

Distinguishing cyanobacteria from algae in optically complex inland waters using a hyperspectral radiative transfer inversion algorithm

Mark William Matthews^{1,2}, Stewart Bernard³, Hayley Evers-King⁴, Lisl Robertson Lains

Abstract

A hyperspectral inversion algorithm was used to distinguish between cyanobacteria and algal blooms in optically complex inland waters. A framework for the algorithm is presented that incorporates a bio-optical model, a solution for the radiative transfer equation using the EcoLight-S radiative transfer model, and a non-linear optimization procedure. The natural variability in the size of phytoplankton populations was simulated using a two-layered sphere model that generated size-specific inherent optical properties (IOPs). The algorithm effectively determined the type of high-biomass blooms in terms of the relative percentage species composition of cyanobacteria. It also provided statistically significant estimates of population size (as estimated by the effective diameter), chlorophyll-*a* (chl-*a*) and phycocyanin pigment concentrations, the phytoplankton absorption coefficient, and the non-algal absorption coefficient. The algorithm framework presented here can in principle be adapted for distinguishing between phytoplankton groups using satellite and *in situ* remotely sensed reflectance.

Keywords: remote sensing, cyanobacteria, harmful algal blooms, bio-optics, algorithm, hyperspectral

This paper appears in the following journal publication:

Matthews, M. W., Bernard, S., Evers-King, H., & Lain, L. R. (2020). Distinguishing cyanobacteria from algae in optically complex inland waters using a hyperspectral radiative transfer inversion algorithm. *Remote Sensing of Environment*, 248, 111981.

¹ Corresponding author: mark@cyanolakes.com. Other authors: sbernard@csir.co.za; lisl.lain@uct.ac.za; hayley.eversking@eumetsat.int.

² CyanoLakes (Pty) Ltd, 22 Midwood Avenue, Bergvliet, 7945, Cape Town, South Africa.

³ Council for Scientific and Industrial Research, 15 Lower Hope Street, Rosebank, 7700, Cape Town, South Africa.

⁴ EUMETSAT, Eumetsat Allee 1, D-64295, Darmstadt, Germany.

⁵ Department of Oceanography, University of Cape Town, Rondebosch, 7701, Cape Town, South Africa.

1. Introduction

The development of a method for remote sensing for distinguishing between blooms of cyanobacteria and algae is of considerable importance due to the large potential negative effects that toxin producing cyanobacteria may have on the health of humans, animals and aquatic ecosystems (see Matthews et al. 2012). Hyper- and multi-spectral satellite missions present the opportunity to detect species-specific spectral features as postulated in the 1990s by Laurie L. Richardson (Richardson, 1996), and are becoming a reality through several planned hyperspectral satellite missions such as Germany's Environmental Mapping and Analysis Program (EnMAP), Italy's PRecursores IperSpettrale della Missione Applicativa (PRISMA), India's Hyperspectral Imaging Satellite (HySIS), NASA's Plankton, Aerosol, Cloud and ocean Ecosystem (PACE), and ESA's FLuorescence EXplorer (FLEX). Studies focused on differentiating cyanobacteria using ground or satellite spectral measurements have been limited to empirical methods detecting diagnostic spectral features of species such as *Microcystis* (e.g. Zhou et al. 2018; Matthews and Odermatt 2015; Stumpf et al. 2012) and *Trichodesmium* (e.g. Hu et al. 2010; Dupouy et al. 2011), and the retrieval of accessory pigments such as phycocyanin (PC) (e.g. Simis et al. 2005; Hunter et al. 2010). Up till the present study, there have been very few approaches using a physically-based inverse modeling approach to distinguishing cyanobacteria from algae (e.g., Xi et al., 2015; 2017). However, all the studies so far have ignored the natural variability in the spectral inherent optical properties (IOPs), the species percentage composition of cyanobacteria to algae, and the phytoplankton population cell-size, which are considered important diagnostic drivers of the resulting remote sensing reflectance and are also indicators of harmful species. This study addresses some gaps identified by Bracher et al. (2017) related to phytoplankton type identification. Firstly, it presents a useful *in situ* matchup dataset from a selection of diverse inland waters; secondly, it presents new measurements of spectral IOPs for two phytoplankton groups (including backscattering derived using a two layered sphere model); and thirdly, it implements an inversion based on the direct solution of the radiative transfer equation through EcoLight-S (Sequoia Scientific, Inc.). It also presents a new framework for an approach that exploits hyperspectral information.

This study tests the hypothesis that cyanobacteria can be systematically differentiated from algae using remote sensing reflectance on the basis of diagnostic IOPs resulting from pigmentation, cell size and internal structure. It uses the results from an earlier investigation on the effect of intracellular gas vacuoles on spectral scattering in cyanobacteria (Matthews and Bernard 2013a). It develops a novel inversion algorithm framework based on the equivalent algal populations (EAP) model (Robertson Lain et al. 2014)

and the inverse method used in Evers-King et al. (2014) that incorporates a direct solution of the radiative transfer equation with EcoLight-S (see also Rehm and Mobley, 2013). The overall aim is to present a novel spectral-matching inversion algorithm that accounts for variability in the size and type of cyanobacteria and algal populations.

2. Materials and Methods

2.1. Description of sample sites and water types

The details of the sampling areas and the methods used to determine the bio-geochemical parameters and absorption coefficients are provided in Matthews and Bernard (2013b). Briefly, measurements were made in three mid-latitude African freshwater reservoirs: Loskop Dam (LK, 25.42 S, 29.35 E), Hartbeespoort Dam (HB, 25.74 S, 27.86 E), and Theewaterskloof Dam (TW, 34.03 S, 19.26 E). Each of the systems had differing phytoplankton populations: LK was dominated by the large celled dinoflagellate *Ceratium hirundinella* (equivalent spherical diameter (ESD) = 40 μm) with a wide range of biomass (chl-*a* of 0.5 to 500 mg m^{-3}); HB was dominated by a mono-specific *Microcystis aeruginosa* cyanobacteria bloom (ESD = 5 μm) with chl-*a* up to 13 000 mg m^{-3} ; and TW had a mixed phytoplankton population co-dominated by the large dinoflagellate *Sphaerodinium fimbriatum* (ESD = 40 μm) and the filamentous vacuolate cyanophyte *Anabaena ucrainica* (ESD = 16 μm) and various diatom species. Therefore, the data can be summarised as being derived from a) a large-celled dinoflagellate bloom, b) a small-celled gas-vacuolate cyanobacteria bloom, and c) a mixed bloom of large-celled dinoflagellates and intermediate-celled gas vacuolated cyanobacteria. For the purposes of analysis, the data were grouped into the following cases based on OECD trophic classes: oligotrophic (chl-*a* < 10 mg m^{-3} , oligo); meso-eutrophic dinoflagellate-dominant (chl-*a* 10 to 30 mg m^{-3} , meso_dino); meso-eutrophic mixed (meso_mixed); hyper-eutrophic dinoflagellate-dominant (chl-*a* > 30 mg m^{-3} , dino); and hyper-eutrophic cyanobacteria-dominant (cyano).

2.2. Remote sensing reflectance

A total of 63 remote sensing reflectance (R_{rs}) spectra from the three systems were used in the algorithm evaluation (Fig. 1). R_{rs} was measured using an ASD FieldSpec™ 3 Portable Spectroradiometer (ASD Inc., Boulder, CO) using the protocols outlined in Mueller et al. (2003). Measurements were only performed under mostly clear sky conditions (cloud cover < 20%) since cloud cover is known to impart large errors from shadows and other effects (Mobley 1999; Doxaran et al. 2004). Ten radiance spectra were collected in sequence for a Spectralon™ plaque, sky and water targets in order to minimise the effects of wind and waves and temporal variability in surface reflectance. Measurements were performed between 9 am and 12 pm using a viewing zenith angle of $\theta = 40^\circ$ away from the sun azimuthally at $\phi = 135^\circ$. Care

was taken to ensure that the plaque was free from shadow or reflectance effects from any source. Sky radiance was measured using the same viewing angle to the zenith. The measurement procedure was performed three times in sequence with dark readings taken between each sequence. The mean of the radiance spectra for each target was then computed, taking care to exclude contaminated or outlying spectra by visual inspection. R_{rs} was then calculated using the mean spectrum for each target according to Mobley (1999):

$$R_{rs} = (L_t - \rho L_s) / \left(\frac{\pi}{R_g} L_g \right) \quad (1)$$

where L_t is the water surface radiance, L_s is the sky radiance, ρ is the proportionality factor for the sky radiance to the reflected sky radiance on the water surface, π is 3.14, L_g is the radiance measured from the Spectralon™ plaque, and R_g is the Spectralon™ bi-directional reflectance function. The Spectralon™ plaque was assumed to be Lambertian and a perfect reflector ($R_g = 0.99$). Since the wind speed of all measurements was less than 5 m s⁻¹ a value for ρ of 0.025 was used (see Fig. 9 in Mobley, 1999).

