure 5). The remote sensing reflectance below the water, $r_{rs}(0^-, \lambda)$, was estimated through the ratio of $L_u(0^-, \lambda)$ to $E_d(0^-, \lambda)$ and the associated uncertainty was estimated by propagating uncertainties in $L_u(0^-, \lambda)$ and $E_d(0^-,\lambda)$ using standard quadrature. Figure 5c shows an example $r_{rs}(\lambda)$ spectrum with corresponding wavelength dependent uncertainties. We note that additional factors such as temperature dependence, stray light and accuracy of cosine response have not been considered here. As such our estimates of radiometric uncertainty should not be considered as a complete analysis, but do provide a first order estimate of uncertainty associated with the processing steps outlined above.

3.3. Laboratory Measurements

The absorption of all dissolved and suspended components minus water was measured using a Point Source Integrating Cavity Absorption Meter (PSICAM), with an estimated accuracy of $\pm2\%$ even in highly turbid waters (Röttgers et al., 2005, 2007). Measurements were made using fresh Milli-Q ultrapure water references and then corrected for salinity and temperature effects on water absorption (Röttgers & Doerffer, 2007; Röttgers et al., 2014). Measurements of absorption by colored dissolved organic matter, $a_{cdom}(\lambda)$, were performed using a 1 m liquid waveguide capillary cell (LWCC) with an Ocean Optics USB2000 minispectrometer, with a noise range of ±0.0001 m^{-1} at 532 nm (Lefering et al., 2017). In this study, the concentration of colored dissolved organic material is represented by its absorption at 440 nm, $CDOM = a_{cdom}(440)$ following for example, Kirk (2011). Total particulate absorption, $a_p(\lambda)$ was derived by subtracting $a_{cdom}(\lambda)$ from PSICAM non water absorption. Total particulate absorption was also obtained by measuring the optical density of particles, $OD_p(\lambda)$ (nondim), retained on a 25 mm GF/F filter using a Shimadzu UV-2501 PC benchtop spectrophotometer in transmission mode. After applying dilute bleach and rinsing with filtered seawater to remove algal pigments, the measurement was repeated for each sample to provide the detrital optical density of non-algal particles, $OD_{det}(\lambda)$. The particulate and detrital absorption coefficients, $a_p(\lambda)$ and $a_{det}(\lambda)$ respectively, were obtained from

$$a_x(\lambda) = 2.303 \frac{A_{fp} OD_x(\lambda)}{V_f \, \beta} \tag{8}$$

where $a_x(\lambda)$ is the required absorption coefficient, A_{fp} is the exposed area of the filter pad, V_f is the volume of sample filtered (m^3), β is the path length amplification factor (nondim) and $OD_x(\lambda)$ is the required optical density. A linear regression approach (Lefering, Röttgers, et al., 2016; McKee et al., 2014) using a combination of PSICAM and LWCC data was employed for the determination of path length amplification factors (β) and scattering offset corrections for both filter pad absorption measurements. Absorption by phytoplankton, $a_{ph}(\lambda)$, was derived by subtracting absorption of detrital particles from the total particulate absorption, $a_{ph}(\lambda) = a_p(\lambda) - a_{det}(\lambda)$.
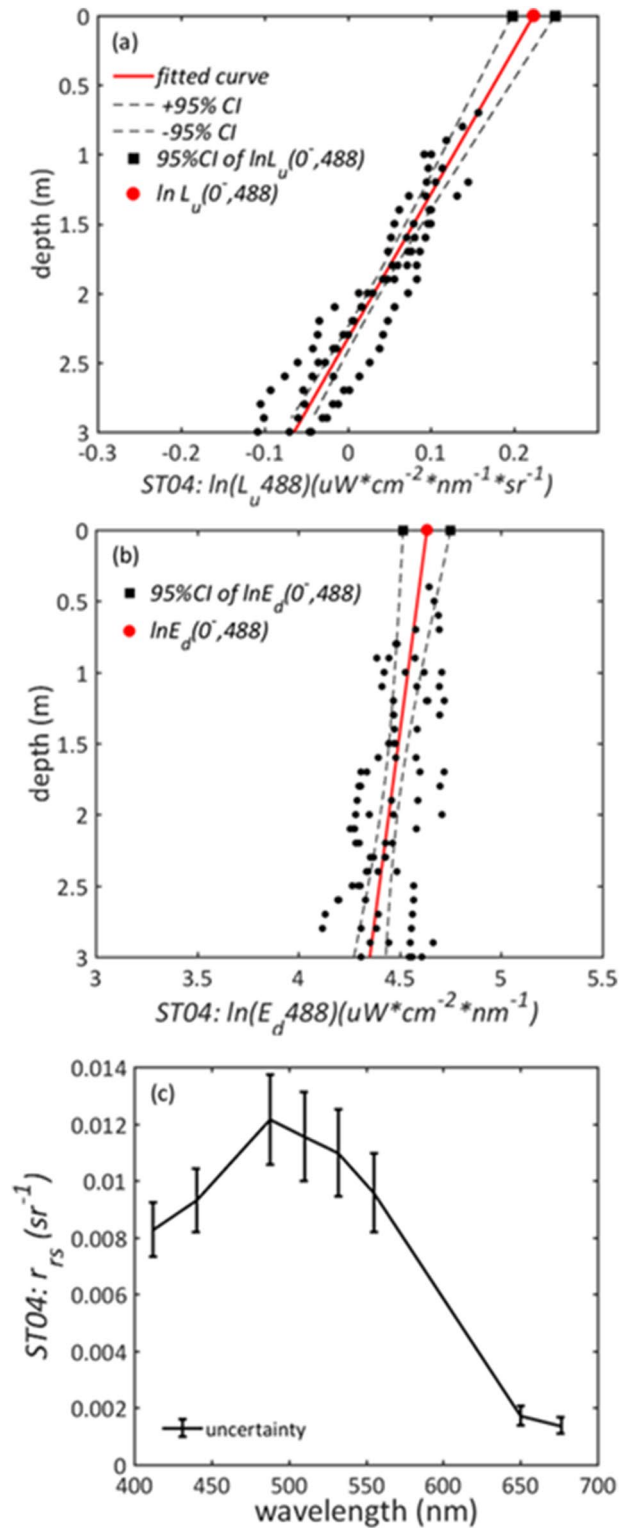
Figure 5. Example of radiometric measurements at one station (ST04): multicast data ensemble for (a) upwards radiance, L_u, and (b) downwards irradiance, E_d, at 488 nm, with best fit regression line (red lines) and 95% confidence intervals (black dashed lines). The red dot is the intercept of the regression line for (a) $L_u(0^-)$ and (b) $E_d(0^-)$, and black squares are associated 95% CIs on the intercepts. (c) The remote sensing reflectance spectrum below the water, $r_{rs}(0^-, \lambda)$, and associated uncertainty are shown as a black line and vertical error bars, respectively.

Concentrations of chlorophyll-*a* (CHL) and total suspended solids (TSS) were measured in triplicate by colleagues from Management Unit of the North Sea Mathematical Models (MUMM) with averaged results reported here. CHL concentrations were determined by standard HPLC measurements: triplicate samples were analyzed using a reversed phase, acetone-based method with a C18 column and a Jasco FP-1520 fluorescence detector. TSS concentrations were estimated using pre-ashed, rinsed and pre-weighed 47 mm GF/F filters, careful rinsing each sample with ultrapure water. Filters were stored frozen and returned to the lab where they were dried and reweighed. A simple relationship between TSS and CHL for Case 1 waters, obtained by least squares regression forced through zero, gives biogenic TSS (TSS_{bio}) per unit CHL ($TSS_{bio} = 0.234$ CHL). This is used to estimate the concentration of biogenic TSS for Case 2 waters, with mineral (i.e., non-biogenic) suspended solids (MSS) obtained for Case 2 waters by subtracting the estimate of TSS_{bio} from total TSS (Bengil et al., 2016).

SIOPs were obtained using the linear regression approached initially proposed by McKee and Cunningham (2006) and then applied by Ramirez-Perez et al. (2018) for an analogous IOP inversion exercise. Spectral SIOPs were calculated as the slopes of simple linear regressions forced through the origin of partitioned IOPs against associated constituent concentrations using natural samples from the surface layer of the ocean. SIOP uncertainties were estimated as the 95% CIs for regression slopes. More details and a complete explanation of the partitioning methodology can be found in Ramirez-Perez et al. (2018). The full set of SIOPs obtained in this manner are shown in Lo Prejato et al. (2020, Figure 2).

4. Results

Development and testing of the SDM is presented in multiple stages below. In the first stage (Section 4.1) a synthetic data set is used to test performance of the SDM across a wide range of OSC concentration ranges and to establish optimal numbers of spectral bands and bootstrap iterations. Real world performance of the SDM is then tested using first in situ IOP measurements (Section 4.2) followed by in situ radiometry (Section 4.4), with Section 4.3 focusing on conversion of r_{rs} to b_b/a using in situ data. The performance of the SDM is then compared with those of alternative algorithms using the Ligurian Sea data set (Section 4.5). Finally, the general applicability of the SDM approach is tested using the NOMAD global data set (Section 4.6).

4.1. Sensitivity Analysis Using Synthetic Data Set

The performance of the Case 2 SDM was evaluated through a sensitivity analysis based on a synthetic data set of 1,690 combinations of fixed log-spaced constituent concentrations ($0.01 < CHL < 100$ mg m^3, $0.01 < MSS < 100$ g m^3, $0.01 < CDOM < 10$ m^{-1}) used to generate total absorption and total backscattering coefficients and to populate Equation 6 (see Lo Prejato et al., 2020 for details). This synthetic test data set was put through the Case 2 SDM with bootstrapping used to propagate uncertainties. The bootstrap median (median$_{boot}$) for each OSC concentration was adopted as the best estimate, while 95% confidence intervals (CI) of the bootstrap distribution was used to estimate the effect of error propagation on estimates of OSC concentrations. For the sake of comparison, the relative CI (CI%) and the bias percentage error (Bias%) have been calculated using

$$CI\% = \left(\frac{1.96\sigma_{boot}}{mean_{boot}} \right) \times 100 \qquad (9)$$

$$Bias\% = \left(\frac{median_{boot} - input\ value}{input\ value} \right) \times 100 \qquad (10)$$

Bias% provides an estimate of how close the median of the bootstrap distribution lies to the known *input value* from the synthetic data set. CI% gives a measure of the spread of the bootstrap distributions.