2.3. Phytoplankton Type and Size

Two estimates of the phytoplankton population size were determined, namely the effective diameter (D_{eff}) and the equivalent spherical diameter (ESD). Phytoplankton identification and counts were performed by microscopy. These were used to calculate ESD using literature estimates of cell volumes for individual species mainly from Reynolds (2006). Particle size distributions (PSDs) were determined using a Multisizer-4™ particle analyzer (BeckmanCoulter®). The 140 µm aperture allowed measurements of particles between 2.8 and 84 µm. Samples were kept cool and in the dark until analysis which was performed on the same day as collection. Fresh water samples were diluted using Isoflow solution (BeckmanCoulter®), after which 20 ml was counted maintaining a concentration of between 2.5 and 10% with correction for particle coincidence. Blank particle counts were measured using freshly 0.2 µm filtered and sample water that was diluted identically. PSDs were corrected by subtracting the mean blank particle counts that were scaled for dilution. PSDs in cells per liter were median filtered to reduce spiking and interpolated onto linear spaced bins of 1 µm diameter through calculation of the spectral density.

PSDs were partitioned into algal and non-algal components using a numerical technique (cf. Bernard et al. 2001). The detrital distribution was estimated as a Jungian distribution with slope of -4 and scaled to the minimum volume of the PSD between 1 and 7 µm. The detrital distribution was then subtracted from the PSD to give the estimated phytoplankton size distribution. The effective radius (R_{eff}) and variance (V_{eff}) of the phytoplankton size distributions were calculated as follows (Hansen and Travis, 1974):

$$r_{eff} = \frac{\int \pi r^3 F(r) d(r)}{\int \pi r^2 F(r) d(r)} = \frac{1}{G} \int \pi r^3 F(r) d(r) \quad (5)$$

$$v_{eff} = \frac{1}{Gr_{eff}^2} \int (r - r_{eff})^2 \pi r^2 F(r) d(r) \quad (6)$$

where r is the particle radius in m, $F(r)$ is the number of particles per unit volume (cells per m³), and $d(r)$ is the difference between the size bins in meters. The effective diameter or D_{eff} ($= 2R_{eff}$) was used to describe a mean particle size.

2.4. Phytoplankton IOPs

The IOPs for cyanobacteria and dinoflagellates were measured (see Matthews & Bernard, 2013b), and the size-specific absorptions and backscattering coefficients were generated using a two-layered sphere model (see Matthews and Bernard, 2013a for details). For cyanobacteria the model was configured with an internal gas vacuole occupying 50% of the cell volume, surrounded by a chromatoplasm. For dinoflagellates the model consisted of an internal cytoplasm surrounded by a chloroplast layer. The complex refractive indices (m), consisting of the imaginary (n') and real (n) components for the absorbing layers, were derived from measurements of the particle size distribution (PSD) and absorption of natural populations of *M. aeruginosa* and *C. hirundinella* (see Fig. 2).

Using the refractive index data, the two-layer model was run for hypothetical populations of cells characterized by cell size. For the purposes of this study, standard size distributions with D_{eff} ranging from 1 to 50 μm with a 1 μm resolution were used (Bernard et al. 2001, 2007), with the value of V_{eff} set to 0.6. The population of cells had diameters ranging from 1 to 100 μm at a 1 μm size interval. The intracellular chlorophyll concentration (c_i) was used to normalise the PSD and volume coefficients to produce chl-*a* specific IOPs. The value of c_i for *Microcystis sp.* cyanobacteria was 2.1 kg m⁻³ (Zhou et al. 2012) and 3.2 kg m⁻³ for *C. hirundinella* calculated from experimental data.

The output of the two layered model was the size-and-chl-*a*-specific IOPs for the hypothetical cyanobacteria and dinoflagellate populations (Fig. 3). The phytoplankton IOPs were therefore characterized in terms of both size (D_{eff}) and concentration of chl-*a* (C) in mg m⁻³.

2.5. Algorithm framework

The algorithm uses an implicit spectral matching approach consisting of three components: a bio-optical model for estimating the IOPs; a radiative transfer model for calculating R_{rs} ; and a non-linear optimization algorithm for matching the estimated and observed R_{rs} (Fig. 4).

2.6. Bio optical model

Water constituents were partitioned into living phytoplankton (ϕ), chromophoric dissolved organic matter here referred to as *gelbstoff* (g), and non-algal particles (NAP) which includes minerals and detritus (also referred to as d), not neglecting water itself (w). The IOPs of other water constituents such as bubbles, viruses, bacteria and very small particles e.g. colloids, are not explicitly determined for this study, but may

also be significant contributors to the total IOPs (e.g. Stramski et al. 2004). The four-component bio-optical model was used to calculate the total volume coefficients according to:

$$a_t = a_\phi + a_{gd} + a_w \quad (2)$$

$$b_t = b_\phi + b_{nap} + b_w \quad (3)$$

$$b_{bt} = b_{b\phi} + b_{bnap} + b_{bw} \quad (4)$$

where a , b and b_b , are the absorption, scattering and backscattering volume coefficients, respectively.

The total phytoplankton coefficients were calculated as an admixture of the cyanobacteria and dinoflagellates types. The phytoplankton admixture coefficient, T , varying between 0 and 1, represented the relative contributions of cyanobacteria and dinoflagellates to the overall phytoplankton population (1 = 100% cyanobacteria, 0 = 100% dinoflagellates). The concentration of C was used to calculate the total phytoplankton IOPs:

$$i_\phi = C \times [Ti_{\phi c}^*(D_{effc}) + (1 - T)i_{\phi d}^*(D_{effd})] \quad (7)$$

where i is a , b or b_b at the relevant effective diameter, and c and d represent cyanobacteria or dinoflagellates, respectively. The size parameters (D_{eff}) for dinoflagellates and cyanobacteria were allowed to vary independently of each other. *Gelbstoff* and detrital (including minerals) absorption (a_{gd}) were coupled because of their similar spectral shapes:

$$a_{gd} = a_{gd}(442)e^{-S(\lambda-442)} \quad (8)$$

where S is the slope coefficient where the mean value of 0.013 for the study areas was used (see Matthews and Bernard 2013b).

An independent non-algal particle (back)scattering coefficient was used as it accounted for other particles (e.g. very small particles, bubbles and bacteria), and because there was variability in the mass-specific *tripton* coefficients between the study areas (see Matthews and Bernard 2013b). A power-law function has been determined to provide a close fit with measurements of the particulate backscattering in coastal and inland waters (e.g. Sun et al. 2009; Snyder et al. 2008):

$$b_{bnap} = b_{bnap}(560) \times (\lambda/560)^\gamma \quad (9)$$

where γ is the slope coefficient, the value of which typically ranges from 0 to -2 nm⁻¹ with a mean near -1 nm⁻¹ (*ibid.*). A spectrally flat γ (= 0 nm⁻¹) was used as it was found to increase the sensitivity to phytoplankton related spectral features.

Based on a review of studies performed in complex waters, from sediment-dominated coastal waters to turbid eutrophic lakes, the value of the backscattering ratio, \tilde{b}_{bp} , ranges from approx. 0.5 to 7% (Snyder et al. 2008; Sun et al. 2009; McKee et al. 2009; O'Donnell et al. 2010; Neukermans et al. 2012). The cited studies refer to the bulk particulate matter (back)scattering and not non-algal particle backscattering as used here (Equation 9). A spectral dependency of \tilde{b}_{bp} has been noted by some authors (Snyder et al. 2008; McKee et al. 2009) although this is somewhat disputed (Whitmire et al. 2007). In productive turbid waters, \tilde{b}_{bp} has been reported as being inversely proportional to the inorganic or mineral component of suspended matter, and poorly correlated with chl-*a* (Sun et al. 2009), typically varying between 1 and 2%. As phytoplankton scatter poorly in comparison with minerals, b_{nap} is likely to differ little from b_p , except in highly productive waters. It is also probable that b_{nap} that excludes the phytoplankton component has less spectral dependence than b_p . Therefore, a spectrally flat \tilde{b}_{bp} with values ranging from 1 to 5% was assessed in the calculation of b_{nap} :

$$b_{nap} = \frac{1}{\tilde{b}_{bnap}} \times b_{bnap} \quad (10)$$

where \tilde{b}_{bnap} is the ratio $b_{bnap}:b_{nap}$.

Phycocyanin (PC) is commonly used as an indicator pigment for cyanobacteria blooms, and various algorithms have been developed for its detection (e.g. Simis et al. 2005; Hunter et al. 2010). PC was retrieved from $a_\phi(\lambda)$. The absorption by PC at 620 nm, $a_{pc}(620)$, was calculated as $a_\phi(620) - 0.24 \times a_\phi(665)$ after Simis et al. (2005). PC was then retrieved using the relationship determined in Matthews and Bernard (2013b):

$$PC = (a_{pc}(620)/0.0146)^{1.076} \quad (11)$$

2.7. Radiative transfer model

EcoLight-S is a radiative transfer model that solves the radiative transfer equation with very fast run times. The use of EcoLight-S avoids the assumptions and uncertainties associated with reflectance approximations and the associated bi-directional function (f/Q), which up to the present has not been

adequately characterised in high-biomass, turbid and optically complex waters (Lee et al. 2011; Chami et al. 2006). In highly scattering waters, the single scattering approximation breaks down, resulting in large changes in the radiance distribution and the f/Q factor (Piskozub and McKee 2011; Aurin and Dierssen 2012).

The inputs to EcoLight-S are the total absorption, scattering and backscattering volume coefficients (a_t , b_t , b_{bt}). Fluorescence is not calculated by EcoLight-S. The required incident downwelling irradiance (E_d) was computed using the Radtran Sky irradiance model (Gregg and Carder 1990). Atmospheric pressure, total precipitable water vapour column in cm (WV), and aerosol optical thickness (AOT) (Microtops II sun photometer) were measured co-incident to R_{rs} and used as input. Horizontal visibility was calculated from AOT at 550 nm according to $3.9449 = (AOT(550) - 0.08498)$ (Retalis et al. 2010). Relative humidity was estimated from WV according to $18WV + 40$ (calculated using data in Raj et al. 2004). Cloud cover was estimated in % as observed *in situ*, and a continental aerosol model was used.

In EcoLight-S the IOPs are assumed to be constant within homogeneous layers. In this study, a homogeneous optically-deep water profile was used with no bottom reflectance. Although the vertical profile of the IOPs significantly changes R_{rs} (e.g. Kutser et al. 2008), the turbid waters under investigation in this study have very shallow optical depths. On average Secchi disk depth was less than 2 m (maximum of approx. 8 m for a few clear water samples). A spectrally variant Fournier and Forand (1994) phase function dependent on the value of $b_b:b$ and shown to provide improved optical closure in optically complex waters was used (e.g., Mobley et al. 2002; Tzortziou et al. 2006; Gallegos et al. 2008).

2.8. Non-linear optimization technique

A Nelder-Mead (1965) downhill simplex algorithm was used to fit the measured R_{rs} spectrum between 400 and 800 nm, using the Euclidian distance and a wavelength weighting function, $f(\lambda)$:

$$d(x, y) = \sum_{i=1}^N f(\lambda_i) \times (|x_i - y_i|)^2 \quad (12)$$

where x and y are the measured and modelled spectra, and $f(\lambda)$ is a spectral function used to assign a reduced weight (1%) to wavelengths in the chl-*a* fluorescence domain (680 to 695 nm). The complex fluorescence effects at high concentrations are thereby not taken into account by the algorithm. The simplex algorithm iterates until the estimated variables change by less than the tolerance level, which was 1×10^{-6} , or the maximum number of iterations is reached.

Six unknowns were solved for: chl-*a*, the admixture coefficient (T), D_{eff} for cyanobacteria and

dinoflagellates, $b_{bnap}(560)$ and $a_{gd}(442)$. Various initial values for each of the unknowns were tested (Table 1). The initial value for chl- a was estimated using an empirically-derived relationship between chl- a and the $R_{rs}(710):R_{rs}(665)$ reflectance ratio (R):

$$Chl = -6.1R^2 + 91.3R - 47.7. \quad (13)$$

The starting value of T was determined using a binary flag based on spectral features (Eq. 4 in Matthews et al. 2012) and the initial chl- a estimate (Table 2). The initial values for the effective diameters for cyanobacteria and dinoflagellates were 5 and 31 μm simulating a small and large species, respectively. The initial value of $a_{gd}(440)$ was 2.5 m^{-1} and ranged between 0.2 to 6.0 m^{-1} .

The non-parametric (Spearman's rank) coefficient of determination (R_2) and log-scaled root mean square error (log-RMSE) was used to assess the performance of the algorithm estimates, as the data were non-normally distributed and varied on a log scale:

$$\text{log-RMSE} = \sqrt{\frac{\sum_{i=1}^N [\log(\hat{x}_i) - \log(x_i)]^2}{N-2}} \quad (16)$$

where x is the measured value, \hat{x} is the estimated value of the unknown, and N is the sample size.

3. Results

3.1. Retrieval of phytoplankton type and size

In general, there was a close agreement between the measured and modeled reflectance across the cases (Fig. 5 and 6). Spectral fitting was most challenging in mineral rich (4A) and very clear waters (4D) (note the order of magnitude difference in spectral reflectance). In general, good fitting in low-to-mid biomass waters did not depend on accurate determination of type. For hyper-eutrophic cyanobacteria-dominant and dinoflagellate-dominant cases, where type (T) was correctly determined, good fitting was obtained even at extreme biomass levels (5C). However, where type was inconclusive, the fitting was correspondingly poor (5D).

By comparing the values of the estimated type parameter and the water classes, it was determined that in general the algorithm correctly differentiated between cyanobacteria and dinoflagellate blooms (Fig. 7C, Table 3). The average type parameter for cyanobacteria-dominant waters was 0.93, while that for dinoflagellate-dominant waters was 0.11. Oligotrophic and mixed waters had a median value near 0.5, indicating dominance by neither type. Successful type detection was largely determined by the algorithm's ability to fit spectral features in the 500 to 650 nm spectral range (Fig 6D). The linear correlation coefficient

between measured and modelled R_{rs} was typically larger than 0.8, except at wavelengths < 450 nm (possibly caused by residual sun glint), at the narrow chl-*a* fluorescence band near 685 nm, and at the fluorescence-related 761 nm feature caused by fill-in effects of Telluric lines noticeable in dinoflagellate-dominant cases (Lu et al. 2016).

The algorithm effectively determined the size of the small-celled and large-celled blooms (Fig. 7A, B). D_{eff} for the large-celled dinoflagellate bloom was estimated as ranging between 20 and 42 μm , while that for the small-celled cyanobacteria bloom ranged from two to 15 μm . There was a tendency to overestimate the size in oligo-mesotrophic cases (Table 3). Fitting was significantly better for hyper-eutrophic cases than oligo-mesotrophic cases likely caused by the spatial distribution of the data (Table 4). D_{eff} was smaller than ESD due to the inclusion of smaller particles by the coulter-counter measurement technique and was not available for some samples.

3.2. Retrieval of pigment concentrations and absorption coefficients

The algorithm provided fair estimates of chl-*a*, PC and a_{ϕ} , but poor estimates for a_{gd} (Fig 8, Table 4). The extreme range of concentrations makes statistical evaluation more challenging. Chl-*a* tended to be overestimated in mesotrophic cases and underestimated in hyper-eutrophic cases. The extreme concentration range (from 0.5 to 10 000 mg m^{-3}) likely reduced the overall log-RMSE which was 1.09 ($R_2 = 0.64$). PC was more challenging to determine than chl-*a* as indicated by the higher log-RMSE value of 1.22 ($R_2 = 0.35$). PC was only estimated for cases where cyanobacteria were present or dominant, resulting in a smaller sample size. The determination of a_{ϕ} was slightly better in oligo-mesotrophic cases (log-RMSE = 1.0, $R_2 = 0.82$) than in hyper-eutrophic cases (log-RMSE = 1.09, $R_2 = 0.41$).

The retrieval of a_{gd} was satisfactory, albeit consistently underestimated, in oligo-mesotrophic cases (log-RMSE = 0.89, $R_2 = 0.66$). By contrast there was no significant correlation in hyper-eutrophic cases: it is most likely impossible to separate absorption from dissolved substances from that of phytoplankton in extreme bloom conditions.

The log-RMSE values in this study are noticeably high when compared to other published studies (e.g. IOCCG, 2006). However, other similar published studies do not handle similar extreme bloom conditions and pigment concentrations over five orders of magnitude. Further, the focus of the present study is to develop an approach for bloom discrimination, while accurate determination of pigment concentrations and IOPs are of secondary importance.

4. Discussion

4.1. The bio-optical basis for distinguishing cyanobacteria from algae

In this study, cyanobacteria and algal blooms were accurately distinguished on the basis of their

spectral reflectance using a non-linear spectral inversion algorithm based on a four-component bio-optical model. This is possible because of cellular-level differences in size, pigmentation, and internal structure between cyanobacteria and algae, briefly discussed here.

The large *C. hirundinella* cells (mean diameter 40 μm) had considerably larger absorption efficiency relative to the smaller cyanobacteria cells. The increased pigment packaging resulted in a considerably smaller mean value of $a_{\phi}^*(675)$ of $0.0089 \text{ m}^2 \text{ mg}^{-1}$ in contrast to that of $0.023 \text{ m}^2 \text{ mg}^{-1}$ for cyanobacteria. The small cell size and intracellular gas vacuole structure used to model cyanobacteria resulted in $b_{b\phi}^*(510)$ values more than two orders of magnitude higher for cyanobacteria ($5.5 \times 10^{-3} \text{ m}^{-1}$) than for the dinoflagellates ($8 \times 10^{-5} \text{ m}^{-1}$). R_{rs} spectra measured for cyanobacteria blooms had considerably larger magnitude than those measured for dinoflagellate blooms of similar biomass. $R_{rs}(560)$ values ranged from 0.013 to 0.048 sr^{-1} for cyanobacteria, an order of magnitude larger than those for the dinoflagellate bloom that ranged from 0.003 to 0.011 sr^{-1} . The larger magnitude can also be attributed to the tendency of cyanobacteria to be vertically stratified in the upper surface layers (Kutser et al. 2008).

Diagnostic phycocyanin and phycoerythrin pigments present in cyanobacteria resulted in marked troughs in R_{rs} between 600 and 650 nm, and a shift in the green peak to wavelengths less than 550 nm. By contrast, dinoflagellate blooms were characterized by strong absorption from peridinin carotenoid and diadinoxanthin and diatoxanthin xanthophylls (Schluter et al. 2006; Richardson 1996) resulting in characteristic absorption troughs in the 440 to 550 nm region, and a green-peak near 560 nm. Diagnostic spectral features in R_{rs} that can be used to discriminate cyanobacteria from algae are changes in the position of the green peak towards lower wavelengths near 550 nm, the presence of a peak near 650 nm from proximal PC and chl-*a* absorption bands, and a noticeable lack of chl-*a* fluorescence related features near 685 nm (see Seppälä et al. 2007).