The fundamental performance of the Matlab solver was established by solving the SDM with zero uncertainties in the IOPs and SIOPs, with the resulting bootstrap results being both accurate and unbiased (both Bias% and CI% $<10^{-12}$). This indicates that in the absence of measurement uncertainties, and assuming the bio-optical model is properly representative of the optical properties of the sample, the solver would always return essentially perfect results. In practice, of course, there will always be uncertainties in both the input measurements and the SIOPs, there may be real natural variability in SIOPs (e.g., Bricaud et al., 1995, but also see McKee et al., 2014), and absence or presence of specific OSCs may impact on the structure of the bio-optical model leading to variable quality of estimated OSC concentrations.
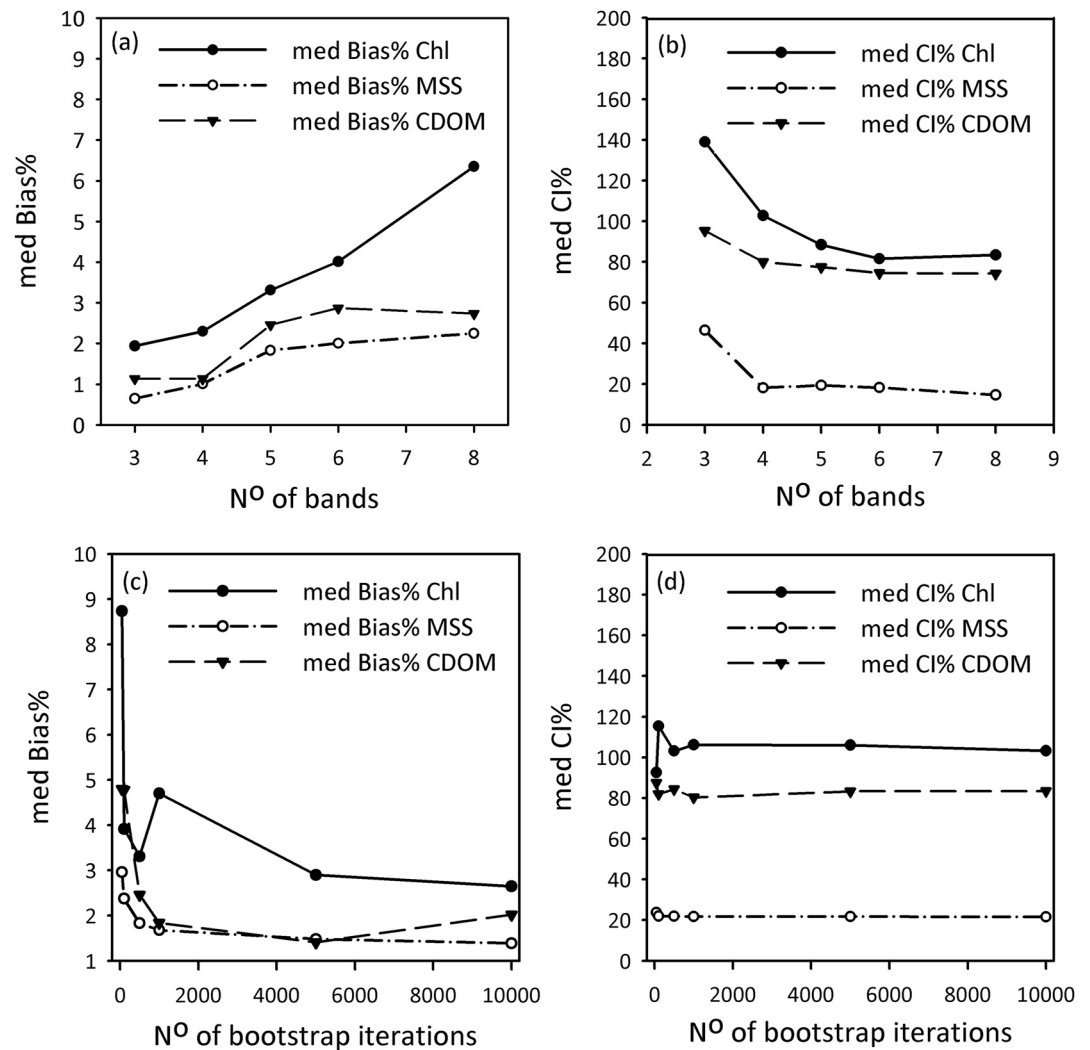
Figure 6. Median value of (a) bias percentage error (med Bias%) on CHL, CDOM, MSS, and (b) median value of 95% Confidence Intervals (med CI%) when considering different subsets of spectral signal and 500 bootstrap iterations. Trend of (c) med Bias% and (d) med CI% for increasing number bootstrap iterations, considering the spectral subgroup including 5 bands.

Applying the Case 2 SDM to the full synthetic data set and propagating uncertainties in IOPs and SIOPs demonstrated that the Bias% of each OSC increased dramatically as the contribution of that OSC to formation of optical signals diminishes at low concentrations compared to other OSCs, analogous to results previously shown by Ramirez-Perez et al. (2018) for the model operating on $b(\lambda)/a(\lambda)$. The full synthetic data set covers a very large range of variability in OSC concentrations that goes far beyond that encountered in the BP09 field experiment. In an effort to constrain the sensitivity analysis to more representative ranges, a reduced synthetic data set was selected ($0.2 < CHL < 5$ mg m^{-3}, $0.1 < MSS < 5$ g m^{-3}, $0.02 < CDOM < 0.1$ m^{-1}) corresponding to the BP09 ranges of concentrations. This subset of synthetic data was used to test the impact of choosing both the number of wavebands used in the SDM and the number of bootstrap iterations. Five different waveband subgroups were selected (412–440–488–510–532–555–650–676 nm; 412–440–488–532–555–650 nm; 412–440–488–555–650 nm; 412–488–555–650 nm; 412–488–555 nm) with a maximum of 8 bands, excluding the signal at 715 nm, and a minimum of 3 bands, required to solve a system with 3 unknowns. The selection of the spectral bands included in each subgroup was based on optical characteristics of the OSCs, taking into consideration the possibility that if some of the bands are highly correlated with each other, a subset of spectral information can be as effective as the full spectral range available for the derivation of the required OSC concentrations (Lee & Carder, 2002; Sathyendranath et al., 1994; Wernand et al., 1997). Figure 6 (a and b) shows the impact of varying

the number of wavebands on the retrievals of CHL, MSS, and CDOM. The SDM has been applied to the reduced synthetic data set with the median values of Bias% and CI%, medBias% and medCI% respectively, showing the overall effect on each OSC. Increasing the number of wavebands increases the medBias% almost linearly though at relatively small values, while medCI% rapidly decreases but ultimately plateaus. From this analysis we conclude that 5 wavebands provide an optimal combination of Bias% and CI% for this SDM. The 5 wavebands version of the model was subsequently tested by varying the number of bootstrap iterations ($B = 50, 100, 500, 1,000, 5,000, 10,000$). Figure 6 (c and d) demonstrate that increasing the number of iterations has essentially no impact on medCI% but significantly reduces medBias% until $B = 500$, beyond which there is no further improvement. From this analysis we conclude that the 5 wavebands SDM operating with 500 bootstrap iterations provides optimal output quality and computational efficiency and is used hereafter throughout the paper in Case 1 and Case 2 versions of the SDM.

The distribution of Bias% and CI% for CHL, MSS, and CDOM concentrations within the LS range is shown in Figure 7. The highest values of Bias% ($\sim20\%$) for CHL occurred when the model was used to retrieve the smallest value (~0.1 mg m^{-3}) of CHL in the range (Figure 7a). For both MSS and CDOM, highest Bias% ($\sim25\%$) occurs when retrieving small values of these quantities with high levels of CHL (Figures 7c and 7e). CI% estimates exhibit similar behavior, with highest values occurring when trying to retrieve low concentrations of an OSC when other OSCs are high. This is a physically realistic result. If an OSC is making only a relatively small contribution to the optical signals, there ought to be more uncertainty in retrieving its concentration. It is unrealistic to expect to be able to retrieve small concentrations of CHL accurately in the presence of high concentrations of MSS and/or CDOM, as discussed by for example, Doerffer and Schiller (1994). Note, however, that this is not necessarily a property of optical complexity per se. If it was, the results of the analysis of synthetic data with zero noise would not have been essentially perfect. Instead, it is important to realize that the ability to resolve small contributions to the reflectance signal is limited by the associated uncertainties in measurements and modeling parameters, in this case b_b/a and the SIOPs.

4.2. Performance of SDM for b_b/a From in Situ IOPs

The Case 2 SDM was applied to in situ AC-9 absorption and BB-9 backscattering measurements with resulting estimates of CHL, CDOM and MSS shown in Figure 8 (a, b and c respectively). Figure 8c illustrates the generation of non-zero MSS estimates for offshore (Case 1 waters) when the Case 2 model is applied. Implementing the Case 1 SDM forces MSS to zero a priori. The resulting concentrations of CHL and CDOM for the Case 1 SDM applied in offshore waters are shown as gray triangles in Figures 8a and 8b.

The estimated concentration values are medians of the bootstrap distributions after applying the SDMs with $B = 500$ and using the spectral information from five wavebands (412–440–488–555–650 nm). The vertical error bars represent the 95% CI of the bootstrap distribution and originate from the propagation of the combined uncertainty on both the optical measurements and the SIOPs. The Mean Absolute Error (MAE) is presented as a metric to evaluate the performance of the SDMs.

$$\text{MAE} = \frac{\sum_{i=1}^{n} |y_i - x_i|}{n} \tag{11}$$

where x_i and y_i are measured and estimated values of each parameter respectively and n is the total number of available matchups. MAE has been calculated separately for onshore and offshore sampling stations in order to allow comparisons between the Case 1 and Case 2 versions of the SDM (Table 1). The Case 2 SDM applied to onshore waters returned MAEs of 0.60 mg m^{-3} for CHL, 0.012 m^{-1} for CDOM and 0.18 g m^{-3} for MSS. For offshore waters, the Case 2 SDM produced slightly lower accuracy CHL and CDOM values compared to the Case 1 SDM, but the differences are fairly marginal suggesting that the Case 2 SDM could be used in all waters with limited problems. Interestingly, the Case 2 SDM returns low but non-zero MSS values for offshore waters that are inconsistent with our assumption of zero MSS for these stations. This may indicate the presence of small amounts of inorganic material in these waters, possibly associated with diatoms forming the bulk of the offshore algal bloom, but as discussed below, this may also reflect residual issues with the quality of in situ IOP measurements.

4.3. Conversion of r_{rs} to b_b/a

A new set of IOP-reflectance relationships based on extensive Hydrolight simulations has recently been published (Lo Prejato et al., 2020) allowing retrieval of IOPs in the form of the ratio $b_b(\lambda)/a(\lambda)$ from under water
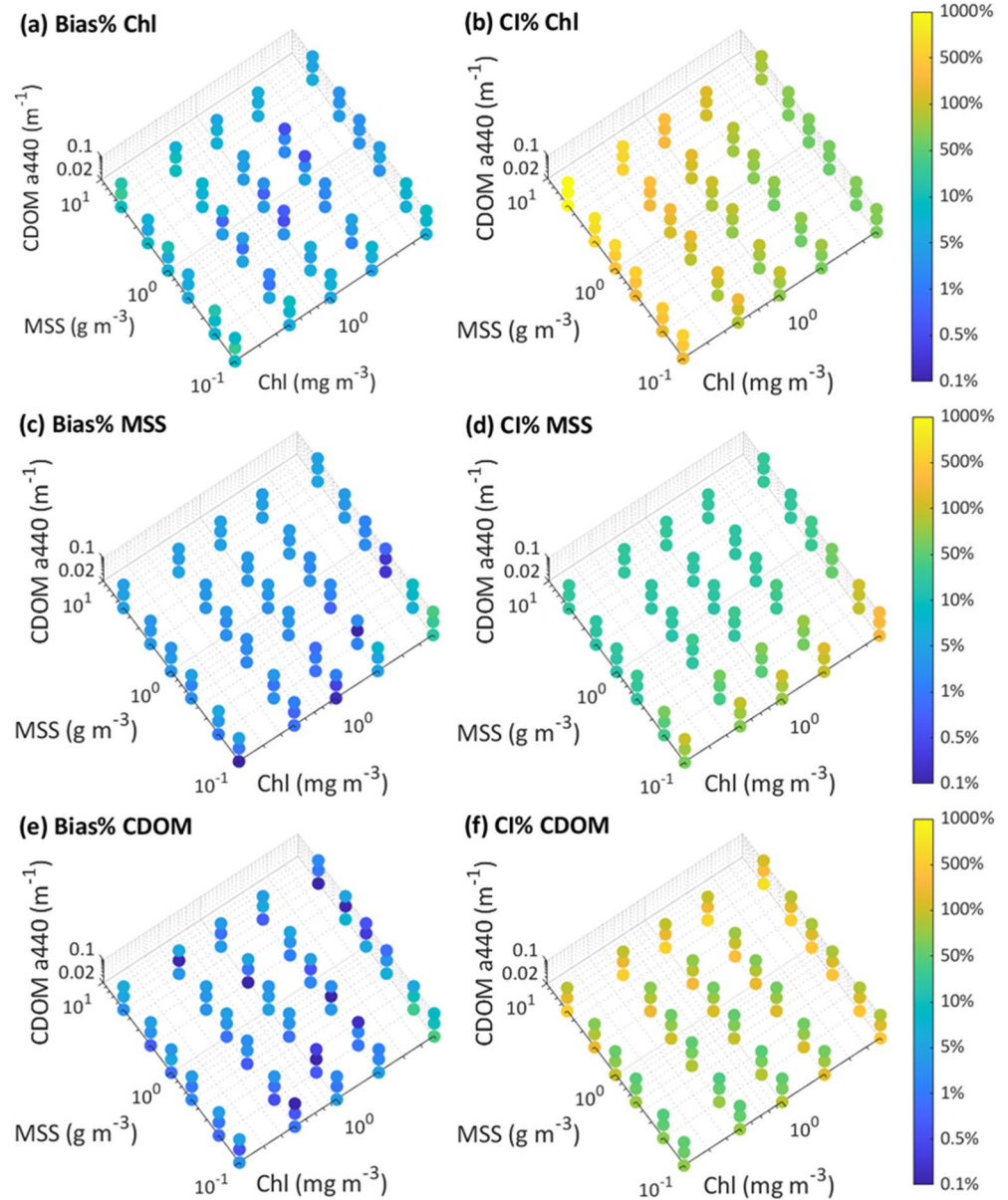
Figure 7. Distribution of the offset percentage error (Bias%) for (a) CHL, (c) MSS, and (e) CDOM and the corresponding relative 95% confidence intervals (CI%) of the bootstrap mean distribution for (b) CHL, (d) MSS, and (f) CDOM. The LS range of concentrations, 500 bootstrap iterations and 5 wavebands (412; 440; 488; 555; 650 nm) have been considered.