5. Conclusion and recommendations

The study has demonstrated how a hyperspectral inversion algorithm incorporating a radiative transfer model can be used to differentiate between cyanobacteria and algal blooms in inland waters, in waters with an extreme range of phytoplankton biomass. This study demonstrated how a modified EAP algorithm framework (see Evers-King et al. 2014, Robertson Lain et al. 2014) can be used to differentiate phytoplankton groups on the basis of type-specific IOPs (in this case cyanobacteria and dinoflagellates). It is the first study to demonstrate how cyanobacteria may be distinguished from eukaryotic algae using an algorithm incorporating a direct solution to the radiative transfer equation through EcoLight-S.

The algorithm can likely be adapted to distinguish between other phytoplankton groups (e.g. diatoms or cryptophytes) to enable more general application, and in principle be used to resolve phytoplankton groups (and possibly size) from satellite-based sensors. This is, however, dependent upon the availability

of appropriate hyper or multi-spectral resolution instruments (such as EnMAP or FLEX) with adequate signal to noise ratios, and accurate estimation of the water-leaving reflectance through atmospheric correction applied to these instruments.

6. Acknowledgements

M.W.M. would like to gratefully acknowledge those who assisted in the collection of field data including Nobuhle Majozi, Heidi van Deventer, Marie Smith, the Department of Water Affairs, Nadene Slabbert, the Theewaterskloof Sports Club, the Loskop Nature Reserve, and Paul Oberholster. The Council for Scientific and Industrial Research (CSIR) Safe Waters Earth Observation Systems project funded the field data collection. The University of Cape Town and the CSIR funded M.W.M.'s PhD research that forms the basis of this work. The early reviews of Emmanuel Boss, and three anonymous reviewers helped greatly to focus and improve the work presented here. The authors declare no conflict of interest.

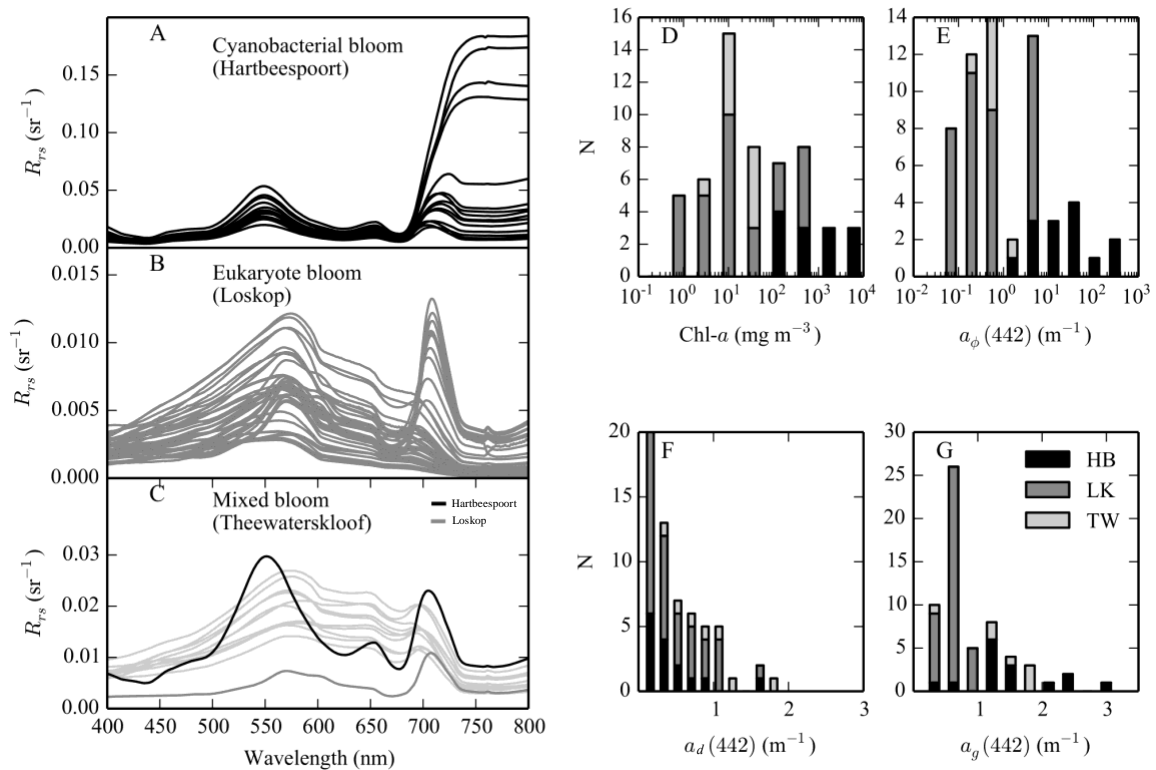


Fig. 1 R_{rs} measured in cyanobacteria-dominant waters (Hartbeespoort) (A), dinoflagellate-dominant waters (Loskop) (B), and mixed waters (Theewaterskloof) (C). The measured distributions of chl-*a* (D), $a_{\phi}(442)$

(E), $a_d(442)$ (F) and $a_g(442)$ (G). R_{rs} spectra from (A) and (B) are overlaid in (C) for comparison of shape and magnitude.

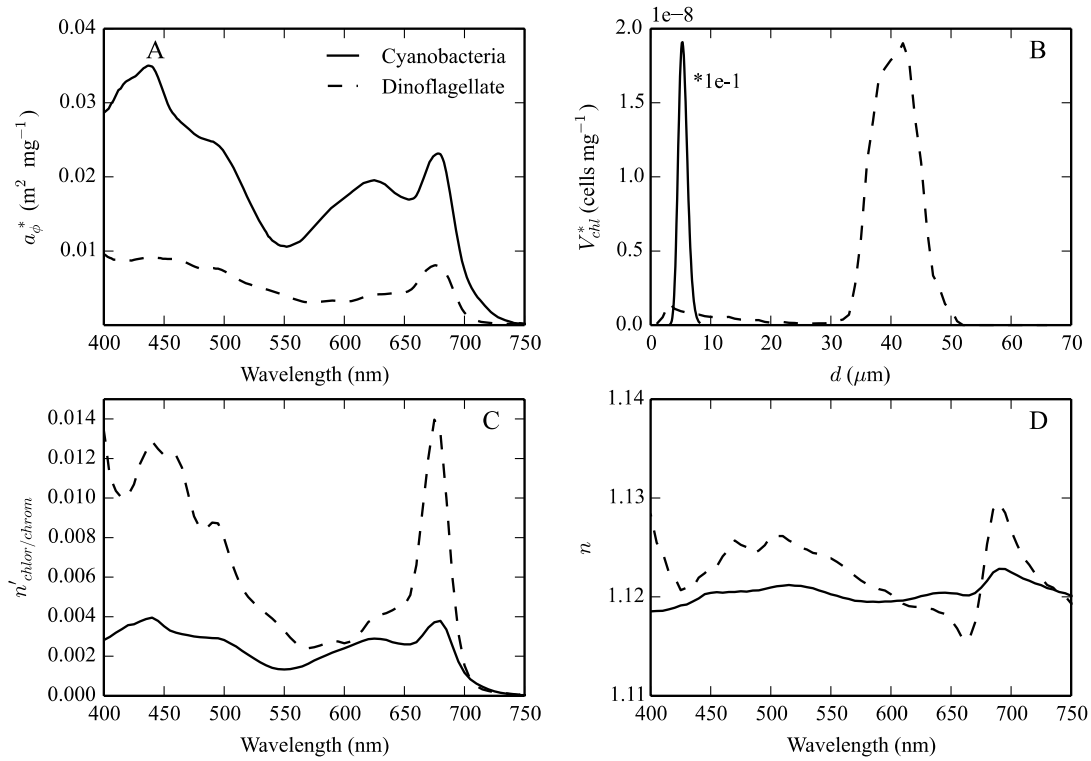


Fig 2. Properties of *M. aeruginosa* (cyanobacteria) and *C. hirundinella* (dinoflagellate) used in the two-layered model simulations. (A) Measured chl-*a* specific absorption coefficients; (B) the chl-*a* specific phytoplankton volume distribution (V_{chl}^*); (C) imaginary refractive index for cyanobacteria chromatoplasm (n'_{chrom}) for cell with 50% gas vacuole volume, and dinoflagellate chloroplast (n'_{chlor}) with 30% cell volume, (D) real refractive index (n) for cyanobacteria chromatoplasm and dinoflagellate chloroplast layers.

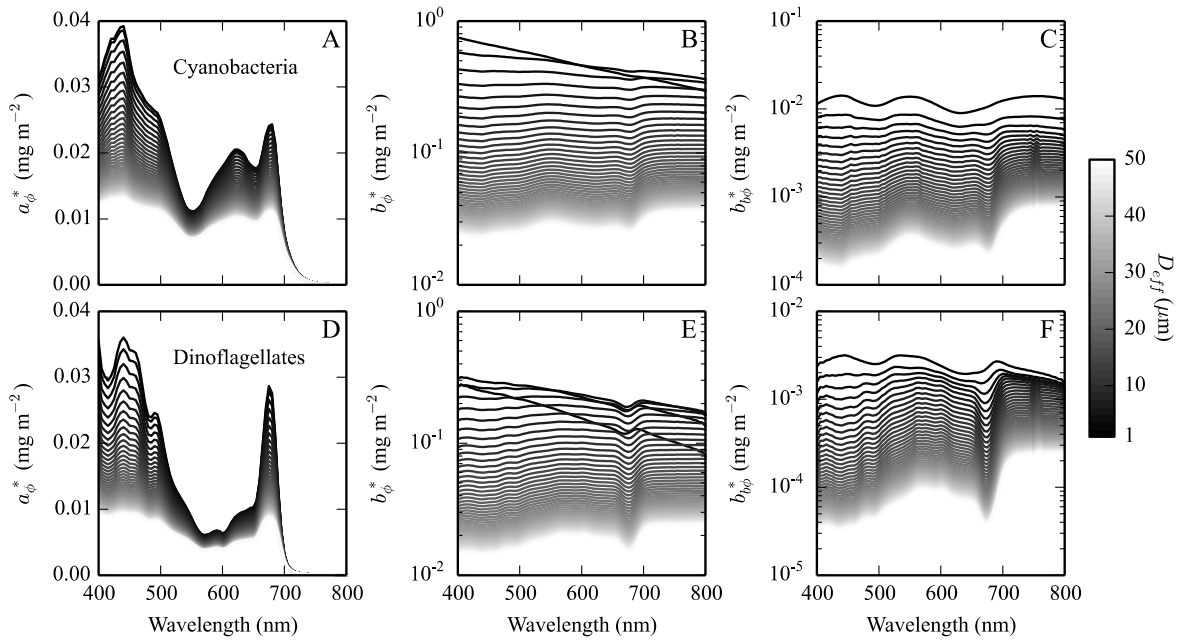


Fig. 3. Chl-*a* specific volume coefficients modelled using a two-layered sphere for cyanobacteria and dinoflagellates populations with D_{eff} from 1 to 50 μm.

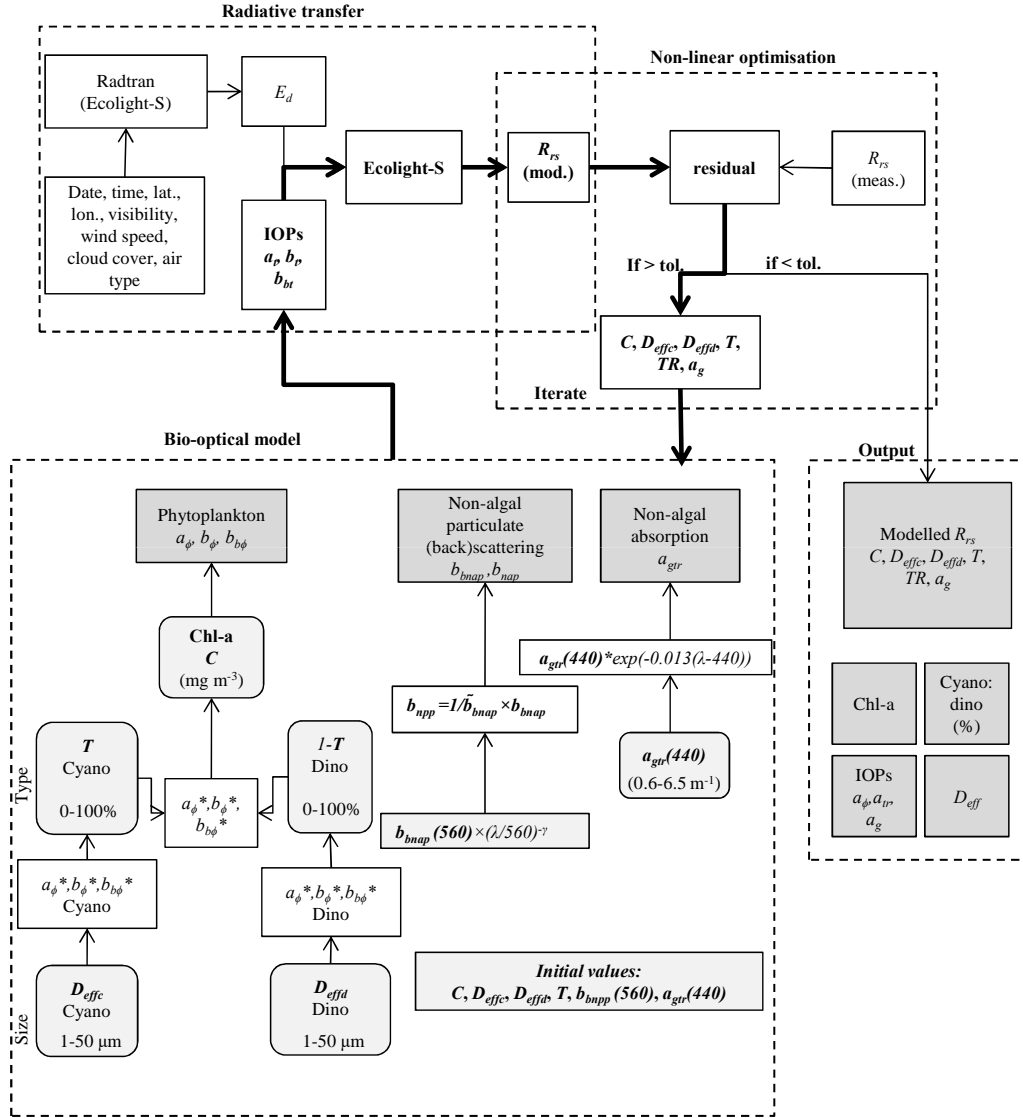


Fig. 4. Framework for the radiative transfer inversion algorithm. Light grey shading represents variables solved for iteratively, and darker grey shading represents outputs. The bold text and arrows represent the iterative process followed by the algorithm. $tol.$ = tolerance. See text for abbreviations.

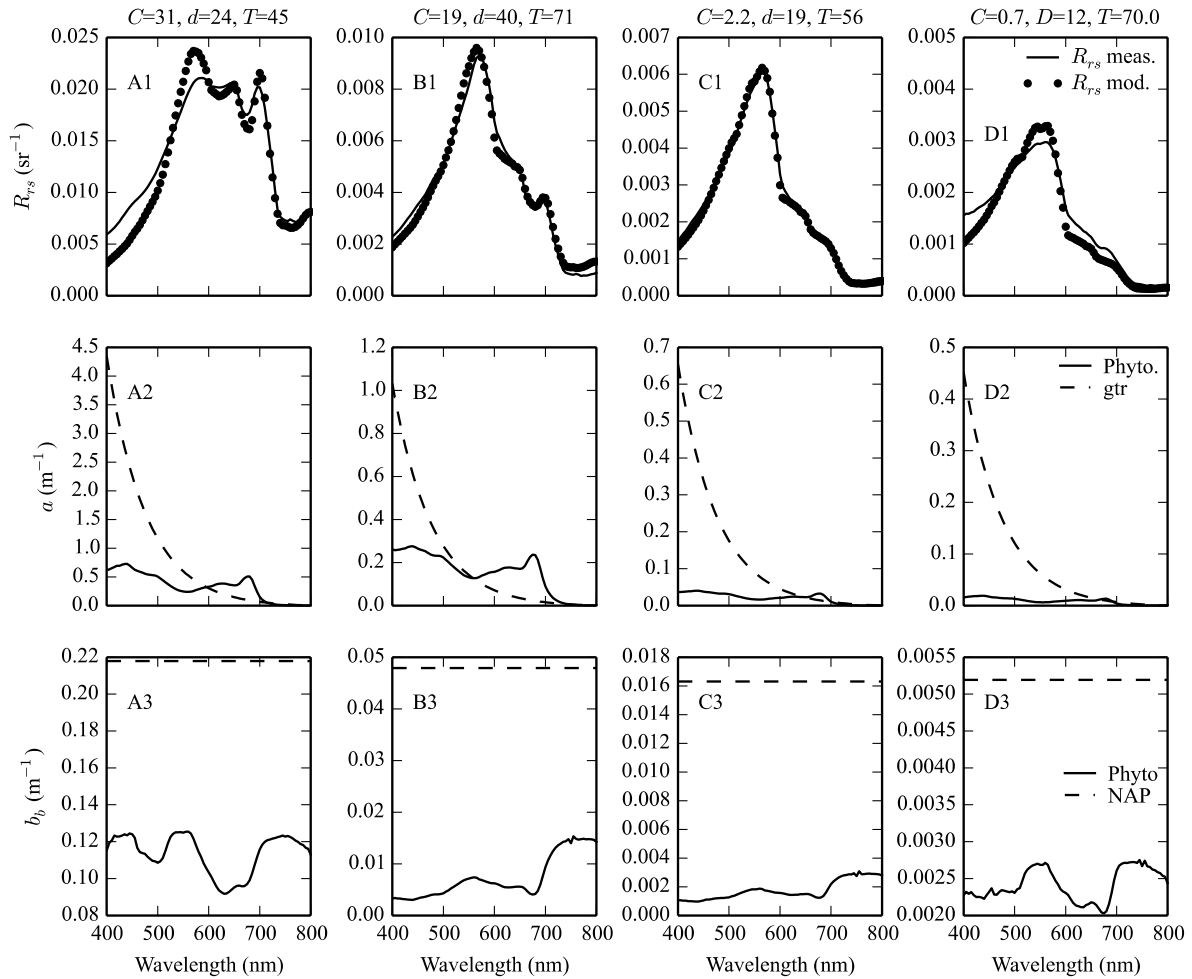


Fig. 5. Examples of spectral fitting in a mixed medium-biomass bloom (A1), mesotrophic dinoflagellate-dominant bloom (B1), and oligotrophic clear waters (C1, D1). Rows 2-3 show corresponding modelled a and b_b coefficients. The titles show the estimated concentration of chl- a (mg m^{-3}) = C , D_{eff} (μm) = d , and type = T as % cyanobacteria.