remote sensing reflectance, $r_{rs}(\lambda, 0^-)$. The 5th order polynomial relationship suggested in Table 1 of Lo Prejato et al. (2020) has been applied to the radiometric data of this study and the resulting data set of $b_b(\lambda)/a(\lambda)$ values has been used as input for the SDM. The associated uncertainties in $r_{rs}(\lambda)$, estimated from real samples (see Section 3.2) have been used in a bootstrapping framework to quantify uncertainties associated with resulting $b_b(\lambda)/a(\lambda)$ values. The uncertainty attributable to the polynomial relationship itself has been considered negligible compared to the uncertainties derived from $r_{rs}(\lambda)$ measurements.

The comparison between measured and retrieved $b_b(\lambda)/a(\lambda)$ presented in Figure 9 shows a clear bifurcation, with one branch deviating strongly from the 1:1 line for higher values of the ratio $b_b(\lambda)/a(\lambda)$. The source of this overestimation of the retrieved ratios is not clear. Analysis of systematic errors in AC-9 data through application of alternative scattering correction methods for absorption and attenuation measurements (not shown) did not significantly improve consistency with modeled data, broadly in line with findings from Lefering, Bengil,
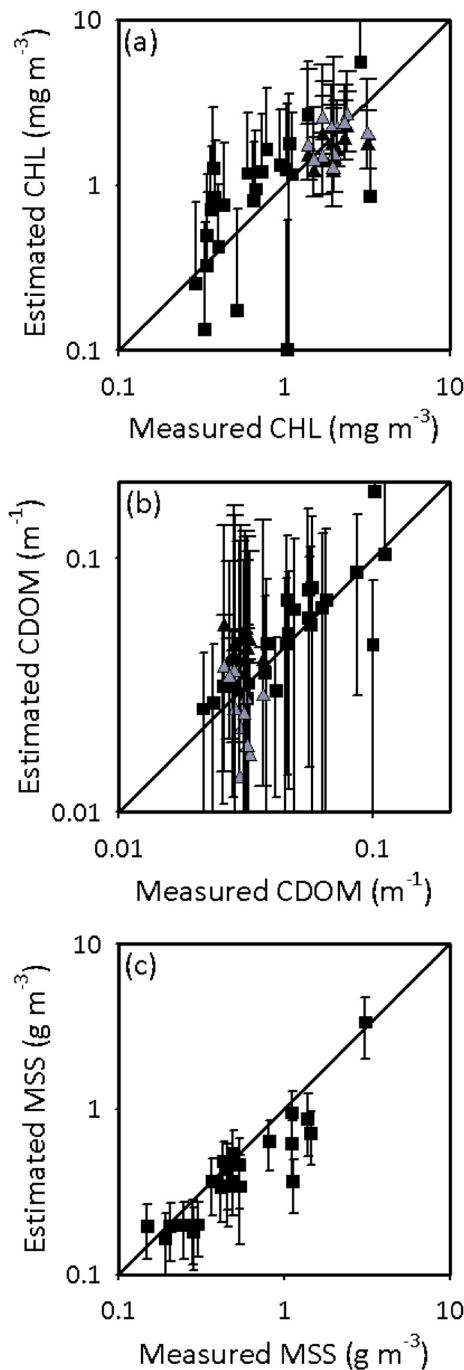
et al. (2016) which showed that light field calculations were not strongly influenced by this potential source of error. Other potential sources of error include problems with correction of backscattering data and errors in radiometric measurements for example, performance of cosine collectors, accuracy of radiometric calibrations. Unfortunately none of these are currently well characterized and the obvious disagreement between measured and modeled relationships is both disappointing and concerning. However, the retrievals of the main optical compounds from in situ radiometry are at least as good as the retrievals from optical field data (see next section) which suggests that the underpinning polynomial relationship between $r_{rs}(\lambda, 0^-)$ and $b_b(\lambda)/a(\lambda)$ is performing well and is unlikely to be a major source of error.

4.4. Performance of SDM for b_b/a From in Situ Radiometry

The SDMs were applied to estimates of $b_b(\lambda)/a(\lambda)$ derived from in situ measurements of remote sensing reflectance below the water, $r_{rs}(\lambda, 0^-)$, for both onshore and offshore stations. The Case 2 SDM was applied to the full data set and then the Case 1 SDM was applied to offshore stations only. The results (Figure 10) show that the MAE (=0.31 mg m^{-3}) for CHL in onshore waters is smaller compared to the analogous case obtained from in situ IOP data. In offshore waters, the Case 1 SDM improves the results for CHL retrievals compared to the Case 2 SDM, but the results for CDOM deteriorate (Table 1). The Case 2 SDM returns smaller MSS values for offshore waters using in situ radiometry inputs than it did using in situ IOPs, suggesting that (a) there are indeed unresolved issues with the in situ IOP data, and (b) application of the Case 2 SDM in open Case 1 waters is a reasonable prospect with no need for active partitioning. In onshore waters, the MAE values for all constituents are smaller or equivalent to the values obtained using in situ IOPs as input. The error bars obtained by using radiometric data for input encompass the 1:1 line, but are generally smaller than those found using in situ IOPs as input. Overall, it is remarkable that the performance of the SDM appears, if anything, to be superior when used with in situ radiometry based inputs than with in situ IOPs. Together with the deviation between measured and derived IOPs observed in Figure 9, it is tempting to suggest that the balance of error remains with in situ IOP measurements.

4.5. Comparison With Other Algorithms for the Ligurian Sea

The Case 2 SDM is unusual in that it can be used to simultaneously predict CHL, CDOM, and MSS. There are, of course, many other possible algorithms to obtain these constituent concentrations from radiometry individually. Here we compare the performance of the Case 2 SDM algorithm with a variety of previously published algorithms focusing on other SAAs, standard algorithms (for CHL) and where possible, algorithms that have been specifically tuned for Mediterranean waters.

4.5.1. Ligurian Sea CHL

The quasi-analytical algorithm (QAA) developed by Lee et al. (2002) has been widely used during the last 2 decades in ocean-color inversions (IOCCG report 2006) for CHL retrievals. Subsequent updated versions of the original model have been published and evaluated (Brewin et al., 2015; Lee, 2014; Lee et al., 2005, 2007, 2009; Mishra et al., 2014). It is a multi-band algorithm in 10 steps which mixes a set of well-known empirical, semi-analytical, and analytical models in order to retrieve several IOPs of the water body from (below) above-surface remote sensing reflectance, (r_{rs}) R_{rs}. It retrieves total absorption, $a_t(\lambda)$, first and then

Figure 8. Comparison of estimated versus measured (a) CHL, (b) CDOM, and (c) MSS in onshore (black squares) and offshore (black triangles) stations using the Case 2 SDM operating on in situ IOP (AC-9 and BB-9) measurements. Error bars are 95% CI of the bootstrap distributions. CHL and CDOM retrieved in offshore stations using the Case 1 SDM are shown as gray triangles. 1:1 lines are shown as solid black lines in each plot. Data shown for 23 onshore and 10 offshore stations.

Table 1
Mean Absolute Errors (MAE) for Retrievals of Optical Constituents Using In Situ IOPs and Radiometry, for Onshore and Offshore Waters, and Using Case 1 and Case 2 Versions of the SDM

	MAE					
	Result with in situ IOPs ($a(\lambda)$ and $b_b(\lambda)$)			Result with in situ radiometric data ($r_{rs}(\lambda)$)		
	Case 2 SDM		Case 1 SDM	Case 2 SDM		Case 1 SDM
	Onshore	Offshore	Offshore	Onshore	Offshore	Offshore
CHL (mg m^{-3})	0.60	0.54	0.40	0.31	0.73	0.51
CDOM (m^{-1})	0.012	0.013	0.010	0.011	0.004	0.021
MSS (g m^{-3})	0.18	–	–	0.18	–	–

decomposes it into individual absorption components, including the absorption by phytoplankton, $a_{ph}(\lambda)$. An up to date version (QAA_v6) is available on-line (https://www.ioccg.org/groups/Software_OCA/QAA_v6_2014209.pdf). The CHL concentration is not a direct output of the model, therefore, it has been estimated as a function of $a_{ph}(443)$, using the power law relationship proposed by Bricaud et al. (1995) and widely used for CHL retrievals from semi-analytical models (e.g., Brewin et al., 2015):

$$CHL = \left(\frac{a_{ph}(443)}{A} \right)^{1/B} \tag{12}$$

with $A = 0.0497$, $B = 0.7575$ and assuming $a_{ph}(443) \approx a_{ph}(440)$.

CHL concentration can also be estimated using empirical maximum band ratio (MBR) algorithms. These empirical algorithms are developed using a global data set (e.g., NOMAD) of in situ CHL concentrations, representing a wide variety of oceanic and coastal waters. These models are polynomial relationships that relate the log-transformed ratio of blue-green remote sensing reflectances to CHL using 2 to 6 bands (O'Reilly &Wertdell, 2019; O'Reilly et al., 1998, 2000). Details of the methodology can be found at https://oceancolor.gsfc.nasa.gov/atbd/chlor_a/. However, in optically complex waters biases between calculated and in situ concentrations are frequent. This is the case for parts of the Mediterranean Sea and as a result a number of regional variants of standard algorithms have been developed.