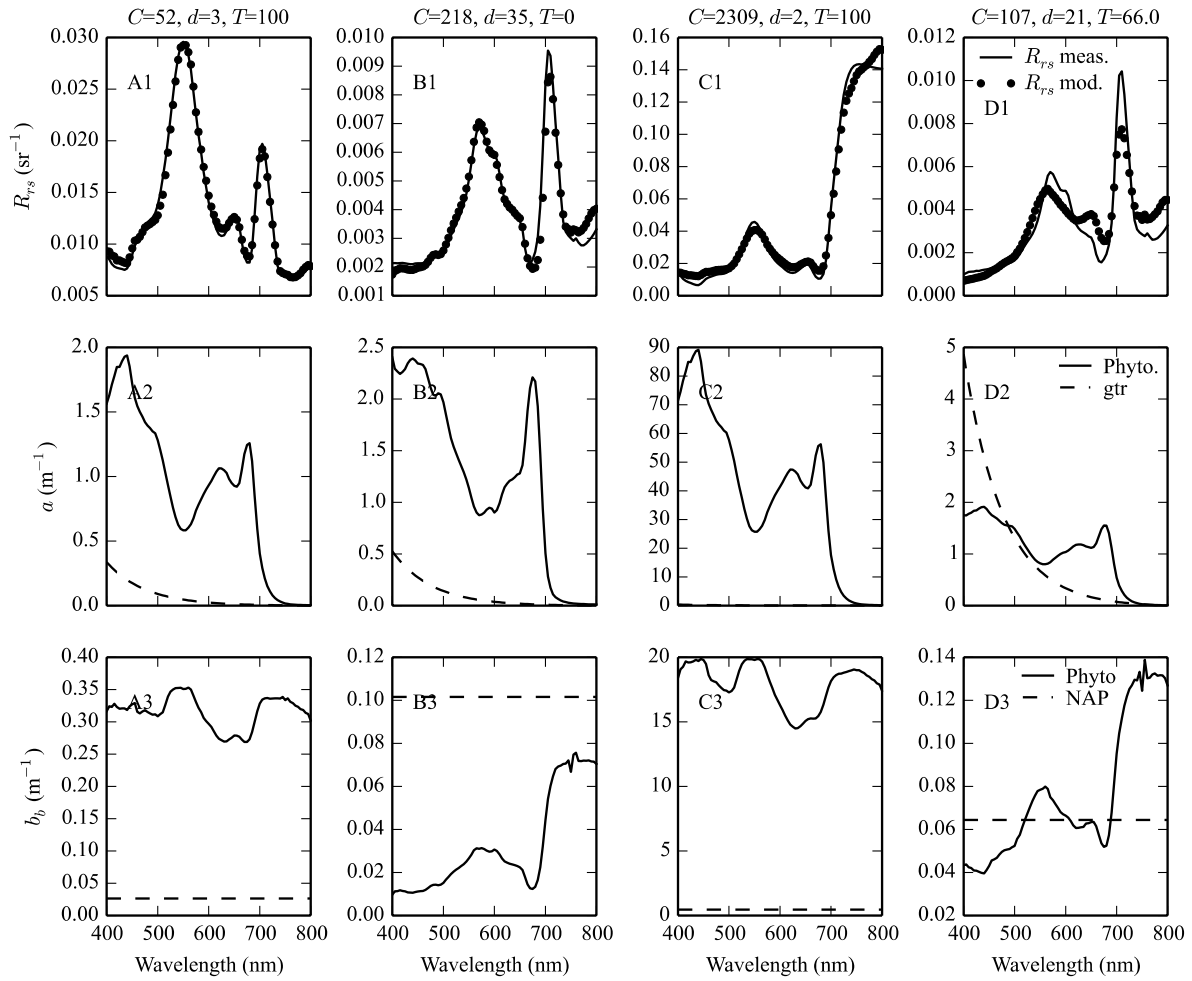


Fig. 6. As for Fig. 4 except for high-biomass cyanobacteria bloom (A), high-biomass dinoflagellate bloom (B), floating cyanobacterial scum (C), and high-biomass dinoflagellate bloom (D).

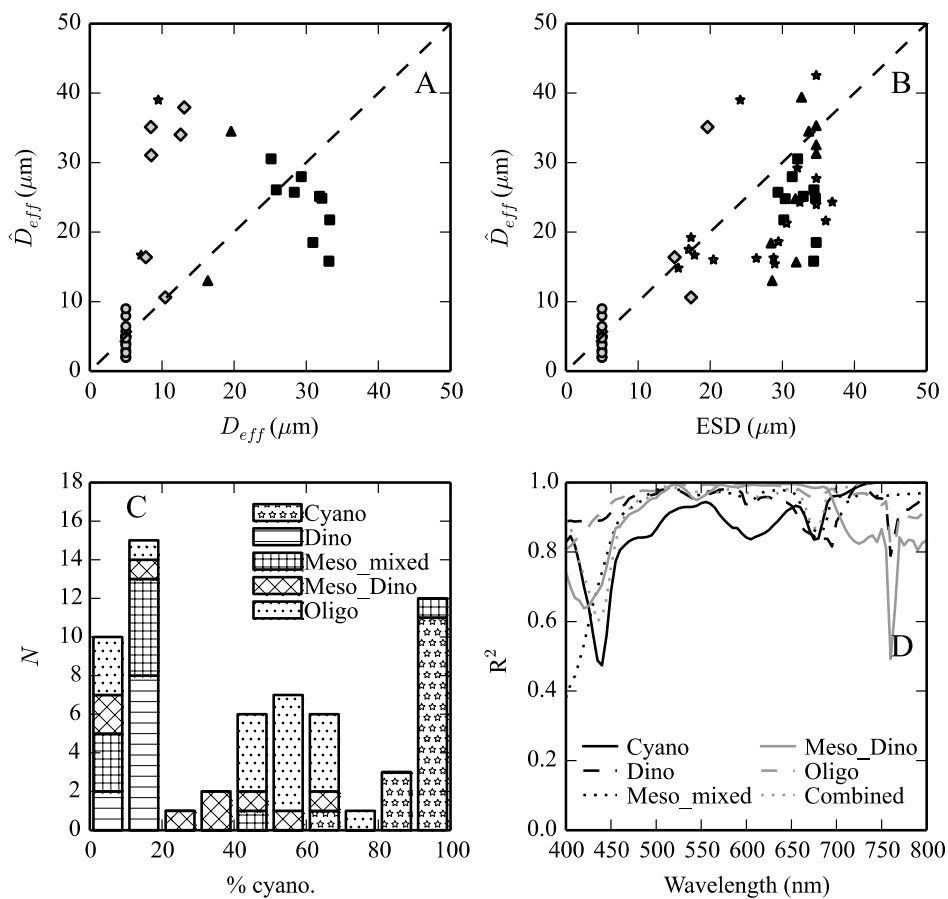


Fig. 7. Modelled (y axis) versus measured (x axis) size parameters D_{eff} and ESD (A - B), histograms where N is the number of spectra binned by percent cyanobacteria (C), and the wavelength-specific correlation coefficients (D).

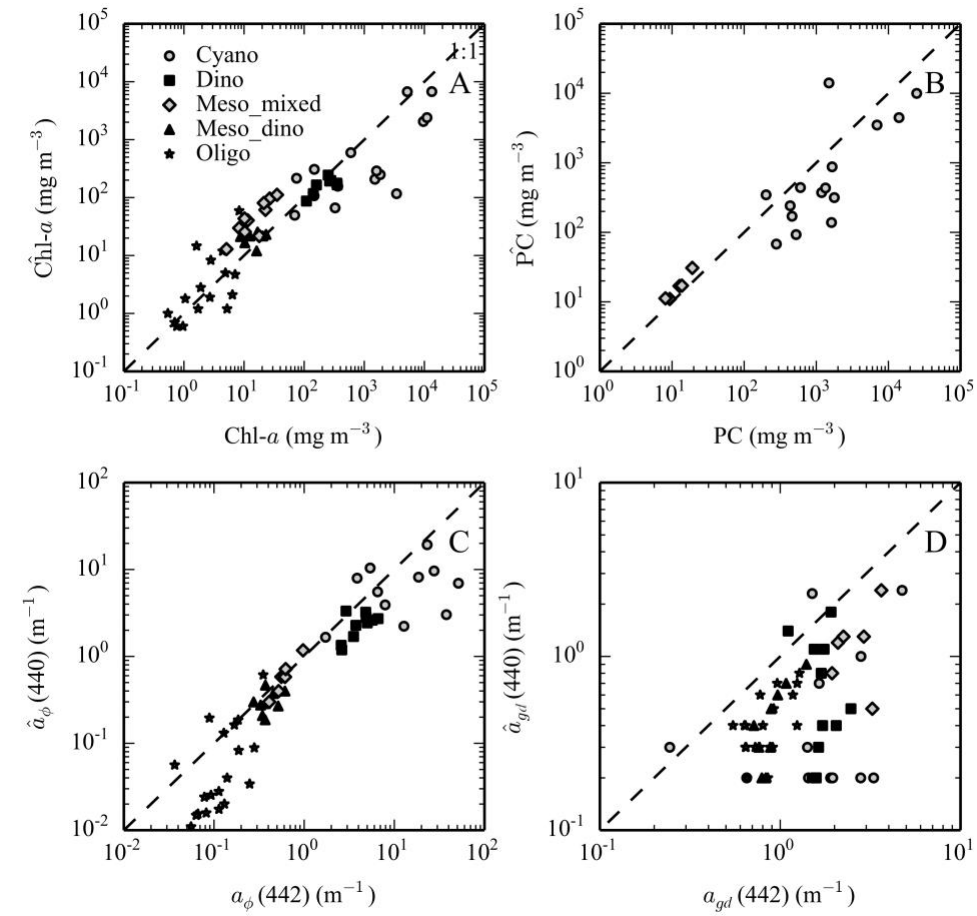


Fig. 8. Modelled (y-axis) versus measured chl-*a* (A), PC (B), $a_\phi(440)$ (C) and $a_{gd}(440)$ (D). Note log scale axes.

Table 1. Experiment testing different sets of initial conditions and resulting change from the default or optimal initial conditions.

| Initial conditions | | | | | | Constants | | Change |
|--------------------|-------------|-----------------|-----------------|------------|------------|------------------|----------------------|--------------------------|
| C | T | $a_{gd}(440)$ | $b_{bnap}(560)$ | D_{effc} | D_{effd} | γ | $1/\tilde{b}_{bnap}$ | |
| mg m ⁻³ | % | m ⁻¹ | m ⁻¹ | μm | μm | nm ⁻¹ | m | |
| Empirical | Flag | 2,5 | 0,1 | 5 | 31 | 0 | 50 | Default / Optimal |
| 10 | 50 | 2,5 | 0,1 | 5 | 31 | 0 | 50 | Large |
| 100 | 50 | 2,5 | 0,1 | 5 | 31 | 0 | 50 | Small |
| empirical | 95 | 2,5 | 0,1 | 5 | 31 | 0 | 50 | Large |
| empirical | 5 | 2,5 | 0,1 | 5 | 31 | 0 | 50 | Large |
| empirical | 50 | 0.5 | 0,1 | 5 | 31 | 0 | 50 | Medium |
| empirical | 50 | 4,0 | 0,1 | 5 | 31 | 0 | 50 | Medium |
| empirical | 50 | 2,5 | 0,001 | 5 | 31 | 0 | 50 | Large |
| empirical | 50 | 2,5 | 0,01 | 5 | 31 | 0 | 50 | Large |
| empirical | 50 | 2,5 | 0,5 | 5 | 31 | 0 | 50 | Large |
| empirical | 50 | 2,5 | 0,1 | 5 | 31 | -0,5 | 50 | Large |
| empirical | 50 | 2,5 | 0,1 | 5 | 31 | -1.2 | 50 | Large |
| empirical | 50 | 2,5 | 0,1 | 5 | 31 | 0 | 10 | Medium |
| empirical | 50 | 2,5 | 0,1 | 5 | 31 | 0 | 100 | Small |

Table 2. Conditions for determining the starting value of the admixture coefficient T .

| C (mg m ⁻³) | Cyanobacteria flag | T |
|---------------------------|--------------------|-----|
| > 20 | True | 0.9 |
| > 20 | False | 0.1 |
| < 20 | True or False | 0.5 |

Table 3. Mean estimated and measured size parameters (μm) and type in percent cyanobacteria.

| Case | Measured | | Modelled | |
|------------|-----------|-------|-----------------|------|
| | D_{eff} | ESD | \bar{D}_{eff} | T |
| Oligo | 9 | 27 | 23 | 0.44 |
| Meso_dino | 18 | 32 | 27 | 0.32 |
| Meso_mixed | 10 | 17 | 32 | 0.22 |
| Dino | 30 | 32 | 24 | 0.11 |
| Cyano | 5 | 5 | 5 | 0.93 |

Table 4. log-RMSE and non-parametric Spearman's rank regression results arranged by water type. ESD = equivalent spherical diameter, PC= phycocyanin.

| | Statistic | D_{eff} | ESD | C | PC | $a_{\phi}(440)$ | $a_{gd}(440)$ |
|-------------------|------------------------|---------------|---------------|--------------------|--------------------|-----------------|-----------------|
| Case | | μm | μm | mg m^{-3} | mg m^{-3} | m^{-1} | m^{-1} |
| Oligo-mesotrophic | log-RMSE | 1.09 | 0.40 | 0.99 | 0.41 | 1.00 | 0.89 |
| | R ₂ | 0.02 | 0.23 | 0.67 | 0.95 | 0.82 | 0.66 |
| | <i>p</i> | 0.69 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 |
| | Slope (<i>m</i>) | 0.40 | 0.60 | 2.63 | 1.82 | 1.07 | 0.46 |
| | Intercept (<i>c</i>) | 22.34 | 6.94 | -2.45 | -5.44 | -0.07 | 0.03 |
| | N | 10 | 30 | 33 | 5 | 34 | 34 |
| Hyper-eutrophic | log-RMSE | 0.44 | 0.44 | 1.27 | 1.43 | 1.09 | 1.66 |
| | R ₂ | 0.83 | 0.88 | 0.59 | 0.31 | 0.41 | 0.12 |
| | <i>p</i> | 0.00 | 0.00 | 0.00 | 0.03 | 0.00 | 0.08 |
| | Slope (<i>m</i>) | 0.73 | 0.69 | 0.38 | 0.34 | 0.37 | 0.27 |
| | Intercept (<i>c</i>) | 1.44 | 1.42 | 84.56 | 1062.86 | 10.42 | 0.17 |
| | N | 24 | 25 | 23 | 15 | 25 | 25 |
| Combined | log-RMSE | 0.66 | 0.41 | 1.09 | 1.22 | 1.02 | 1.25 |
| | R ₂ | 0.30 | 0.67 | 0.64 | 0.35 | 0.46 | 0.29 |
| | <i>p</i> | 0.00 | 0.00 | 0.00 | 0.01 | 0.00 | 0.00 |
| | Slope (<i>m</i>) | 0.62 | 0.74 | 0.39 | 0.36 | 0.39 | 0.34 |
| | Intercept (<i>c</i>) | 7.94 | 2.04 | 41.89 | 739.55 | 4.01 | 0.12 |
| | N | 34 | 55 | 56 | 20 | 59 | 59 |

7. References

- Aurin, D. A., and H. M. Dierssen. 2012. Advantages and limitations of ocean color remote sensing in CDOM-dominated, mineral-rich coastal and estuarine waters. *Remote Sens. Environ.* **125**: 181-197.
- Bernard, S., T. A. Probyn, and R. G. Barlow. 2001. Measured and modelled optical properties of particulate matter in the southern Benguela. *S. Afr. J. Sci.* **97**: 410-420.
- Bernard, S., F. A. Shillington, and T. A. Probyn. 2007. The use of equivalent size distributions of natural phytoplankton assemblages for optical modeling. *Opt. Express.* **15**:1995-2007.
- Bracher, A., Bouman, H. A., Brewin, R. J., Bricaud, A., Brotas, V., Ciotti, A. M., ... and N. J. Hardman-Mountford. 2017. Obtaining phytoplankton diversity from ocean color: a scientific roadmap for future development. *Frontiers in Marine Science.* **4**: 55.
- Chami, M., D. McKee, E. Leymarie, and G. Khomenko. 2006. Influence of the angular shape of the volume-scattering function and multiple scattering on remote sensing reflectance. *Appl. Optics.* **45**: 9210-9220.
- Dierssen, H. M., R. M. Kudela, J. P. Ryan, and R. C. Zimmerman. 2006. Red and black tides: Quantitative analysis of water-leaving radiance and perceived color for phytoplankton, colored dissolved organic matter, and suspended sediments. *Limnol. Oceanogr.* **51**: 2646-2659.
- Doxaran, D., N. C. Cherukuru, S. J. Lavender, and G. F. Moore. 2004. Use of a Spectralon panel to measure the downwelling irradiance signal: case studies and recommendations. *Appl. Optics.* **43**: 5981-5986.
- Dupouy, C., D. Benielli-Gary, J. Neveux, Y. Dandonneau, and T. K. Westberry. 2011. An algorithm for detecting *Trichodesmium* surface blooms in the South Western Tropical Pacific. *Biogeosciences.* **8**: 3631-3647.
- Evers-King, H., S. Bernard, L. Robertson Lain, and T.A. Probyn. 2014. Sensitivity in reflectance attributed to phytoplankton cell size: forward and inverse modelling approaches. *Opt. Express.* **22**: 11536-11551.
- Fournier, G. R. and L. J. Forand. 1994. Analytic phase function for ocean water. *In* J. F. Jaffe [ed.], *Proc. SPIE 2258, Ocean Optics XII.* **194**: 194-201.
- Gallegos, C. L., R. J. Davies-Colley, and M. Gall. 2008. Optical closure in lakes with contrasting extremes of reflectance. *Limnol. Oceanogr.* **53**: 2021-2034.
- Gregg, W. W. and L. K. Carder. 1990. A simple spectral solar irradiance model for cloudless maritime atmospheres. *Limnol. Oceanogr.* **35**: 1657-1675.
- Hansen, J. and L. Travis. 1974. Light scattering in planetary atmospheres. *Space Sci. Rev.* **16**: 527-610.