After converting in situ r_{rs} into R_{rs} using a constant transmission factor ($T = 0.54$) following Lee et al. (1998), standard and regional OCx algorithms were applied to the LS data set, using the available blue and green wavebands (440, 488, 510, and 555 nm). Here it assumed that differences caused by using for example, 440 nm instead of 443 nm have minimal impact on algorithm performance. Algorithms used here are: OC3M (https://oceancolor.gsfc.nasa.gov/atbd/chlor_a/), the standard CHL algorithm for MODIS, calibrated on the NOMAD global data set, with coefficients a0 = 0.2424, a1 = −2.7423, a2 = 1.8017, a3 = 0.0015, a4 = −1.2280; OC4S (https://oceancolor.gsfc.nasa.gov/atbd/chlor_a/), the standard CHL algorithm for SeaWiFS, with coefficients a0 = 0.3272, a1 = −2.9940, a2 = 2.7218, a3 = −1.2259, a4 = −0.5683; MedOC3, developed by Santoleri et al. (2008) and calibrated on a representative bio-optical data set collected in the Mediterranean area. Coefficients are: a0 = 0.380; a1 = −3.688; a2 = 1.036; a3 = 1.616; a4 = −1.328; and AD4, originally developed for the Adriatic Sea (Berthon & Zibordi, 2004; D'Alimonte & Zibordi, 2003), with coefficients: a0 = 0.236, a1 = −3.331, a2 = 2.286, a3 = 4.2834, a4 = −5.816.

Figure 11 shows results for CHL retrieval for the Ligurian Sea data set using each of the CHL algorithms listed above applied to in situ radiometry, with data partitioned further into onshore and offshore stations. The performance of the Case 2 SDM is generally at least comparable with other established
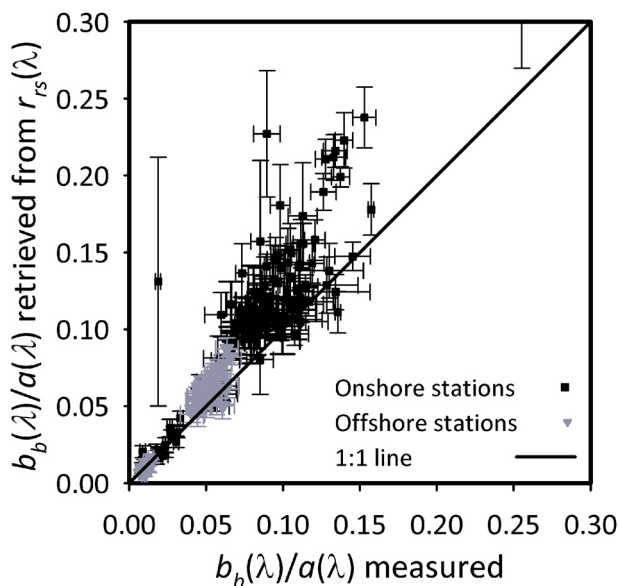

Figure 9. Retrieved versus measured data in onshore (black squares) and offshore (gray triangles) stations, at 5 wavebands (412–440–488–555–650 nm). Horizontal error bars are observed in situ IOP measurement uncertainties. Vertical error bars are the result of propagating associated $r_{rs}(\lambda)$ measurement uncertainties through a 5th order polynomial IOP-reflectance relationship (see text) and represent the 95%CI of the bootstrap distribution.
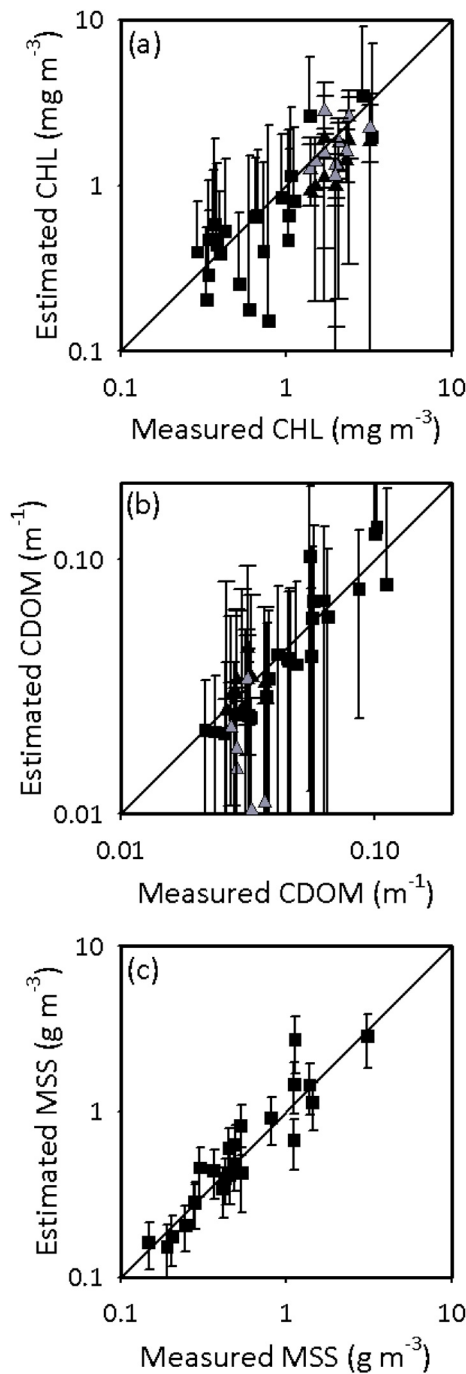
Figure 10. Comparison of estimated versus measured (a) CHL, (b) CDOM, and (c) MSS in onshore (black squares) and offshore (black triangles) stations, using the SDM operating on $b_b(\lambda)/a(\lambda)$ derived from radiometric data. Error bars are 95% CI of the bootstrap distributions. CHL and CDOM retrieved in offshore stations using the Case 1 SDM are shown as gray triangles. 1:1 lines are shown as solid black lines in each plot. Data shown for 23 onshore and 10 offshore stations.

algorithms. The majority of Case 2 SDM CHL estimates fall within the ±35% mission target for ocean color CHL retrievals for this data set (Figure 11a). QAAv4 has a general tendency to overestimate CHL though performance is generally better offshore (Figure 11b). Conversely, QAAv6 tends to underestimate CHL with the majority of points lying close to the lower threshold of the mission target range (Figure 11b). OC3M and OC4S both perform well in offshore waters, but have a clear tendency to overestimate CHL for onshore stations (Figure 11c). MEDOC3 and AD4 also tend to overestimate CHL in onshore waters and have slightly more varied performance than OC3M and OC4S for offshore waters (Figure 11d). Case2 SDM performance is most similar to QAAv6 and it is noticeable that most of the other algorithms return overestimates in a number of cases, possibly reflecting residual issues in dealing with the optical complexity of the LS data set.

4.5.2. Ligurian Sea CDOM

Absorption by CDOM is strongest in the UV-blue and decreases approximately exponentially into the red-NIR (Schwarz et al., 2002). In Case 1 waters, CDOM originates mainly from phytoplankton while in coastal waters CDOM can be highly variable in composition and concentration and can absorb up to 90% of sunlight between 400 and 500 nm (Bélanger et al., 2008). Absorption by CDOM can be derived from reflectances using both SAAs and empirical band ratio algorithms. In Figure 12 we compare the performance of the Case 2 SDM CDOM product with a selection of previously published CDOM algorithms including: the 2 versions of the QAA algorithm (QAAv4, QAAv6) discussed previously (Lee et al., 2002) and a third version (QAAZY) optimized for CDOM retrieval in turbid estuarine and coastal waters (Zhu & Yu, 2013); two band ratio algorithms (M2008, M2014) based on observations from US coastal waters (Mannino et al., 2008, 2014); and a further two band ratio algorithms proposed by Berthon et al. (2006) for the Adriatic Sea region (B_AD) and European coastal regions (B_EC) for CDOM retrieval. NB. The Berthon et al. algorithms recover absorption by CDOM at 400 nm while the rest of our CDOM values are reported at 440 nm. We have applied a best-fit conversion factor of 2.5 to the B_AD and B_EC data to establish comparability with other algorithms.

All bar one of the Case 2 SDM data points falls within the ±35% mission target boundary indicating robust performance for the data set from which the model was developed. QAAv4 performs well for offshore waters but is less reliable onshore, while the opposite is true for QAAv6. QAAZY appears to overestimate CDOM in all cases tested here. The Mannino et al. band ratio algorithms typically overestimate CDOM for these waters but the general approach appears to be reasonably robust as demonstrated by the strong performance of the two Berthon et al. algorithms. It seems possible that the differences in performance between these two sets of band ratio algorithms might be associated with regional differences in the data sets used to generate them. Overall it would appear that retrieval of CDOM with the Case 2 SDM is at least on a par with previously published algorithms.

4.5.3. Ligurian Sea MSS

Mineral particles are well known to be have potential to influence remote sensing reflectance signals in the red-NIR through relatively efficient generation of optical backscattering signals (e.g., Doxaran et al., 2002). Neil et al. (2011) derived polynomial expressions for computation of the maximum and minimum values of suspended mineral particles of terrigenous origin in the presence of independently varying concentrations of CHL and CDOM, based on R_{rs}667. Figure 13 shows these two relationships applied to onshore Ligurian Sea in situ R_{rs}(663)
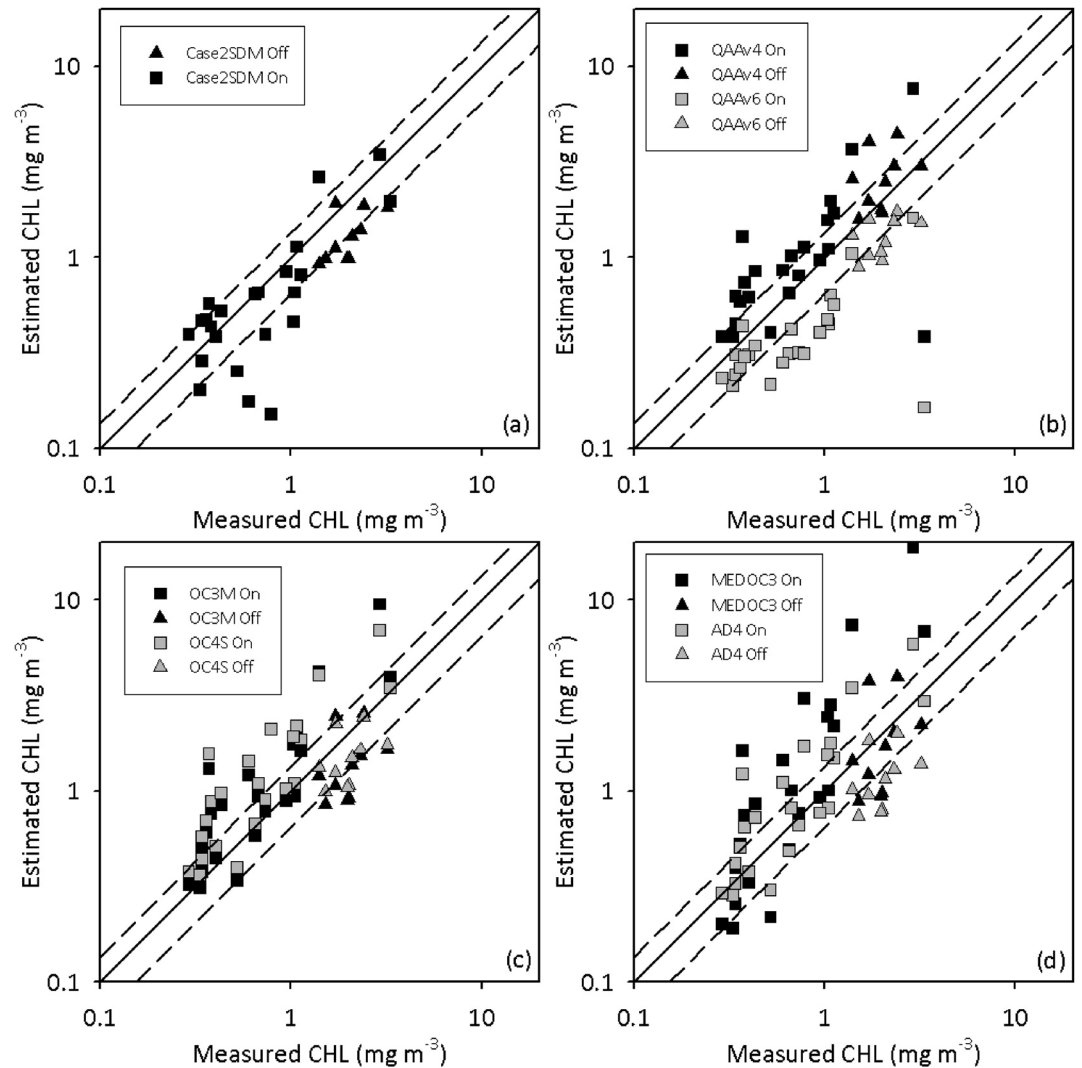
Figure 11. Comparison of estimated versus measured CHL for the Ligurian Sea data set using (a) the Case 2 SDM applied to in situ radiometry, (b) two versions of the QAA algorithm, (c) two standard MBR CHL algorithms, and (d) two regionally tuned versions of standard MBR algorithms. 1:1 lines are shown as solid black lines and ±35% mission targets are shown as dashed lines in each plot. Squares and triangles represent onshore and offshore stations respectively. Data shown are for 23 onshore and 10 offshore stations.

data (in lieu of $R_{rs}(667)$) along with estimates of MSS from the Case 2 SDM. It is clear that the Case 2 SDM estimates of MSS are generally consistent with previous predictions of the relationship between MSS and red reflectance given by Neil et al., with the SDM model directly accounting for the influence of other materials on red reflectance rather than leaving this factor as an unknown.

4.6. Application to a Global Data Set

In the previous section the Case 2 SDM was found to perform at least as well as other existing algorithms for CHL, CDOM, and MSS, with the SDM approach having the additional merit of directly producing estimates for all three constituent simultaneously and, as shown earlier, with associated uncertainties. However it would be reasonable to expect that the Case 2 SDM approach would work particularly well for the data set from which it was derived. In this section we compare the performance of the SDM approach with other CHL algorithms using a global data set. The intention is to establish a more robust understanding of general algorithm performance.

Data were selected from the NOMAD data set, which is a community generated set of in situ bio-optical data compiled by the NASA Ocean Biology Processing Group at Goddard Space Flight Center, Maryland, USA
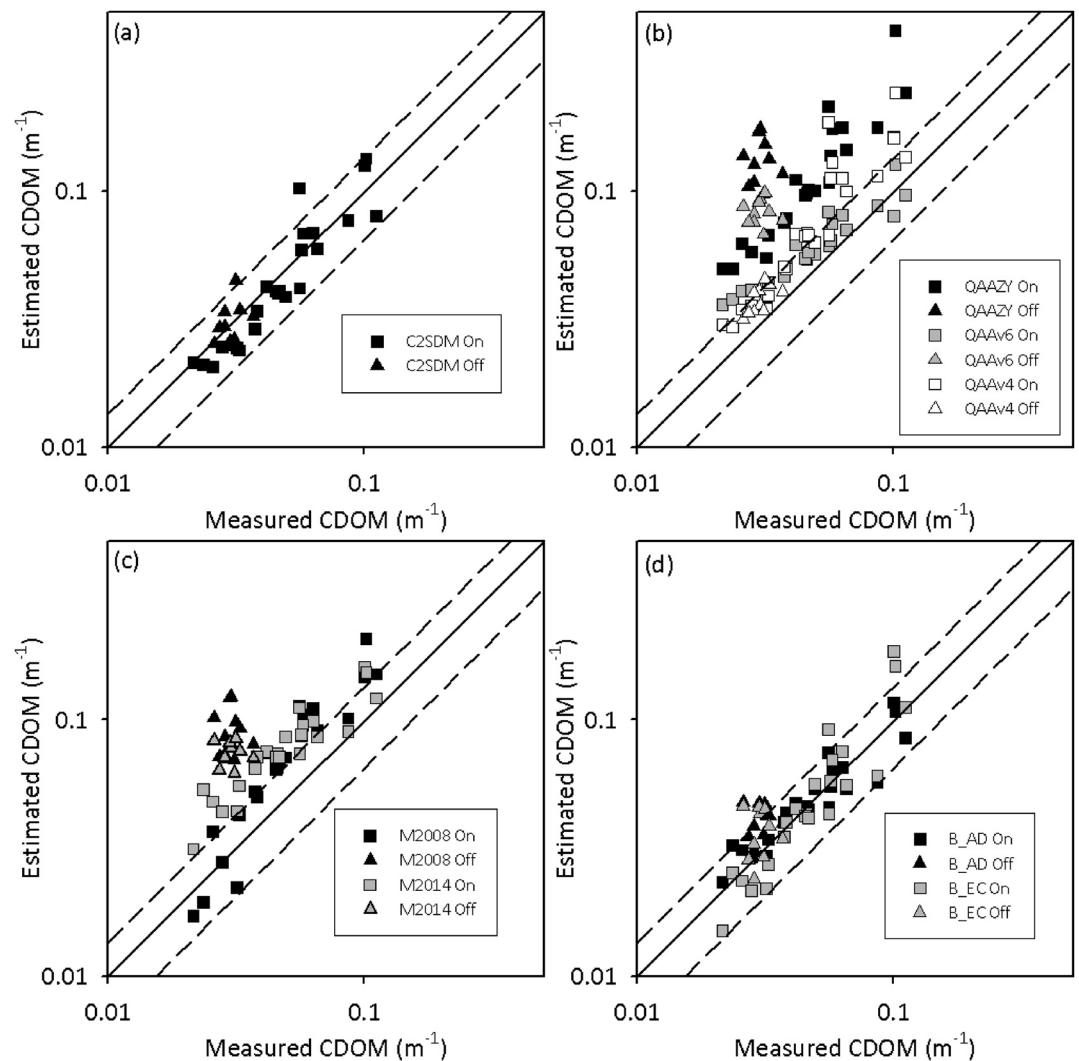
Figure 12. Comparison of estimated versus measured CDOM for the Ligurian Sea data set using (a) the Case 2 SDM applied to in situ radiometry, (b) three versions of the QAA algorithm, (c) two 2-wavelength band ratio CDOM algorithms presented by Manino and co-workers based on observations from the USA, and (d) two 2-wavelength band ratio CDOM algorithms presented Berthon and co-workers based on observations from the Adriatic Sea and from other European coastal waters. 1:1 lines are shown as solid black lines and $\pm 35\%$ mission targets are shown as dashed lines in each plot. Squares and triangles represent 23 onshore and 10 offshore stations respectively.

(Werdell & Bailey, 2005). This provided ~500 stations with CHL data available between 0 and 10 m and remote sensing reflectance data available at 4 wavebands (412, 440, 488, and 555 nm). Note the 650 nm band that was previously used in the Case 2 SDM was not available from this data set, so in what follows the Case 2 SDM operates on the 4 wavebands identified in this paragraph only. Figures 14a–14d shows estimation of CHL for the 4 waveband version of the SDM approach along with estimates from QAAv4, OC3M, and OC4S. The Case 2 SDM and QAAv4 perform very similarly, with both having a tendency to underestimate CHL at high concentrations and to exhibit more spread than is observed with the two band ratio algorithms. This is consistent with the findings reported by Brewin et al. (2015) who noted that performance of SAA algorithms was generally not as strong as for MBR algorithms. That said, in all cases there are large numbers of CHL estimates that fall well outside the $\pm 35\%$ mission target lines indicating that general performance levels are not ideal. Figure 14 (e and f) shows distributions of CDOM and MSS estimates from the Case 2 SDM. Note the log-spaced histogram bins. These results suggest that there is a great deal of optical complexity within this subset of the NOMAD data set and that the Case 2 SDM approach returns low values of CDOM and MSS which is expected as much of the data set is from relatively clear offshore waters. Overall it would appear that the Case 2 SDM performs similarly to other
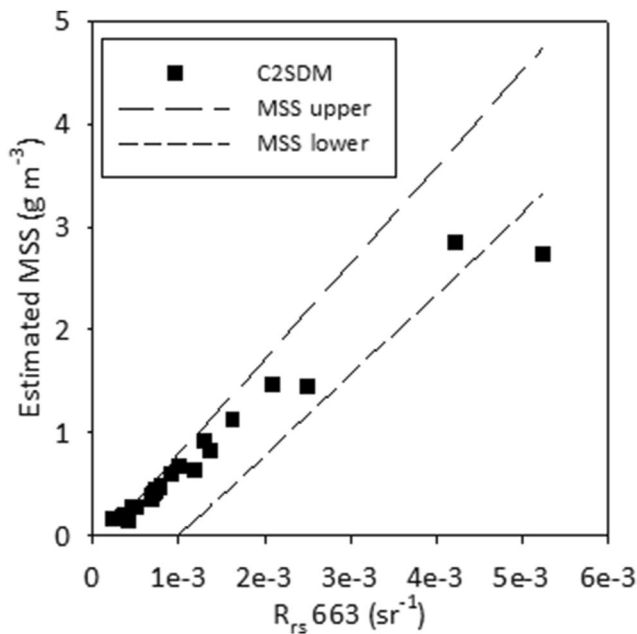
Figure 13. Dashed lines show predicted upper and lower limits for the relationship between MSS and red reflectance taking into account the influence of CHL and CDOM. All bar one of the Case 2 SDM MSS predictions fall within these boundaries.

established algorithms for CHL and has the additional merit of providing plausible estimates of CDOM and MSS for regions well outside the original training data set.

5. Discussion and Conclusions

5.1. SDM Construction Considerations

The spectral deconvolution model presented here is a modified version of the model presented by Ramirez-Perez et al. (2018) with the fundamental change being transition from use of scattering coefficients to backscattering coefficients. This is an important modification that facilitates transfer of the SDM from in situ AC-9/S IOP measurements toward remote sensing applications where the relationship between remote sensing reflectance and the ratio of backscattering to absorption is particularly relevant. Recent results presented by Lo Prejato et al. (2020) reaffirm the potential to easily relate remote sensing reflectance to b_b/a in both clear oceanic and optically complex coastal waters. Thus this version of the SDM has been formulated to operate on the ratio of b_b/a rather than b_b and a separately as the ratio is easily derived from remote sensing reflectance whereas a and b_b would be much more difficult to determine separately. There is therefore a clear path to use this SDM with either direct in situ measurements of absorption and backscattering or from either in situ or remotely sensed values of R_{rs}. Here we have chosen to focus on in situ primary data sources as a stepping stone to allow us to fully understand practical performance characteristics of the SDM before tackling the more complicated case of remote sensed data inputs.

The approach we have developed here shares a heritage with a number of previous studies in this area, notably models presented by Hoge and Lyon (1996) and Wang et al. (2005). Like Hoge and Lyon, our model uses a form of linear matrix inversion to solve a system of linear equations. Where Hoge and Lyon solved for three parameters using three wavebands, we have extended the analysis to include over-constrained systems, where the expectation is that inclusion of additional wavebands provides more spectral information to improve retrieval of OSC concentrations. We have shown that this is indeed the case, but that the improvement is not without limitations. Another novel feature of our approach comes from having access to an unusually rich set of SIOPs that allows us to fully populate a simple bio-optical model (Equation 6). This was made possible by the particular distribution of stations occupied during the BP09 cruise and by the development of a careful partitioning approach outlined in Ramirez-Perez et al. (2018) and extended in Lo Prejato et al. (2020). As a result, our bio-optical model is fully based on observed SIOPs, does not rely on assumed models for spectral backscattering and does not bundle detrital particle absorption with CDOM absorption for example, Wang et al. (2005) and many others. The SIOPs used in our SDM have the further potential advantages of (a) being based on highly accurate PSICAM measurements for some of the absorption coefficients, and (b) having been generated by the linear regression technique which reduces sensitivity to the impact of measurement uncertainties (McKee et al., 2014). Estimation of propagated measurement uncertainties for SIOPs further enables estimation of uncertainties for the final products of our SDM. The bootstrapping approach taken in this paper is simple to implement and allows us to combine uncertainties on all inputs. This SDM has been developed with this aspect in mind from the outset and we believe that the general simplicity of the SDM construction through Equation 6 makes this particularly tractable.