Hu, C., J. Cannizzaro, K. L. Carder, F. E. Muller-Karger, and R. Hardy. 2010. Remote detection of *Trichodesmium* blooms in optically complex coastal waters: Examples with MODIS full-spectral data. *Remote Sens. Environ.* **114**: 2048-2058.

Hunter, P. D., A. N. Tyler, L. Carvalho, G. A. Codd, and S. C. Maberly. 2010. Hyper-spectral remote sensing of cyanobacterial pigments as indicators for cell populations and toxins in eutrophic lakes. *Remote Sens. Environ.* **114**: 2705-2718.

IOCCG. 2006. Remote sensing of inherent optical properties: Fundamentals, tests of algorithms, and applications. International Ocean-Colour Coordinating Group.

Kutser, T., L. Metsamaa, and A. G. Dekker. 2008. Influence of the vertical distribution of cyanobacteria in the water column on the remote sensing signal. *Estuar. Coast. Shelf S.* **78**: 649-654.

Lee, Z., K. Du, K. Voss, G. Zibordi, B. Lubac, R. Amone, A. Weidemann, and R. Arnone. 2011. An inherent-optical-property-centered approach to correct the angular effects in water-leaving radiance. *Appl. Optics*, **50**: 3155-3167.

Lu, Y., Li, L., Hu, C., Zhang, M., Sun, S., and C. Lv. 2016. Sunlight induced chlorophyll fluorescence in the near-infrared spectral region in natural waters: Interpretation of the narrow reflectance peak around 761 nm. *J. Geophys. Res. Oceans*, **121**(5): 5017–5029.

Matthews, M. W. and S. Bernard. 2013a. Using a two-layered sphere model to investigate the impact of gas vacuoles on the inherent optical properties of *Microcystis aeruginosa*. *Biogeosciences*. **10**: 8139-8157.

Matthews, M. W. and S. Bernard. 2013b. Characterizing the Absorption Properties for Remote Sensing of Three Small Optically-Diverse South African Reservoirs. *Remote Sens.* **5**: 4370-4404.

Matthews, M. W., S. Bernard, and L. Robertson. 2012. An algorithm for detecting trophic status (chlorophyll-a), cyanobacterial-dominance, surface scums and floating vegetation in inland and coastal waters. *Remote Sens. Environ.* **124**: 637-652.

McKee, D., M. Chami, I. Brown, V. S. Calzado, D. Doxaran, and A. Cunningham. 2009. Role of measurement uncertainties in observed variability in the spectral backscattering ratio: a case study in mineral-rich coastal waters. *Appl. Optics*. **48**: 4663-4675.

Mobley, C. (1999). Estimation of the remote-sensing reflectance from above-surface measurements. *Appl. Optics*. **38**: 7442-7455.

Mobley, C. D., L. K. Sundman, and E. Boss. 2002. Phase function effects on oceanic light fields. *Appl. Optics*. **41**: 1035-1050.

Mueller, J. L., A. Morel, R. Frouin, C. Davis, R. Arnone, K. Carder, Z. Lee, R. Steward, S. Hooker, C. D. Mobley, S. McLean, B. Holben, M. Miller, C. Pietras, K. D. Knobelspiesse, G. S. Fargion, J. Porter, and K. Voss. 2003. Radiometric Measurements and Data Analysis Protocols, p. 1-78. *In* J. L. Mueller, G.

S. Fargion, and C. R. McClain [eds.], Ocean Optics Protocols For Satellite Ocean Color Sensor Validation, Revision 4, Volume III. National Aeronautical and Space Administration.

Nelder, J. A., and R. Mead. 1965. The downhill simplex method. *Comput. J.* **7**: 308-310.

Neukermans, G., H. Loisel, X. Meriaux, R. Astoreca, and D. McKee. 2012. In situ variability of mass-specific beam attenuation and backscattering of marine particles with respect to particle size, density, and composition. *Limnol. Oceanogr.* **57**: 124-144.

O'Donnell, D., S. W. Effler, C. M. Strait, and G. A. Leshkevich. 2010. Optical characterizations and pursuit of optical closure for the western basin of Lake Erie through in situ measurements. *J. Great Lakes Res.* **36**: 736-746.

Piskozub, J. and D. McKee. 2011. Effective scattering phase functions for the multiple scattering regime. *Opt. Express.* **19**: 4786-4794.

Raj, E. P., P. C. S. Devara, R. S. Mahes Kumar, G. Pandithurai, K. K. Dani, S. K. Saha, S. M. Sonbawne, and Y. K. Tiwari. 2004. Results of Sun Photometer-Derived Precipitable Water Content over a Tropical Indian Station. *J. Appl. Meteorol.* **43**: 1452-1459.

Rehm, E. and C. D. Mobley. 2013. Estimation of hyperspectral inherent optical properties from in-water radiometry: error analysis and application to in situ data. *Appl. Optics.* **52**: 795-817.

Retalis, A., D. G. Hadjimitsis, S. Michaelides, F. Tymvios, N. Chrysoulakis, C. R. I. Clayton, and K. Themistocleous. 2010. Comparison of aerosol optical thickness with in situ visibility data over Cyprus. *Nat. Hazards Earth Syst. Sci.* **10**: 421-428.

Reynolds, C. S. 2006. *The Ecology of Phytoplankton*. Cambridge University Press.

Richardson, L. L. (1996). Remote Sensing of Algal Bloom Dynamics. *BioScience.* **46**: 492-501.

Robertson Lain, L., S. Bernard, and H. Evers-King. 2014. Biophysical modelling of phytoplankton communities from first principles using two-layered spheres: Equivalent Algal Populations (EAP) model. *Opt. Express.* **22**: 16745–16758.

Schluter, L., T. L. Lauridsen, G. Krogh, and T. Jorgensen. 2006. Identification and quantification of phytoplankton groups in lakes using new pigment ratios - a comparison between pigment analysis by HPLC and microscopy. *Freshwater Biol.* **51**: 1474-1485.

Seppälä, J., P. Ylöstalo, S. Kaitala, S. Hällfors, M. Raateoja, and P. Maunula. 2007. Ship-of-opportunity based phycocyanin fluorescence monitoring of the filamentous cyanobacteria bloom dynamics in the Baltic Sea. *Estuar. Coast. Shelf S.* **73**: 489-500.

Simis, S. G. H., S. W. M. Peters, and H. J. Gons. 2005. Remote sensing of the cyanobacterial pigment phycocyanin in turbid inland water. *Limnol. Oceanogr.* **50**: 237-245.

Snyder, W. A., R. A. Arnone, C. O. Davis, W. Goode, R. W. Gould, S. Ladner, G. Lamela, W. J. Rhea, R. Stavn, M. Sydor, and A. Weidemann. 2008. Optical scattering and backscattering by organic and inorganic particulates in U.S. coastal waters. *Appl. Optics*. **47**: 666-677.

Stramski, D., A. Bricaud, and A. Morel. 2001. Modeling the inherent optical properties of the ocean based on the detailed composition of the planktonic community. *Appl. Optics*. **40**: 2929-2945.

Stumpf, R. P., T. T. Wynne, D. B. Baker, and G. L. Fahnenstiel. 2012. Interannual variability of cyanobacterial blooms in Lake Erie. *PLoS ONE*. **7**: e42444.

Sun, D., Y. Li, Q. Wang, J. Gao, H. Lv, C. Le, and C. Huang. 2009. Light scattering properties and their relation to the biogeochemical composition of turbid productive waters: a case study of Lake Taihu. *Appl. Optics*. **48**: 1979-1989.

Tzortziou, M., J. R. Herman, C. L. Gallegos, P. J. Neale, A. Subramaniam, L. W. Harding, and Z. Ahmad. 2006. Bio-optics of the Chesapeake Bay from measurements and radiative transfer closure. *Estuar. Coast. Shelf S.* **68**: 348-362.

Whitmire, A. L., E. Boss, T. J. Cowles, and W. S. Pegau. 2007. Spectral variability of the particulate backscattering ratio. *Opt. Express*. **15**: 7019-7031.

Xi, H., Hieronymi, M., Krasemann, H., and R. Röttgers, 2017. Phytoplankton group identification using simulated and in situ hyperspectral remote sensing reflectance. *Frontiers in Marine Science*. **4**: 272.

Xi, H., Hieronymi, M., Röttgers, R., Krasemann, H., and Z. Qiu, 2015. Hyperspectral differentiation of phytoplankton taxonomic groups: a comparison between using remote sensing reflectance and absorption spectra. *Remote Sensing*. **7(11)**: 14781-14805.

Zhou, B., Shang, M., Wang, G., Zhang, S., Feng, L., Liu, X., Wu, L. and Shan, K., 2018. Distinguishing two phenotypes of blooms using the normalised difference peak-valley index (NDPI) and Cyano-Chlorophyta index (CCI). *Science of the Total Environment*. **628**: 848-857.

Zhou, W., G. Wang, Z. Sun, W. Cao, Z. Xu, S. Hu, and J. Zhao. 2012. Variations in the optical scattering properties of phytoplankton cultures. *Opt. Express*. **20**: 11189-11206.