5.2. Optimisation of the SDM

As was the case with the previous version presented by Ramirez-Perez et al. (2018), the new version of the SDM works perfectly if no errors are introduced into the data inputs. This is true for any number of wavebands from the minimum, equal to the number of OSCs to be retrieved, for example, Hoge and Lyon (1996), up to the maximum of 8 wavebands available for this study. However, incorporating realistic errors into both IOPs and SIOPs inevitably introduces errors in retrieved OSCs. Unsurprisingly OSC errors are most significant when the contribution of the OSC in question to total IOPs is small compared to other constituents. This is an important point: given the
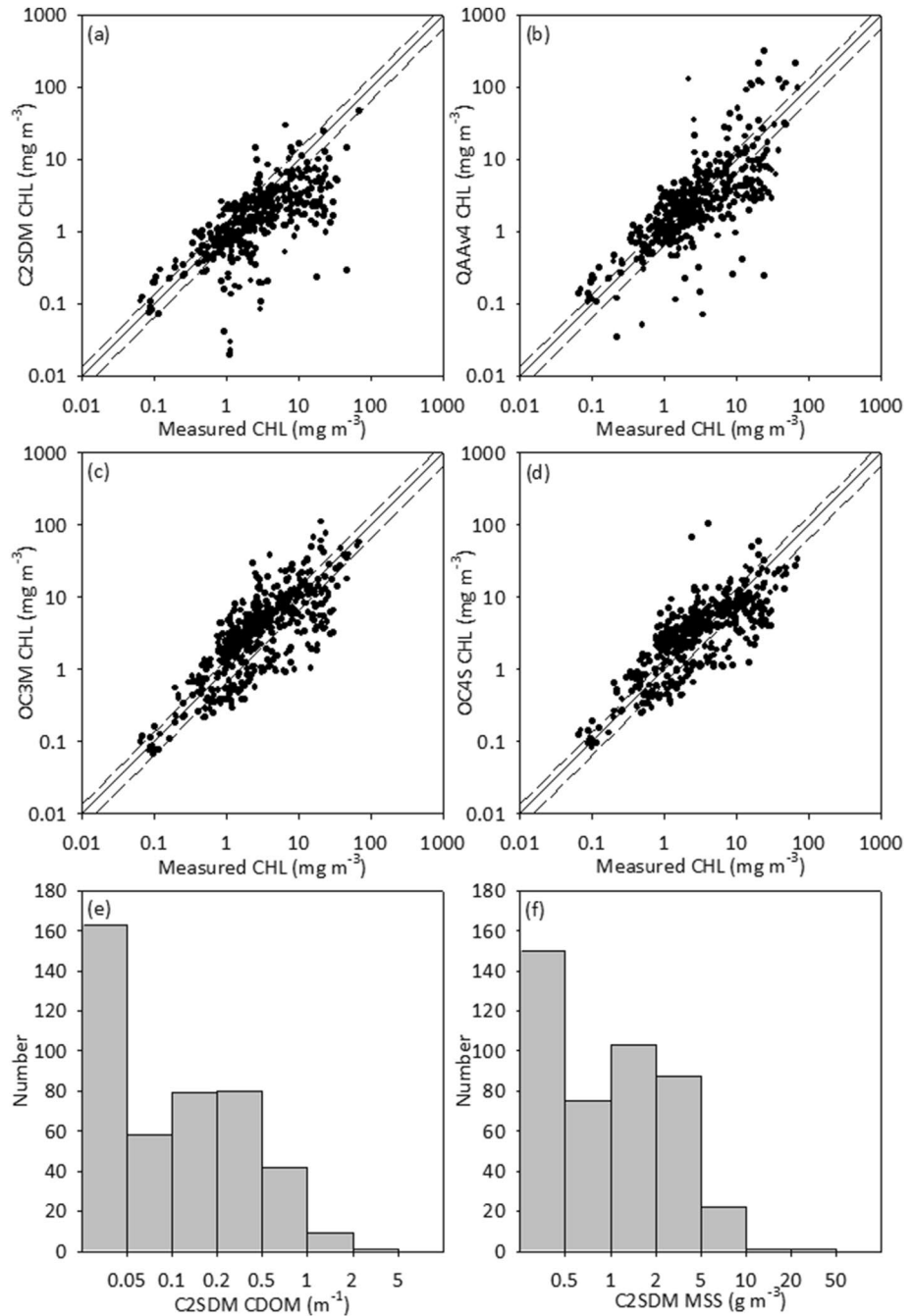
Figure 14. Comparison of (a) Case 2 SDM CHL performance against (b) QAAv4, (c) OC3M and (d) OC4S for the subset of the NOMAD data set providing R_{rs} data compatible with the Case 2 SDM algorithm requirements. Solid lines show 1:1 and dashed lines represent $\pm 35\%$ mission targets. Distributions of Case 2 SDM (e) CDOM and (f) MSS concentrations for the NOMAD data set indicate that there is a wide range of optical complexity in this data set.

broad spectral features associated with marine OSCs, there is no reason to expect to be able to accurately retrieve small OSC concentrations in the presence of overwhelming concentrations of other components. On the other hand, it is absolutely essential that OSC concentrations are provided to end users with reasonable estimates of uncertainties and, preferably, with information about the concentration of other components to inform judgment on likely data quality. The SDM approach presented here achieves both of these goals. In a similar manner, it is important to understand the impact of measurement uncertainties on algorithm performance. The synthetic data set results illustrate that measurement uncertainties may play as important a role in SDM performance breakdown

as variability in SIOPs. Whilst many existing algorithms do breakdown in Case 2 waters due to the effects of optical complexity, to a large extent the SDM approach is robust. It does however remain sensitive to uncertainties in inputs (IOPs or radiometry), natural variability in SIOPs and the presence of additional OSCs with distinctive properties for example, coccolithophores. Importantly, the SDM could be adapted in future to accommodate additional OSCs if suitable libraries of SIOPs can be established and classification schemes developed to identify these occurrences.

The SDM is designed to exploit spectral information content provided by either in situ IOPs or remote sensing reflectance data. A naïve approach would suggest that including more spectral information would presumably improve retrieved data quality. However, analysis of SDM performance showed that simply increasing the number of wavebands does not lead to ever greater accuracy in OSC retrieval, with the spread error (%CI) plateauing for 5 or more wavebands and the bias error (%Bias) appearing to increase at least up to the maximum of 8 wavebands available here. This result is consistent with others in the literature (e.g., Sathyendranath et al., 1994; Wernand et al., 1997; Lee & Carder, 2002) and suggests that hyperspectral data may not be necessary or even beneficial for retrieving this level of OSC information. Indeed, these results raise the intriguing possibility that relatively small numbers of wavebands could be effective for retrieving CHL, MSS, and CDOM, a result that is potentially useful for designing sensors for very small remote sensing platforms for example, remotely piloted aircraft and cubesats. However, that said, there may be potential to expand the SDM to include additional components for example, specific algal groups with characteristic spectral features such as cyanobacteria. In this case, access to additional wavebands from hyperspectral sensors would almost certainly prove useful (see e.g., Giardino et al., 2019).

A key benefit of the SDM is ability to easily propagate errors in order to provide OSC concentrations with uncertainties. Computational efficiency of the SDM is largely driven by selection of an appropriate number of iterations for the bootstrap process and it has been shown that this also impacts on the production of stable estimates of CI% and Bias% values. Optimal OSC retrieval was reached for runs using 500 iterations, but the number of iterations could be reduced with only minimal impact on CI% and Bias%. This is not problematic when applying the SDM to relatively small numbers of in situ data points, but could easily become a major consideration in the case of applying the SDM to remote sensing imagery. Being able to reduce the number of iterations to much lower values without significantly reducing data integrity is therefore an important feature for future applications.

5.3. Performance of the SDM

The SDM has been tested using both offshore (Case 1) and onshore (Case 2) data sets from the Ligurian Sea, and using both in situ IOP and in situ radiometric data. As Table 1 illustrates, the MAEs for each OSC are broadly comparable under each of these variations. Ignoring the use of the Case 2 SDM in Case 1 waters, the MAE for CHL is ≤ 0.6 mg m^{-3}, the MAE for CDOM is ≤ 0.021 m^{-1} and the MAE for MSS is 0.18 g m^{-3}. How useful these levels of accuracy are depends on the application. In the first instance it is clearly a benefit to be able to retrieve all three parameters simultaneously. Knowledge of the three OSCs, and knowing that performance is limited when the contribution of an OSC is low compared to others, helps to inform judgment of data quality. Second, provision of an estimate of uncertainty for each retrieved OSC concentration is a further aid to establishing data quality. Note that the error bars shown in Figures 8 and 10 represent 95% CIs (rather than standard deviations–therefore ∼2x greater), and are presented to highlight that there is in fact considerable uncertainty beyond that expressed by the MAE values. In most (but not all) cases, the 95% CI encompasses the 1:1 line. In all cases there is no sign of major systematic deviations from the 1:1 line. Taking into consideration the turbid nature of the onshore stations, this is potentially a major achievement when considered against the performance of for example, standard blue-green reflectance ratio algorithms in turbid waters (e.g., McKee et al., 2007). It is likely that the level of performance presented here is of more value for Case 2 applications than for Case 1 waters, but the ability to provide a consistent data product across both water types may be useful as well. The fact that performance is similar using both in situ IOPs and in situ radiometry, despite the discrepancy between the two noted in Figure 9, is a very welcome sign that the approach has potential to be transferred to remote sensing applications. As demonstrated clearly here, however, the quality of input data is crucial in determining the quality of retrieved OSCs.

Comparison of the SDM performance with other single parameter and SAA algorithms was generally favorable, with the SDM producing OSC estimates that were at least of similar quality for the Ligurian Sea data set from which the SIOPs used to populate the SDM were derived. Similar results were obtained for the NOMAD global data set, with the SDM performing consistently with the QAA algorithm and only a little worse that the OCx

algorithms. In all cases, the delivery of all three OSC concentrations and provision of associated uncertainties suggests that the SDM approach has general merit, particularly for optically complex coastal waters.

5.4. Future Perspectives

The SDM approach presented here has demonstrated that it is possible to retrieve information on major OSCs using a relatively small number of wavebands across a wide range of water types and to obtain estimates of OSC uncertainties in the process. The next stage of development will be to apply the technique to satellite-derived remote sensing reflectance data. Looking further into the future, a major attraction of the SDM approach is that it is based on a logical, easily understood bio-optical model that can potentially be adapted as circumstances require/permit. The SDM approach has previously been shown to have potential to be incorporated into an adaptive scheme to address features such as the package effect for chlorophyll-specific absorption and natural variability in other SIOPs (Brando et al., 2012). It may also be possible to add additional components (e.g., cyanobacteria) or adapt it toward a phytoplankton functional type algorithm where the limiting factors are being able to provide a suitable suite of SIOPs and sufficient spectral input data to the SDM. Establishing libraries of SIOP spectra for different water types/OSCs and potentially further developing the solving approach to support analysis of more complex mixtures of OSCs (e.g., Zhang et al., 2011) are potentially fruitful lines of research to build on this work.

Conflict of Interest

The authors declare no conflicts of interest relevant to this study.

Data Availability Statement

All data underpinning this publication are openly available from the University of Strathclyde KnowledgeBase at https://doi.org/10.15129/3424d06f-053c-43b7-aed0-532ffeebe6b4. The authors would like to thank three anonymous reviewers for constructive comments that greatly improved the quality of the paper.

References

Bélanger, S., Babin, M., & Larouche, P. (2008). An empirical ocean color algorithm for estimating the contribution of chromophoric dissolved organic matter to total light absorption in optically complex waters. *Journal of Geophysical Research*, *113*(C4), C04027. https://doi.org/10.1029/2007JC004436

Bengil, F., McKee, D., Besiktepe, S. T., Sanjuan Calzado, V., & Trees, C. (2016). A bio-optical model for integration into ecosystem models for the Ligurian Sea. *Progress in Oceanography*, *149*, 1–15. https://doi.org/10.1016/j.pocean.2016.10.007

Berthon, J.-F., & Zibordi, G. (2004). Bio-optical relationships for the northern Adriatic Sea. *International Journal of Remote Sensing*, *25*(7–8), 1527–1532. https://doi.org/10.1080/01431160310001592544

Berthon, J.-F., Zibordi, G., Vandepitte, D., Canuti, E., & Eker-Develi, E. (2006). *Regional bio-optical relationships and algorithms for the Adriatic Sea, the Baltic Sea and the English channel/North Sea suitable for ocean colour sensors*. European Commission, Directorate-General Joint Research Centre.

Bosc, E., Bricaud, A., & Antoine, D. (2004). Seasonal and interannual variability in algal biomass and primary production in the Mediterranean Sea, as derived from 4 years of SeaWiFS observations. *Global Biogeochemical Cycles*, *18*, 1–17. https://doi.org/10.1029/2003GB002034

Boss, E., & Pegau, S. (2001). The relationship of scattering in an angle in the back direction to the backscattering coefficient. *Applied Optics*, *40*(30), 5503–5507. https://doi.org/10.1364/AO.40.005503

Brando, V. E., Dekker, A. G., Park, Y. G., & Schroeder, T. (2012). Adaptive semianalytical inversion of ocean color radiometry in optically complex waters. *Applied Optics*, *51*(15), 2808–2833. https://doi.org/10.1364/AO.51.002808

Brewin, R. J. W., Sathyendranath, S., Müller, D., Brockmann, C., Deschamps, P.-Y., Devred, E., et al. (2015). The Ocean colour climate change initiative: III. A round-robin comparison on in-water bio-optical algorithms. *Remote Sensing of Environment*, *162*, 271–294. https://doi.org/10.1016/j.rse.2013.09.016

Bricaud, A., Babin, M., Morel, A., & Claustre, H. (1995). Variability in the chlorophyll-specific absorption coefficient of natural phytoplankton: Analysis and parameterization. *Journal of Geophysical Research*, *100*(C7), 13321–13332. https://doi.org/10.1029/95jc00463

Cherukuru, N., Brando, V. E., Schroeder, T., Clementson, L. A., & Dekker, A. G. (2014). Influence of river discharge and ocean currents on coastal optical properties. *Continental Shelf Research*, *84*, 188–203. https://doi.org/10.1016/j.csr.2014.04.022

D'Alimonte, D., & Zibordi, G. (2003). Phytoplankton determination in an optically complex coastal region using a multilayer perceptron neural network. *IEEE Transactions on Geoscience and Remote Sensing*, *41*(12), 2861–2868. https://doi.org/10.1109/TGRS.2003.817682

Davison, A., & Hinkley, D. (1997). *Bootstrap methods and their application*. Cambridge University Press.

Devred, E., Turpie, K., Moses, W., Klemas, V., Moisan, T., Babin, M., et al. (2013). Future retrievals of water column bio-optical properties using the hyperspectral infrared imager (HyspIRI). *Remote Sensing*, *5*(12), 6812–6837. https://doi.org/10.3390/rs5126812

Doerffer, R., & Schiller, H. (1994). Inverse modeling for retrieval of ocean color parameters in Case II coastal waters: An analysis of the minimum error. In *Proc. SPIE 2258, Ocean Optics XII* (p. 26). https://doi.org/10.1117/12.190042

Doxaran, D., Froidefond, J.-M., Lavender, S., & Castaing, P. (2002). Spectral signature of highly turbid waters. Application with SPOT data to quantify suspended particulate matter. *Remote Sensing of Environment*, *81*(1), 149–161. https://doi.org/10.1016/S0034-4257(01)00341-8

Fan, C., & Warner, R. A. (2014). Characterization of water reflectance spectra variability: Implications for hyperspectral remote sensing in estuarine waters. *Marine Science*, *4*(1), 1–9. https://doi.org/10.5923/j.ms.20140401.01

Giardino, C., Brando, V. E., Gege, P., Pinnel, N., Hochberg, E., Knaeps, E., et al. (2019). Imaging spectrometry of inland and coastal waters: State of the art, achievements and perspectives. *Surveys in Geophysics*, *40*(3), 401–429. https://doi.org/10.1007/s10712-018-9476-0

Gitelson, A. A., Dall'Olmo, G., Moses, W., Rundquist, D. C., Barrow, T., Fisher, T. R., et al. (2008). A simple semi-analytical model for remote estimation of chlorophyll-a in turbid waters: Validation. *Remote Sensing of Environment*, *112*(9), 3582–3593. https://doi.org/10.1016/j.rse.2008.04.015

Gordon, H. R., Brown, O. B., Evans, R. H., Brown, J. W., Smith, R. C., Baker, K. S., & Clark, D. K. (1988). A semianalytic radiance model of ocean color. *Journal of Geophysical Research*, *93*(D9), 10909–10924. https://doi.org/10.1029/JD093iD09p10909

Gordon, H. R., & Morel, A. Y. (1983). *Remote assessment of ocean color for interpretation of satellite visible imagery: A review* (Vol. 114). Springer-Verlag.

Hoge, F. E., & Lyon, P. E. (1996). Satellite retrieval of inherent optical properties by linear matrix inversion of oceanic radiance models: An analysis of model and radiance measurement errors. *Journal of Geophysical Research*, *100*(C7), 16631–16648. https://doi.org/10.1029/96JC01414

Hooker, S. B., Esaias, W. E., Feldman, G. C., Gregg, W. W., & McClain, C. R. (1992). (Eds.), In S. B. Hooker, & E. R. Firestone (Eds.), *An overview of SeaWiFS and ocean-color. NASA Tech. Memo. 104566* (Vol. 1, pp. 24). NASA Goddard Space Flight Center.

IOCCG. (2000). Remote sensing of ocean colour in coastal, and other optically-complex waters. In S. Sathyendranath (Ed.), *Reports of the international ocean-colour coordinating group. no 3*. IOCCG. https://doi.org/10.25607/OBP-95

IOCCG. (2006). Remote sensing of inherent optical properties: Fundamentals, tests of algorithms, and applications. In Z.-P. Lee (Ed.), *Reports of the international ocean-colour coordinating group, no. 5*. IOCCG. https://doi.org/10.25607/OBP-96

IOCCG. (2019). Uncertainties in ocean colour remote sensing. In F. Mélin (Ed.), *Reports of the international ocean-colour coordinating group, No. 18*. IOCCG. https://doi.org/10.25607/OBP696

IOCCG Protocol Series. (2018). Inherent optical property measurements and protocols: Absorption coefficient. In A. R. Neeley & A. Mannino (Eds.), *IOCCG ocean optics and biogeochemistry protocols for satellite ocean colour sensor validation* (Vol. 1.0). IOCCG. https://doi.org/10.25607/OBP-119

Kirk, J. (2011). *Light and photosynthesis in aquatic ecosystems*. Cambridge University Press.

Lee, Z., Arnone, R., Hu, C., Werdell, P. J., & Lubac, B. (2010). Uncertainties of optical parameters and their propagations in an analytical ocean color inversion algorithm. *Applied Optics*, *49*(3), 369–381. https://doi.org/10.1364/AO.49.000369

Lee, Z., Carder, K. L., & Arnone, R. A. (2002). Deriving inherent optical properties from water color: A multiband quasi-analytical algorithm for optically deep waters. *Applied Optics*, *41*(27), 5755–5772. https://doi.org/10.1364/AO.41.005755

Lee, Z., Lubac, B., Werdell, P. J., & Arnone, R. (2009). An update of the Quasi-analytical algorithm (QAA_v5). Technical report: International ocean colour coordinating group (IOCCG).

Lee, Z. P. (2014). *An update of the quasi-analytical algorithm (QAA_v6)*. IOCCG. Retrieved from http://www.ioccg.org/groups/Software_OCA/QAA_v6_2014209.pdf

Lee, Z. P., & Carder, K. L. (2002). Effect of spectral band numbers on the retrieval of water column and bottom properties from ocean color data. *Applied Optics*, *41*(12), 2191–2201. https://doi.org/10.1364/AO.41.002191

Lee, Z. P., Carder, K. L., Mobley, C. D., Steward, R. G., & Patch, J. S. (1998). Hyperspectral remote sensing for shallow waters. I. A semianalytical model. *Applied Optics*, *37*(27), 6329–6338. https://doi.org/10.1364/AO.37.006329

Lee, Z. P., Darecki, M., Carder, K. L., Davis, C., Stramski, D., & Rhea, W. J. (2005). Diffuse attenuation coefficient of downwelling irradiance: An evaluation of remote sensing methods. *Journal of Geophysical Research*, *110*(C2), C02017. https://doi.org/10.1029/2004JC002573

Lee, Z. P., Weidemann, A., Kindle, J., Arnone, R., Carder, K. L., & Davis, C. (2007). Euphotic zone depth: Its derivation and implication to ocean color remote sensing. *Journal of Geophysical Research*, *112*(C3), C03009. https://doi.org/10.1029/2006JC003802

Lefering, I., Bengil, F., Trees, C., Röttgers, R., Bowers, D., Nimmo-Smith, A., et al. (2016b). Optical closure in marine waters from in situ inherent optical property measurements. *Optics Express*, *24*(13), 14036–14052. https://doi.org/10.1364/OE.24.014036

Lefering, I., Röttgers, R., Utschig, C., & McKee, D. (2017). Uncertainty budgets for liquid waveguide CDOM absorption measurements. *Applied Optics*, *56*(22), 6357–6366. https://doi.org/10.1364/AO.56.006357

Lefering, I., Röttgers, R., Weeks, R., Connor, D., Utschig, C., Heymann, K., & McKee, D. (2016). Improved determination of particulate absorption from combined filter pad and PSICAM measurements. *Optics Express*, *24*(22), 24805–24823. https://doi.org/10.1364/OE.24.024805

Lo Prejato, M., McKee, D., & Mitchell, C. (2020). Inherent optical properties-reflectance relationships revisited. *Journal of Geophysical Research: Oceans*, *125*(11), e2020JC016661. https://doi.org/10.1029/2020JC016661

Mannino, A., Novak, M. G., Hooker, S. B., Hyde, K., & Aurin, D. (2014). Algorithm development and validation of CDOM properties for estuarine and continental shelf waters along the northeastern U.S. coast. *Remote Sensing of Environment*, *152*, 576–602. https://doi.org/10.1016/j.rse.2014.06.027

Mannino, A., Russ, M. E., & Hooker, S. B. (2008). Algorithm development and validation for satellite-derived distributions of DOC and CDOM in the U.S. Middle Atlantic Bight. *Journal of Geophysical Research*, *113*(C7), C07051. https://doi.org/10.1029/2007JC004493

Maritorena, S., Siegel, D. A., & Peterson, A. R. (2002). Optimization of a semianalytical ocean color model for global-scale applications. *Applied Optics*, *41*(15), 2705–2714. https://doi.org/10.1364/AO.41.002705

Matsushita, B., Yang, W., Chang, P., Yang, F., & Fukushima, T. (2012). A simple method for distinguishing global Case-1 and Case-2 waters using SeaWiFS measurements. *ISPRS Journal of Photogrammetry and Remote Sensing*, *69*, 74–87. https://doi.org/10.1016/j.isprsjprs.2012.02.008

McKee, D., Chami, M., Brown, I., Calzado, V. S., Doxaran, D., & Cunningham, A. (2009). Role of measurement uncertainties in observed variability in the spectral backscattering ratio. A case study in mineral-rich coastal waters. *Applied Optics*, *48*(24), 4663–4675. https://doi.org/10.1364/AO.48.004663

McKee, D., & Cunningham, A. (2006). Identification and characterisation of two optical water types in the Irish Sea from in situ inherent optical properties and seawater constituents. *Estuarine, Coastal and Shelf Science*, *68*(1–2), 305–316. https://doi.org/10.1016/j.ecss.2006.02.010

McKee, D., Cunningham, A., & Dudek, A. V. (2007). Optical water type discrimination and tuning remote sensing band-ratio algorithms: Application to retrieval of chlorophyll and Kd(490) in the Irish and Celtic Seas. *Estuarine, Coastal and Shelf Science*, *73*(3–4), 827–834. https://doi.org/10.1016/j.ecss.2007.03.028

McKee, D., Piskozub, J., Röttgers, R., & Reynolds, R. A. (2013). Evaluation and improvement of an iterative scattering correction scheme for in situ absorption and attenuation measurements. *Journal of Atmospheric and Oceanic Technology*, *30*(7), 1527–1541. https://doi.org/10.1175/JTECH-D-12-00150.1

McKee, D., Röttgers, R., Neukermans, G., Calzado, V. S., Trees, C., AmpoloRella, M., et al. (2014). Impact of measurement uncertainties on determination of chlorophyll-specific absorption coefficient for marine phytoplankton. *Journal of Geophysical Research: Oceans*, *119*(12), 9013–9025. https://doi.org/10.1002/2014JC009909

Mishra, S., Mishra, D. R., & Lee, Z. (2014). Bio-optical inversion in highly turbid and cyanobacteria-dominated waters. *Transactions on Geoscience and Remote Sensing*, *52*(1), 375–388. https://doi.org/10.1109/TGRS.2013.2240462

Mobley, C. D. (1994). *Light and water: Radiative transfer in natural waters*. Academic Press.

Mobley, C. D., Stramski, D., Bissett, W. P., & Boss, E. (2004). Optical modelling of ocean waters: Is the Case 1 - Case 2 classification still useful? *Oceanography*, *17*(2), 60–67. https://doi.org/10.5670/oceanog.2004.48

Mooney, C. Z., & Duval, R. D. (1993). *Bootstrapping: A nonparametric approach to statistical inference*. Sage.

Morel, A. (1974). Optical properties of pure water and pure seawater. In N. G. Jerlov & E. S. Nielsen (Eds.), *Optical aspects of oceanography*. Academic Press.

Morel, A. (1988). Optical modeling of the upper ocean in relation to its biogenous matter content (Case I waters). *Journal of Geophysical Research*, *93*(C9), 10749–10768. https://doi.org/10.1029/JC093iC09p10749

Morel, A., & Prieur, L. (1977). Analysis of variations in ocean color. *Limnology & Oceanography*, *22*(4), 709–722. https://doi.org/10.4319/lo.1977.22.4.0709

Neil, C., Cunningham, A., & McKee, D. (2011). Relationships between suspended mineral concentrations and red-waveband reflectances in moderately turbid shelf seas. *Remote Sensing of Environment*, *115*(12), 3719–3730. https://doi.org/10.1016/j.rse.2011.09.010

O'Reilly, J. E., Maritorena, S., Mitchell, B. G., Siegel, D. A., Carder, K. L., Garver, S. A., et al. (1998). Ocean color chlorophyll algorithms for SeaWiFS. *Journal of Geophysical Research*, *103*(C11), 24937–24953. https://doi.org/10.1029/98JC02160

O'Reilly, J. E., Maritorena, S., O'Brien, M., Siegel, D., Toole, D., Menzies, D., et al. (2000). Ocean color chlorophyll a algorithms for SeaWiFS, OC2, and OC4: Version 4. In S. B. Hooker & E. R. Firestone (Eds.), *SeaWiFS postlaunch calibration and validation analyses, part 3. NASA Tech. Memo. 2000-206892 11* (pp. 9–23). NASA Goddard Space Flight Center.

O'Reilly, J. E., & Wertdell, P. (2019). Chlorophyll algorithms for ocean color sensors - OC4, OC5 & OC6. *Remote Sensing of Environment*, *229*, 32–47. https://doi.org/10.1016/j.rse.2019.04.021

Organelli, E., Bricaud, A., Gentili, B., Antoine, D., & Vellucci, V. (2016). Evaluation of bio-optical inversion models for the retrieval of Colored Dissolved Organic Matter (CDOM) and Colored Detrital Matter (CDM) light absorption coefficients in the Mediterranean Sea using field and satellite ocean color radiometry. *Remote Sensing of Environment*, *186*, 297–310. https://doi.org/10.1016/j.rse.2016.08.028

Pope, R., & Fry, E. (1997). Absorption spectrum (380–700 nm) of pure water. II. Integrating cavity measurements. *Applied Optics*, *36*(33), 8710–8723. https://doi.org/10.1364/AO.36.008710

Ramirez-Perez, M., Twardowski, M., Trees, C., Piera, J., & McKee, D. (2018). Inversion of in situ light absorption and attenuation measurements to estimate constituent concentrations in optically complex shelf seas. *Journal of Geophysical Research: Oceans*, *123*(1), 720–737. https://doi.org/10.1002/2017JC013453

Roesler, C. S., & Perry, M. J. (1995). In situ phytoplankton absorption, fluorescence emission, and particulate backscattering spectra determined from reflectance. *Journal of Geophysical Research*, *100*(C7), 13279–13294. https://doi.org/10.1029/95JC00455

Röttgers, R., & Doerffer, R. (2007). Measurements of optical absorption by chromophoric dissolved organic matter using a point-source integrating-cavity absorption meter. *Limnology and Oceanography: Methods*, *5*, 126–135. https://doi.org/10.4319/lom.2007.5.126

Röttgers, R., Häse, C., & Doerffer, R. (2007). Determination of the particulate absorption of microalgae using a point-source integrating-cavity absorption meter: Verification with a photometric technique, improvements for pigment bleaching, and correction for chlorophyll fluorescence. *Limnology and Oceanography: Methods*, *5*, 1–12. https://doi.org/10.4319/lom.2007.5.1

Röttgers, R., McKee, D., & Utschig, C. (2014). Temperature and salinity correction coefficients for light absorption by water in the visible to infrared spectral region. *Optics Express*, *22*(21), 25093–25108. https://doi.org/10.1364/OE.22.025093

Röttgers, R., Schönfeld, W., Kipp, P.-R., & Doerffer, R. (2005). Practical test of a point-source integrating cavity absorption meter: The performance of different collector assemblies. *Applied Optics*, *44*(26), 5549–5560. https://doi.org/10.1364/AO.44.005549

Sanjuan Calzado, V., McKee, D., Trees, C., & Zibordi, G. (2011). Multi and single cast radiometric processing and merging in the Ligurian Sea. *Optics of Natural Waters*, *2011*.

Santoleri, R., Volpe, G., Marullo, S., & Nardelli, B. B. (2008). Open waters optical remote sensing of the Mediterranean Sea. In V. Barale & M. Gade (Eds.), *Remote sensing of the European seas*. Springer. https://doi.org/10.1007/978-1-4020-6772-3_8

Sathyendranath, S., Hoge, F. E., Platt, T., & Swift, R. N. (1994). Detection of phytoplankton pigments from ocean colour: Improved algorithms. *Applied Optics*, *33*(6), 1081–1089. https://doi.org/10.1364/AO.33.001081

Schwarz, J. N., Kowalczuk, P., Kaczmarek, S., Cota, G. F., Mitchell, B. G., Kahru, M., et al. (2002). Two models for absorption by coloured dissolved organic matter (CDOM). *Oceanologia*, *44*(2), 209–241.

Smith, R. C., & Baker, K. S. (1981). Optical properties of the clearest natural waters (200–800 nm). *Applied Optics*, *20*(2), 177–184. https://doi.org/10.1364/AO.20.000177

Smyth, t. J., Moore, G. F., Hirata, T., & Aiken, J. (2006). Semianalytical model for the derivation of ocean color inherent optical properties: Description, implementation, and performance assessment. *Applied Optics*, *45*(31), 8116–8131. https://doi.org/10.1364/AO.45.008116

Wang, P., Boss, E., & Roesler, C. (2005). Uncertainties of inherent optical properties obtained from semi-analytical inversions of ocean colour. *Applied Optics*, *44*(19), 4074–4085. https://doi.org/10.1364/AO.44.004074

Werdell, P. J., & Bailey, S. W. (2005). An improved in-situ bio-optical data set for ocean color algorithm development and satellite data product validation. *Remote Sensing of Environment*, *98*(1), 122–140. https://doi.org/10.1016/j.rse.2005.07.001

Werdell, P. J., McKinna, L. I. W., Boss, E., Ackleson, S. G., Craig, S. E., Gregg, W. W., et al. (2018). An overview of approaches and challenges for retrieving marine inherent optical properties from ocean colour remote sensing. *Progress in Oceanography*, *160*, 186–212. https://doi.org/10.1016/j.pocean.2018.01.001

Wernand, M. R., Shimwell, S. J., & De Munck, J. C. (1997). A simple method of full spectrum reconstruction by a five-band approach for ocean colour applications. *International Journal of Remote Sensing*, *18*(9), 1977–1986. https://doi.org/10.1080/014311697217981

Yacobi, Y. Z., Moses, W. J., Kaganovsky, S., Sulimani, B., Leavitt, B. C., & Gitelson, A. A. (2011). NIR-red reflectance-based algorithms for chlorophyll-a estimation in mesotrophic inland and coastal waters: Lake Kinneret case study. *Water Research*, *45*(7), 2428–2436. https://doi.org/10.1016/j.watres.2011.02.002

Zaneveld, J. R. V., Kitchen, J. C., & Moore, C. (1994). Scattering error correction of reflecting-tube absorption meters. In J. S. Jaffe (Ed.), *Proceedings of ocean optics XII*. SPIE.

Zhang, X., Twardowski, M., & Lewis, M. (2011). Retrieving composition and sizes of oceanic particle subpopulations from the volume scattering function. *Applied Optics*, *50*(9), 1240–1259. https://doi.org/10.1364/AO.50.001240

Zhu, W. N., & Yu, Q. (2013). Inversion of chromophoric dissolved organic matter (CDOM) from EO-1 Hyperion imagery for turbid estuarine and coastal waters. *IEEE Transactions on Geoscience and Remote Sensing*, *51*(6), 3286–3298. https://doi.org/10.1109/TGRS.2012.2224117

Zibordi, G., D'Alimonte, D., & Berthon, J.-F. (2004). An evaluation of depth resolution requirements for optical profiling in coastal waters. *Journal of Atmospheric and Oceanic Technology*, *21*(7), 1059–1073. https://doi.org/10.1175/1520-0426(2004)021<1059:AEODRR>2.0.CO;2